Evaluation of Cross-Survey Research Methods for the Estimation of Low-Incidence Populations

Author: Raquel Magidin de Kramer

Boston College Electronic Thesis or Dissertation, 2016

Boston College
Lynch School of Education

Department of
Educational Research, Measurement, and Evaluation

EVALUATION OF CROSS-SURVEY RESEARCH METHODS FOR THE ESTIMATION OF LOW-INCIDENCE POPULATIONS

Dissertation
by

RAQUEL MAGIDIN DE KRAMER

submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

December 2016

Abstract

Evaluation of Cross-Survey Research Methods for the Estimation of the Proportion of Low-Incidence Populations

Raquel Magidin de Kramer, Author

Henry Braun, Dissertation Chair

This study evaluates the accuracy, precision, and stability of three different methods of cross-survey analysis in order to determine their suitability for estimating the proportions of low-incidence populations. Population parameters of size and demographic distribution are necessary for planning and policy development. The estimation of these parameters for low-incidence populations poses a number of methodological challenges. Cross-survey analysis methodologies offer an alternative to generate useful, low-incidence population estimates not readily available in today's census without conducting targeted, costly surveys to estimate group size directly.

The cross-survey methods evaluated in the study are meta-analysis of complex surveys (MACS), pooled design-based cross-survey (PDCS), and Bayesian multilevel regression with post-stratification (BMRP). The accuracy and precision of these methods were assessed by comparing the estimates of the proportion of the adult Jewish population in Canada generated by each method with benchmark estimates. The stability of the estimates, in turn, was determined by cross-validating estimates obtained with data from two random stratified subsamples drawn from a large pool of US surveys.

The findings of the study indicate that, under the right conditions, cross-survey methods have the potential to produce very accurate and precise estimates of low-incidence populations. The study did find that the level of accuracy and precision of these estimates varied depending on the cross-survey method used and on the conditions under which the estimates were produced. The estimates obtained with PDCS and BMRP methodologies were more accurate than the ones generated by the MACS approach. The BMRP approach generated the most accurate estimates. The pooled design-based cross-survey method generated relatively accurate estimates across all the scenarios included in the study. The precision of the estimates was found to be related to the number of surveys considered in the analyses.

Overall, the findings clearly show that cross-survey analysis methods provide a useful alternative for estimation of low-incidence populations. More research is needed to fully understand the factors that affect the accuracy and precision of estimates generated by these cross-survey methods.

Acknowledgements

There are many people to whom I owe thanks. I could not have completed this dissertation without the help and support of my advisors and colleagues, as well as the support and encouragement of my family and friends.

I would like to start by thanking my dissertation committee. I am very grateful for Prof. Henry Braun's willingness to give me a chance with an unconventional topic. I greatly appreciated all of his invaluable guidance and advice as well as his generous patience and continuous support throughout the process. Dr. Michael Russell and Dr. Laura O'Dwyer also offered much thoughtful advice and support for the dissertation, and I am so grateful for the opportunities they gave me to work and learn from them throughout my time at Boston College. I would like to acknowledge also my sincere appreciation to Prof. Leonard Saxe not only for his extensive guidance and support as a member of my committee, but also for his invaluable mentorship, which enabled the realization of this dissertation.

My warmest gratitude to Dr. Elizabeth Tighe, a mentor and friend. I could not have done this without her guidance, critical feedback and continued support and encouragement. I am so grateful for her generosity with her time and knowledge.

Thank you to all of my friends and colleagues at the Steinhardt Social Research for your continuing support.   Special thanks to Dr. Janet Aronson, Rachel Bernstein, Eve Farber, Dr. Annette Koren and Daniel Parmer. To Rachel and Eve, thanks for all your help with the arduous task of data collection and organization. Thank you, Janet Rachel and Daniel, for your invaluable help as I organized my thoughts and put them in writing.

Thank you to Annette and Janet for taking me out and for your time and company when I needed.

My sincere gratitude goes to my editor and friend Judith Filc. Thank you, Judith, for taking the time to understand my work and making it readable to the rest of the world.

I am so grateful to have such a loving, caring, and supportive group of family and friends who saw me through the dissertation process. To my friends, I thank each of you for your continued enthusiasm and support. To my parents and sisters: thank you for being always there and for keeping me sane. To my children Ron and Gal, I do not have words to thank you enough for your love and care. Gal, thanks for the innumerous hours you sat with me helping me put the ideas in writing and revising my drafts. Ron, thanks for always reminding me of what is important. I am so proud of both of you. Lastly to my husband Dario, thank you for always believing I could do it and for never letting me give up. I could not have done this without your love, patience, and continuous support.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Overview

The purpose of this study is to contribute to the existing body of research on cross-survey analysis as a method for estimating the proportions and demographic characteristics of low-incidence populations. Cross-survey analysis refers to the combined analysis of data from different surveys. The study evaluates the operating characteristics of three different methods of cross-survey analysis in order to determine their suitability for estimating the proportions of low-incidence populations. This research serves two main purposes. First, it addresses the need to find additional methods to generate low-incidence population estimates that are currently not readily available. Second, it further advances the understanding of the use and limitations of different cross-survey methods.

While the use of cross-survey methodologies is growing (e.g., Gelman, 2009; Tighe, Livert, Barnett & Saxe, 2010), there is limited research on the accuracy of the different methods. This study extends current research by assessing the accuracy, precision, and stability of estimates that result from the application of three different methods for cross-survey analysis. It assesses accuracy and precision by comparing the estimates generated by each method with benchmark estimates. The proportions of Canadian low-incidence religious groups obtained with these methods are compared with estimates from the Canadian census. To assess stability, this study provides cross-validation of estimates generated from different subsamples drawn from a large pool of surveys of the US population.

Background

The motivations for this research are (a) the need for alternative methodologies to estimate the size and distribution of low-incidence populations, especially those populations that are not measured directly by a public census; (b) the need to further understand the limitations and strengths of the latest methods of cross-survey estimation of low-incidence populations; and (c) the importance and need for systematic data for the study of rare religiously defined groups. At the end of this section I include an outline of the three different cross-survey methods used in the analysis.

The Need to Develop Alternative Methods to Estimate the Size and Demographic Distribution of Low-Incidence Populations

The use of parameter estimates of size and demographic distribution is manifold. Population parameters are needed for planning and policy development, as well as for the analysis of survey research. Decision-makers at the local, national, and international levels use these estimates to allocate funds and make informed policy decisions in the areas of education, healthcare services, and economic development. In the last few decades there has been a growing demand from both the public and private sectors for reliable population parameters of low-incidence populations (Gosh & Rao, 1994; Pfefferman, 2002). Information on population size and demographic distribution is also relevant in the evaluation of survey research studies. This information is often used as auxiliary data to calibrate surveys and correct for coverage, sampling, and non-response errors in survey research (Groves et al., 2004; Kim, Li, & Valliant, 2007; Valliant, Dever, & Kreuter, 2013).

The U.S. Census Bureau is the leading source of demographic and economic data for this country. The Current Population Survey (CPS) and the American Community Survey (ACS) provide extensive data collection capabilities in a vast variety of subject areas, including housing conditions and demographic, social, and economic features. Nevertheless, there are limitations to the information provided by the census in terms of coverage and precision. Regarding coverage, for example, the U.S. Census Bureau does not collect data on religious affiliation (U.S. Census Bureau, 2014). Neither does it gather specific health information. Smoking status, for instance, is only available in a limited number of records (U.S. Census Bureau, 2013). In terms of precision, the data collected by the U.S. Census Bureau or other government organizations are often insufficient to provide reliable estimates for small geographical areas or subpopulations. As a result of these limitations, alternative data sources are required to estimate the size and features of low-incidence populations.

Population-based estimates may also be secured from large probability based sampling surveys. Yet, obtaining estimates of low-incidence populations in this way is extremely expensive. The large overall sample size needed for the estimation of a small incidence population in small geographical areas often exceeds the survey's funding and capabilities (e.g., Korn & Graubard, 1999; Rao, 2003).

An alternative methodology called Small Area Estimation (SAE) has been developed over the last 30 years in response to the demand for reliable estimations for small areas (Pfeffeman, 2002; Rao, 2003; Rao & Ghosh, 1994). SAE methodologies have in common the use of a) indirect, model-dependent estimation and b) auxiliary, related data such as administrative or alternative census data to improve estimation

(Pfeffeman, 2002; Rao, 2003; Rao & Ghosh, 1994). One of the cross-survey methods assessed in this study, Bayesian multilevel estimation with post-stratification, is based on this methodology. The drawback of SAE methodologies, however, is their need for auxiliary data, which are not always readily available for low-incidence population groups, especially those groups that are not assessed by censuses.

Cross-Survey Analysis

Traditional meta-analytic methods combine multiple data sources to increase the accuracy of final estimates (Bartolucci & Hillegass, 2010). These types of methods have customarily been developed and applied to combine information from independent clinical trials and observational studies. In the context of this proposal, cross-survey methods refer to meta-analytic methods combining data from multiple surveys.

The potential value of such combination has only received attention in recent years. There are several applications of these methods in the social, political, biomedical, and natural sciences (Gelman, 2009; Korn & Graubard, 1999; Pfeffermann, 2013; Rao et al., 2008; Schenker & Raghunathan, 2007). Cross-survey methods have also been proposed or adopted as an alternative to develop population estimates when demographic data required for post-estimation are unavailable (e.g., Gellman, 2009; Gellman & Hill, 2007; Tighe et al., 2010).

Although there is an increase in the use of cross-survey methods, they are not always properly applied. As Fox (2010) points out, “researchers have started to employ many different techniques including meta-analysis; however, the analysis is often done without reference to a generalized framework or a systematic review and often without an understanding of the methodological differences between surveys and experiments" (p.

527). This study contributes to the understanding and application of these methods through direct comparisons of the different frameworks that examine the methodological differences.

There are different approaches to cross-survey analysis. This dissertation evaluates three of them. These are Meta-Analysis of Complex Surveys, Pooled Design-Based Cross-Survey, and Bayesian Multilevel Regression with Post-Stratification. What follows is a brief description of each of these approaches. The cross-survey methods are further described in Chapters 2 and 3.

Meta-Analysis of Complex Surveys. This method is based on the framework for the meta-analysis of complex survey data described by Fox (2011). The method is appropriate for combining results of survey research studies that (a) employ probability samples and (b) represent the same underlying population. Although this framework builds on the ideas of traditional meta-analysis methods for combining results of experimental studies, it also takes into account the particular characteristics of survey research studies that employ probability samples, such as sample design and weighting. Meta-analysis is most commonly applied to estimate effect size. It has also been used, however, to estimate descriptive quantities. Similar to traditional meta-analysis, the meta-analysis of complex surveys follows an approach whereby estimates are obtained independently from individual surveys and the overall estimator is a function of these separate estimates (Roberts and Binder, 2009).

Pooled Design-Based Cross-Survey. The second approach is based on the framework presented by Korn and Graubard (1999). This framework uses a pooled sample technique whereby individual records from all the surveys are combined into one sample (Roberts & Binder, 2009). The dataset is then treated as a sample from a single population (Thomas & Wannell, 2009). This method includes two main steps, namely, recalculation of sample weights and parameter estimation. Sample weights are recalculated in order to take into account survey-specific sample sizes and designs (Korn & Graubard, 1999).

Bayesian Multilevel Regression with Post-Stratification (BMRP). The third cross-survey approach is based on the multilevel regression and post-stratification framework developed by Park, Gelman, and Bafumi (2004) and described in detail in Gelman and Hill (2007). It is a model-based pooled sample method. As is the case with the pooled design-based cross-survey, individual records from all the surveys are combined into a single sample. Each record in the pooled sample includes both data from individual respondents and information pertinent to the survey. This method is model-based in that sampling and survey design variables are included as factors in the analysis so that their relationship to the population estimates can be controlled for, examined and explained directly (Little, 2004; Tighe et al., 2010).

The Study of Religious Groups

There is substantial theoretical and practical interest in the study of religious groups (Tighe et al., 2010). Religious orientation has been associated with a host of social behaviors, including educational attainment and decision-making (Schieman, 2011), attitudes toward teaching creationism in public schools (Baker, 2013), voting and involvement in politics (e.g., Gelman, 2009; Mattis, 2001), family life, health behavior, and social capital (Sherkat & Ellison, 1999).

Unlike many other countries, the U.S. Census Bureau does not collect information on religion despite the existing interest in the study of religious groups. The absence of population data to describe the distribution and demographic composition of religiously defined populations represents a challenge to scholars of American religion (e.g., Lim, 2013; Perl, Greely and Gray, 2006; Tighe et al., 2010; Tighe et al., 2013). Without known religious population parameters, there are few means of calibrating surveys involving these groups. The absence of such parameters makes it hard to assess the representativeness of samples designed to generalize to specific religious groups. Thus, it is difficult to interpret the findings from targeted surveys of religious groups.

This study evaluates the three different cross-survey methods by estimating the Jewish population in Canada and the US. I chose religious affiliation as the subject for two reasons. First, there is a strong theoretical and practical interest in the study of religious groups, paired with a lack of US related census population estimates. Second, the evaluation of religious population estimates measured by the Canadian census facilitates the evaluation of the accuracy of cross-survey methods by providing an external benchmark not available for analysis of US data. As for the decision to focus on the estimation of Jewish populations, it was prompted by the low incidence of the population in both the US and Canada and the sizable body of research and data about the use of cross-survey methods to estimate this population in the US.

Research Purpose and Research Questions

Research purpose: To examine the feasibility of using cross-survey methods to estimate the proportions of low-incidence populations.

The purpose of this dissertation is to assess the relative utility of different methods of cross-survey analysis for estimating the proportions of low-incidence populations. Given the interest in the study of religion and the lack of related population estimates in the US, this study looks at the use of cross-survey analyses in research on low-incidence religious populations. I examine and compare the operational characteristics of three cross-survey analysis methods for estimating the proportions of the total adult Jewish population in Canadian provinces and metropolitan areas as well as in U.S. metropolitan areas.

Research Questions

This dissertation seeks to answer the following research questions:

- How do cross-survey estimates of the proportions of the adult Jewish population in Canadian provinces and in metropolitan areas compare to the estimates from the Canadian census (2001) and Canadian National Household Survey (NHS)?
- How do the results of the cross-survey approaches compare in terms of their accuracy and precision in estimating the proportions of the total adult Jewish population in the Canadian provinces and metropolitan areas included in the study?
- How do the three cross-survey approaches compare in terms of their stability and precision for estimating the proportions of the total adult Jewish population in metropolitan areas in the continental US?

Research Design and Methods

This section briefly describes the research design and methods used to answer the research questions. Chapter 3 provides more detailed information on the methodology employed.

The study explores the empirical properties of three different methods of cross-survey analysis for the estimation of the proportions of the adult Jewish population in the largest metropolitan areas of the US and Canada. Evaluation and cross-validation studies were conducted for this purpose. The analyses were based on data from nationally representative random samples of the adult population in Canada and the US. The study includes three main sets of surveys: 2001 Canada batch (Batch Ca2001), 2011 Canada batch (Batch Ca2011), and 2011 US batch (Batch US2011). Batch Ca2001 includes data from surveys with nationally representative random samples of the adult population in Canada fielded between 1997 and 2004. Batch Ca2011 comprises data from surveys of nationally representative random samples of the adult population in Canada conducted between 2006 and 2014. Batch US2011 consists of surveys with nationally representative random samples of the adult population in the continental US carried out between 2008 and 2014. Estimates of the proportions of the adult Jewish population were generated for each of the batches and for two random subsamples of the US2011 batch (see detailed description in Chapter 3).

The estimates generated using the three cross-survey analyses are compared and evaluated to answer the research questions. Collating the estimates generated from the Canadian batches with census population estimates provides information as to the accuracy and precision of the different methods. Population parameters of the adult

Jewish population in Canadian provinces and metropolitan areas were obtained from the 2001 Canadian census and from the 2011 National Household Survey (NHS) conducted by Statistics Canada. Cross-validating the estimates obtained from different subsamples of the US batch contributes to assessing the stability of the estimates generated by the different methods. Finally, comparing all the estimates obtained from the different cross-survey methods across the three main batches offers additional information regarding the stability and precision of these methods.

The study consisted of three stages: (a) data preparation, (b) population estimation using cross-survey analysis, and (c) comparative evaluation of cross-survey methods. Chapter 3 includes comprehensive information on each stage of the study.

Significance of the Study

The estimation of the parameters of size and demographic distribution of low-incidence populations presents a number of methodological challenges (Gosh & Rao, 1993; Tighe et al., 2010). As noted earlier, the U.S. Census Bureau is the leading source of demographic and economic data for the country. Nevertheless, there are limitations to the information provided by the census in terms of coverage and precision. In addition, surveying rare populations accurately is very difficult and costly (Kalton, 2009).

Cross-survey analysis methodologies offer an alternative to study low-incidence populations (Gellman, 2009; Gellman & Hill, 2007; Tighe et al., 2010). As Fox (2011) asserts, however, "while the literature on studies that pool survey data is growing at an astounding rate, the literature on how to pool this data is not" (p. 1). This study expands the existing body of research by exploring the empirical properties of three different methods of cross-survey analysis to estimate the proportion of low-incidence population groups. Comparison of the estimates with an external criterion and systematic comparisons among the methods shed light on the appropriateness and limitations of each method.

The object of this study is religious affiliation. Despite substantial theoretical and practical interest in the study of religious groups, the U.S. Census does not collect data on religious identification (U.S. Census Bureau, 2014).

Researchers at Brandeis University's Steinhardt Research Institute have employed one of these approaches – Bayesian multilevel estimation with post-stratification – to estimate the proportion and characteristics of the adult Jewish population in the US. This method has proven useful. Nevertheless, given the lack of census population counts, it is very difficult to validate the estimates that result from this cross-survey approach. Testing the accuracy of this method against external criteria for the Canadian data offers insights regarding its potential to generate parameters of size and demographic distribution of the total adult Jewish population.

Chapter 2: Review of the Literature

Introduction

This literature review covers two main topics: (a) research into cross-survey theory and analysis, with a focus on the cross-survey methods that are used in this study and (b) sociodemographic research on one particular low-incidence religious group in the US and Canada, namely, the Jewish population. These two topics provide the background for the current study.

The first section begins with a discussion of cross-survey methods. First it presents the rationale and theory behind cross-survey analysis, and then it addresses the specific methods that are used in this study. The second section, in turn, explores the history and conditions of sociodemographic research on the Jewish population in the US and Canada.

Cross-Survey Methods

Greater access to data has led to a proliferation of analyses that group or pool results from multiple studies (Fox, 2011). In addition, with the increased availability of more than one survey containing similar variables, the integration of studies has expanded in recent years from combining randomized control trials to combining survey data (Fox, 2011; Roberts & Binder, 2009). The application of methods that combine data from different surveys can be found in a vast variety of fields, ranging from the social and political sciences to the medical and natural sciences (Rao et al., 2008; Gelman, 2009; Korn & Graubard, 1999; Pfeffermann, 2013; Schenker & Raghunathan, 2007). Survey methodologists resort to different names to refer to methods that combine survey data. In this study I use the term *cross-survey methods*.

There are many reasons for combining data from two or more surveys. They can stem from an interest in combining multiple frames from the same survey (e.g., Lohr, 2006; Lohr & Rao, 2006) and rolling samples from different periods (e.g., Kish, 1994, 1999), or a desire to combine surveys that have either a common target population or a common domain, among others. As a result of this diversity, cross-survey methods also differ in the type of data they are intended to combine. Multiple-frame surveys, periodic samples, and rolling samples are some of the surveys whose design is integral to the idea of sample combination. Kish (1994, 1999) and Lohr (2006) developed methodologies to combine these types of surveys. Yet the methodologies they advanced are not entirely suitable for surveys that were not originally developed to be combined. Given the purpose and design of this study, the literature review focuses on cross-survey methods that allow the combination of data from different types of surveys.

Analysts combine data from two or more surveys for a variety of reasons. A common one is to borrow strength and increase coverage and effective sample size (Gelman, 2014; Rao, 2003; Fox, 2011). It is often the case that the population or phenomenon under study is rare. The expectation is that increasing the overall sample size should lead to reduced sampling errors (Roberts & Binder, 2009), lower bias, and greater precision (Shenker, 2014). Additional reasons include taking advantage of the varied strengths of different studies (Shenker, 2014); for instance, by combining data from multiple or complementary frames to increase coverage (Roberts & Binder, 2009; Schenker, Gentleman, Rose, Hing, & Shimizu, 2002).

Considerations regarding the Implementation of Cross-Survey Analyses.

There are a number of aspects that must be taken into account when combining data from different surveys. In their discussion of the combination of health surveys, Rao et al. (2008) stress the importance of paying special attention to sampling design and variance estimation. They argue that unlike random clinical trials (RCT), where data are randomized to conditions, health surveys often yield cross-sectional data from probability samples where each unit in a target population has a positive probability of being sampled for measurement. The survey design of each study has an impact on the variance of the individual study and, in turn, the variances of individual studies have an impact on the summary estimates resulting from combining surveys.

Similarly, Fox (2011) points to the importance of distinguishing between randomization frameworks in experimental and survey designs. In an experimental framework it is the assignment of individuals that is random. Such randomization is intended to reduce confounding factors and obtain internally valid results. In a design-based framework, by contrast, what is random is the selection of individuals. The purpose of randomization in this case is to generalize to the finite population from which the sample was drawn. Fox argues that failing to acknowledge these differences and applying classic meta-analysis methodologies to combine survey data might lead to erroneous conclusions.

An additional characteristic that needs to be considered when combining survey data is the target population. Researchers emphasize the need to review its definition in individual surveys so that the resulting summary estimates will apply to a meaningful population (Rao et al., 2008; Schenker et al., 2002).

Schenker et al. (2002) suggest taking into account the following additional questions when combining surveys: Do the surveys cover similar time periods? Can their data be statistically combined? Do they ask similar questions? Are their sources of information similar? Have the surveys changed over time? Do they ascertain the features of interest accurately and comparably? Do apparently similar questions have different meaning or applicability between surveys? Do the separate estimates have face validity? If the answers to these questions are not clearly formulated and properly interpreted, researchers may reach the wrong conclusions.

Background Definitions

In this section I briefly present some terms that are commonly used in survey analysis in general and in cross-survey analysis in particular. These terms serve to describe the methods assessed in this study.

Finite Population and Superpopulation. Statistically, a finite population can be defined as a finite collection of units to which inferences are to be made (Lepowsky, 2008). A population is called finite when it is possible to count its units. Sampling from this type of population is one of the building blocks of survey sampling theory and practice (Lachan, 2008). Complex design surveys are surveys of finite populations where the samples are selected by way of sampling schemas that are different from simple random sampling. Survey weight variables can be used to obtain approximate design-unbiased estimates of finite population quantities, such as means and totals (Binder, 2011).

"The superpopulation approach treats the value associated with a population unit as the realization of a random variable rather than as a fixed number" (Lachan, 2008, p.284). A finite population of interest can thus be regarded as a realization of the infinite population known as superpopulation (Fox 2011; Mallec, 2008).

Quantity of Interest. In most cross-survey analyses there are three types of quantities of interest: simple descriptive, descriptive under an assumed relationship, and analytical quantities. *Simply descriptive* quantities include features of a single finite population, such as means, proportions, and totals. *Descriptive under an assumed relationship* quantities are based on an assumed relationship among the characteristics of the finite target populations of the different surveys. *Analytic quantities* are characteristics or relationships that hold beyond the specific finite populations surveyed, such as the parameters of a superpopulation (Roberts & Binder, 2009). The current study focuses on the simple descriptive quantities of the population.

Approaches to Estimation: Separate and Pooled Approaches to Combining Data. When conducting a cross-survey analysis, data from different surveys may be combined using either a separate or a pooled approach to estimation. With the separate approach, an estimate is independently obtained for each survey. The combined estimate is then a function of the separate estimates. The type of function that combines these estimates depends on the type of quantity being estimated and on whether the separate survey estimates are independent of each other. The meta-analysis of complex surveys method used in this study employs a separate approach to estimation.

The pooled approach, by contrast, requires combining individual records of all the surveys. The estimate is then obtained with techniques appropriate for a single sample and modified weights (Roberts & Binder, 2009). Both the pooled sample cross-survey and the Bayesian multilevel regression with post-stratification methods use the pooled approach for estimation.

Approaches to Estimation: Design-Based and Model-Based inference. Two alternative frameworks commonly used to make statistical inferences from samples to populations are design-based and model-based frameworks (Binder, 2011; Sterba, 2009). Design-based methods rely on probability sampling principles, and model-based methods, on statistical models (Roberts & Binder, 2009).

The design-based approach was originally developed by J. Neyman (1934). Under this framework, samples depend only on the probabilities used to select units from the finite population (Binder, 2011). Statistical inferences, such as testing hypotheses and constructing confidence intervals, are thus based only on the probabilities used to select the samples (Roberts & Binder, 2009; Binder, 2011). This model is mainly concerned with inferences for finite populations (Fox, 2011) and is often adopted to estimate descriptive quantities such as the total or the mean of a response variable (Ghosh, 2009).

The model-based approach may be attributed to R. A. Fisher. Fisher believed that obtaining a random sample from a given population would not always be feasible, particularly in the case of observational studies in sociology and economics (Sterba, 2009; Husson, Curran, & Bauer, 2013). Under this approach, inferences about population parameters are made by way of a statistical model that links the theory to the sample data. This model describes how the dependent variable(s) is/are thought to have been generated (Sterba, 2009). As Streba (2009) states, “the purpose of the statistical model is to provide a link between the observed units in the sample and the unobserved units in the infinite population, enabling causal or analytic inferences to pertain to these unobserved units as well" (p. 3).

An example of a statistical model, described by Sterba (2009), is the linear regression model: $y_i = \beta_0 + \beta_0 x_i + \varepsilon_i$. Under this approach, the dependent variable y_i is generated as a function of a known, fixed independent variable x_i and a known error. In this example, the superpopulation is made up of all possible y-values that could be generated by the model (Royall, 1988), while the observations $y_1 \ldots y_n$ are presumed to be independent realizations from this superpopulation (Binder & Roberts, 2003).

As Ghosh (2009) asserts, "both design- and model-based approaches can be frequentist, where such procedures do not make an explicit use of priors either for the finite population or the superpopulation parameters. In contrast, the Bayesian approach assumes that the response variable associated with any unit is the realization of a random variable following some specified distribution based on prior information" (p 153).

Approaches to Cross-Survey Methods: Description and Applications of the Cross-Survey Methods Assessed in this Study

I focus on three different cross-survey methods that are suitable for combining data from surveys using probability sampling of the same underlying population in order to estimate the proportions of the adult population that is Jewish in provincial and metropolitan areas in Canada and in metropolitan areas in the US. What follows is the description of the three cross-survey methods: meta-analysis of complex surveys (MACS), pooled design-based cross-survey (PDCS), and Bayesian multilevel regression with post-stratification (BMRP).

Meta-Analysis of Complex Surveys (MACS). This method is based on the framework for the meta-analysis of complex survey data presented by Fox (2011). Although it builds on traditional meta-analysis methods for combining experiments, it takes into account the particular characteristics of survey research that employs probability samples. The meta-analysis method uses a separate approach, whereby estimates are obtained independently from individual surveys and the overall estimator is a function of these estimates (Roberts & Binder, 2009). Fox's framework offers three possible estimation models, namely, two fixed-effect models and one random-effect model. This study applies the first fixed-effect model (case1a). Case1a follows a design-based approach. The model is appropriate for combining estimates from samples from a finite population and for estimating a descriptive quantity. Therefore, it is suitable for estimating the proportions of the adult population that is Jewish in metropolitan areas in the US and Canada (a descriptive quantity from a finite population).

Case1a is based on the idea of combining survey samples of a finite population to obtain an estimate of a population parameter. As Fox (2011) explains, survey data under a design-based framework can be used to obtain estimates of the finite population. If data from a single survey are not sufficient to meet the needs of the research, pooling several estimates of the finite population (calculated with finite inference models) may be considered. In the Case1a model, pooling and variance estimation for the finite population take into account the stochastic error due to sampling, and the inference is based on repeated sampling of the population. Figure 2.1 shows the diagram for this model presented by Fox (2011). In the diagram θ_N refers to the population parameter and $\hat{\theta}_N$ is the estimate of the finite population, while $\hat{\theta}^{F(F)}$ denotes a finite population parameter estimate (F) under a design-based sampling framework (F). (In the figure the superpopulation box is shaded to indicate that it is not being used for inferences.)

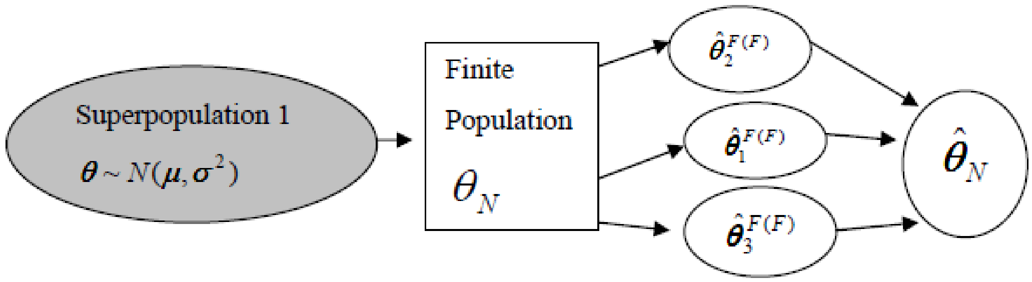

Figure 2.1: Fixed-effect model case1a using sampling concepts. Reprinted from *A framework for the meta-analysis of survey data* (p. 90), by K.M. Fox (2011). (Doctoral dissertation, Queen's University, 2011). Retrieved from http://hdl.handle.net/1974/6900

Under this model, the estimate of a finite population parameter $\hat{\theta}_N$ is the weighted average from all the survey samples (Fox, 2011). If $\hat{\theta}_j$ and w_j are the estimate and relative weight from survey sample *j,* then:

$$\hat{\theta}_N = \frac{\sum_{j=1}^{k} w_j \hat{\theta}_j}{\sum_{j=1}^{k} w_j}$$

As Fox (2011) claims, in a meta-analysis, weights are used to attach greater significance to the more precise estimates when pooling. The author suggests weighting the surveys by their variance and design effect in order to take into account both study precision and sample design. The relative weights (w_j) are thus a function of both the variance (v_j) and the design effect $(deff_j)$ of survey sample *j*. Design effect is a measure of the precision gained or lost due to the use of a complex sample design instead of a simple random sample (SRS) (Kish, 1965):

$$w_j = \frac{1}{\frac{deff_j}{v_j}}$$

The variance of the estimate of the population parameter is calculated as a weighted average of the variances of the different survey samples (Fox, 2011):

$$\hat{v}(\hat{\theta}) = {}^{1}\!\big/_{\sum_{j=1}^{k} w_j}$$

According to Fox (2010), this model makes some assumptions, namely, (a) that the surveys cover the same target population, (b) that individual estimates are unbiased, and (c) that the survey weights are independent of the point estimates. In addition, two less obvious assumptions are being made. First, that the samples are independent, and second, that errors are independent among surveys. Regarding these last assumptions, special attention must be paid when using multiple surveys from large statistics agencies. Large organizations may use overlapping samples for different surveys or different survey designs that show the same methodological flaws (e.g., specific flaws in conducting interviews), plausibly resulting in biased estimates.

In her research Fox (2010) tested the model by means of simulations. She found that under the case1a model the estimator converges to the census parameter as the number of studies grows. This was the case even when the sampling designs of the individual studies were diverse and included surveys with simple random sampling, stratified sampling, and cluster sampling. Fox (2010, 2011) stresses the need to consider individual survey variance when applying the meta-analysis method, in particular when samples are viewed as part of the same finite population (model case1a). This author points out that, although researchers familiar with meta-analysis of experimental design methods may be tempted to apply them to combine survey data, ignoring surveys' individual designs when combining estimates may likely produce erroneous results (she illustrates this point with simulated data).

There are a few examples in the literature that apply a meta-analysis method to the combination of survey data (e.g., Purcell et al., 2012; Thomas & Wannel, 2009; Lansky et al., 2014; Rao et al., 2008; Steel et al., 2009). Nevertheless, in several of these studies it is not clear if the design of the individual surveys was taken into account. Furthermore, when the variance was considered, it is not clear how. Most of the studies that reference meta-analytic methods to combine survey data cite the methodology advanced by Rao et al. (2008). Fox (2013) argues, however, that Rao's proposed methodology fails to "take into account the possibility of clustering in the data with complex sampling designs" (p. 35).

The advantage of using a method based on a separate approach is that it does not require individual data. It can hence be used when there are only survey level estimates available.

Pooled Design-Based Cross-Survey (PDCS). The second method assessed in this dissertation is modeled on the framework presented by Korn and Graubard (1999). This framework follows a design-based pooled sample approach, whereby individual records from all the surveys are combined into a single sample (Roberts & Binder, 2009). The dataset is then treated as a sample from one population (Thomas & Wannell, 2009).

As Korn and Graubard (1999) have underscored, when conducting a pooled analysis of data from survey samples that can be seen as sampling the same population, the original survey weights should be modified. Just adding the survey weights of the individual surveys would lead to overestimation, while averaging the weights (adding them and dividing by the number of individual surveys) might lead to inefficient estimation. Instead, to get unbiased and efficient estimates of a pooled sample, the researchers suggest estimating the population with a weighted mean. This method includes two main steps, namely, recalculation of sample weights and parameter estimation. Recalculation of sample weights is performed in order to take into account survey-specific sample sizes and designs (Korn & Graubard, 1999). The following description was drawn from Korn and Graubard (1999).

Assuming L is the number of survey samples with $n_1, n_2 \ldots n_L$ sample sizes, $\{y_{ij} \mid i=1,\ldots, n_j \ , j = 1, \ldots, L\}$ are the sample values, and $\{w^*_{ij} \mid i = 1, \ldots, n_{j,}\ j = 1, \ldots, L\}$ the revised weights, then the estimated population mean $\bar{y}$ can be written as:

$$\bar{y} = \frac{\sum_{j=1}^{L} \sum_{i=1}^{n_j} w^*_{ij}\, y_{ij}}{\sum_{j=1}^{L} \sum_{i=1}^{n_j} w^*_{ij}}$$

The authors define the revised weight $w^*_{ij} = w_{ij} * k_j$ for some k_j $\{\, k_j \mid\ j = 1, \ldots, L\,\}$. Given that the weighted populations of the survey samples are estimators of the population size:

$$\sum_{j=1}^{L} \sum_{i=1}^{n_j} w^*_{ij} = \sum_{j=1}^{L} \sum_{i=1}^{n_j} w_{ij}/L$$

then:

$$\bar{y} = \frac{(k_1 \sum_{i=1}^{n_1} w_{i1})\bar{y}_1 + \cdots + (k_L \sum_{i=1}^{n_L} w_{iL}\,)\bar{y}_L}{k_1 \sum_{i=1}^{n_1} w_{i1} + \cdots + k_L \sum_{i=1}^{n_L} w_{iL}}$$

To minimize the variance of $\bar{y}$:

$$k_j = \frac{\sum_{i=1}^{n_1} w_{i1} + \cdots + \sum_{i=1}^{n_L} w_{iL}}{L \sum_{i=1}^{n_j} w_{ij}} (1 - \frac{Var(\bar{y}_j)}{Var(\bar{y}_1) + \cdots + Var(\bar{y}_L)})$$

As the authors affirm, the "optimal" k_j depends on the variances of the population estimates ($\bar{y}_j's$) of the different survey samples. Nonetheless, given that estimating the variances may add variability to the weights, they suggest simplifying the equation, instead, by treating the y's as independent and identically distributed with the same variance σ^2:

$$\mathrm{Var}(\bar{y}\,) \cong \left(CV_j^2 + 1\right)\sigma^2/n_j$$

where CV_j is the coefficient of variation of the sample weights in survey sample j. By applying this relation to the previous formula, k_j can be calculated as:

$$k_j = \frac{\sum_{i=1}^{n_1} w_{i1} + \cdots + \sum_{i=1}^{n_L} w_{iL}}{L \sum_{i=1}^{n_j} w_{ij}} (1 - \frac{\left(CV_j^2 + 1\right)/n_j}{(CV_1^2 + 1)/n_1 \; + \cdots + \; (CV_L^2 + 1)/n_L})$$

If the coefficient of variation is approximately equal for all surveys, then the formula can be simplified:

$$k_j = \frac{\sum_{i=1}^{n_1} w_{i1} + \cdots + \sum_{i=1}^{n_L} w_{iL}}{L \sum_{i=1}^{n_j} w_{ij}} (\frac{n_j}{n_1 + \cdots + \; n_L})$$

Since the left fraction in the previous formula is approximately 1, the authors simplify the formula one more time to the form:

$$k_j = \; (\frac{n_j}{n_1 + \cdots + \; n_L})$$

Korn and Graubard (1999) suggest generating the variance of the estimates by way of a replication method of variance estimation. Replication methods repeatedly calculate the parameter estimator on different data subsets. These authors propose using the jackknife method, which recomputes the estimates by removing one primary sampling unit from the dataset at a time. Each replicate provides an estimate of the population parameter of interest, and the variance of the replicated estimates, an estimate of the variance of the point estimate (Hyunshik, 2008). In the context of this study, the primary sampling unit is the individual respondent at the case level.

An additional aspect of Korn and Graubard's work that is being considered for this study is the way they calculate separate estimates for subpopulations (e.g., estimates by age categories). The authors suggest estimating *kj* separately by subgroups.

A number of studies use this methodology, especially when analyzing health surveys (e.g., Baker, Rendall, & Weden, 2015; Chiu, Austin, Manuel, & Tu, 2010; Herrera, 2012).

Bayesian Multilevel Estimation with Poststratification (BMRP). The third cross-survey approach that I evaluate in this study is a model-based pooled sample method. As in the case with the pooled design-based cross-survey method, individual records from all the surveys are combined into a single sample. Each record in the pooled sample includes both data concerning individual respondents and information pertinent to the survey itself. Yet, unlike the previous method, which is design-based, this one is model-based. The model-based approach, as Ghosh (2009) points out, “views the finite population as a sample from a hypothetical superpopulation, and inference for finite population parameters are model-based" (p. 153).

The model-based approach is suitable for the present study. Following Binder and Roberts' (2009) terminology, this approach views the quantity of interest (in this case, the proportions of the adult Jewish population in a metropolitan area in the US or Canada) as an outcome of a logistic regression model. Under this model, the outcome variable is the probability that an adult will be Jewish given specific demographics and geographic location.

In model-based methods, sampling and survey design variables are included in the analysis so that their relationship to the population estimates can be examined and explained directly (Little, 2004; Tighe et al., 2010). A model-based approach may, therefore, account for different sample designs and non-representativeness of key demographic variables (Park, Gelman, & Bafumi, 2004; Tighe et al., 2010).

The model-based approach used in this study builds on the multilevel regression and post-stratification framework (MRP) developed by Park, Gelman, and Bafumi (2004) and on the applied methodology advanced by Gelman and Hill (2007). Additionally, it draws from small area estimation (SAE) methods. SAE methods combine raw or individual data from multiple data sources to estimate small areas or groups (Lohr & Prasad 2003), and often use model-based analysis (Rao, 2003). These researchers suggest combining the modeling approach frequently adopted in SAE with population information used in post-stratification to estimate the mean of a binary variable that is conditional on demographic and geographic predictors in cross-survey analysis.

This approach comprises two steps: first, fitting a Bayesian multilevel logistic regression model to individual responses using demographic and geographical information; and second, estimating the population using census demographics.

Kastellec, Lax, and Phillips (2010) summarize this method as follows: "[It] begins by using multilevel regression to model individual survey responses as a function of demographic and geographic predictors, partially pooling respondents across states to an extent determined by the data. The final step is poststratification, in which the estimates for each demographic-geographic respondent type are weighted (poststratified) by the percentages of each type in the actual state populations" (p. 1). Next is a brief description of the two steps involved in this method, adapted primarily from the work of Gelman and Hill (2007).

In the first step, a Bayesian multilevel logistic regression model is fitted to the individual responses y given demographic characteristics and geographical location. Hierarchical modeling is used to take account of clustering of respondents within surveys and geographical locations. The model estimates the average response θ_l for each cross-classification l of demographics and geographical location. Assuming a binary yes/no survey response y for each respondent i, the probability of $y_{i=1}$ can be modeled using a logistic regression. What follows is an example drawn from Gelman and Hill (2007) to describe this step. The authors developed a Bayesian multilevel logistic regression model to estimate the probability that a respondent will prefer the Republican candidate for president.

To do so, they worked with data from seven news polls conducted in 1988. Defining y_i as 1 for supporters of the Republican candidate and 0 for supporters of the Democratic candidate, the authors modeled the probability of $y_i = 1$ as a function of respondents' demographics (gender, race, age, and education) and state. In the following example, the model includes individual predictors with fixed intercepts (gender and race), individual predictors with varying intercepts (age and education), and a state level predictor.

$Pr\ (y_i = 1) = logit^{-1}(\beta^0 + \beta^{female} \cdot female_i + \beta^{black} \cdot black_i + \alpha^{age}_{k[i]} + \alpha^{edu}_{l[m]} + \alpha^{state}_{j[i]} + \alpha^{poll}_{p[i]})$. For i=1,…,n

The varying coefficients are assigned normal distributions with a mean of 0 and standard deviations estimated from data, given non-informative uniform priors:

$$\alpha^{age}_k \sim N(0, \sigma^2_{age}), for\ k = 1, \ldots, 4$$

$$\alpha^{edu}_l \sim N(0, \sigma^2_{edu}), for\ k = 1, \ldots, 4$$

The state-level predictor is modeled as a function of the region (the authors include indicators for five regions) and a measure of previous Republican votes in the state.

$$\alpha^{state}_j \sim N(\alpha^{region}_m + \beta^{v.prev} \cdot prev_j, \sigma^2_{state})$$

$$\alpha^{region}_m \sim N(0, \sigma^2_{region}), for\ m = 1, \ldots, 5$$

In the second step population estimates are calculated with the help of logistic regression coefficients and census demographics. The model estimates an average response $\hat{\theta}_l$ for each demographic vector and geographical cross-classification (l). Adult population counts (N_l) for each demographic and geographical cross-classification (l) can be obtained from the census. The estimated population average for a geographical location $\hat{\theta}_g$, for example, can be calculated thus:

$$\hat{\theta}_g = \sum_{l \in g} N_l \hat{\theta}_l / \sum_{l \in g} N_l$$

As previously noted, the method has a number of advantages derived from its mixed methodology. Given that the approach is model-based, sampling and survey design variables are included in the analysis so that their relationship to the population estimates can be examined and explained directly (Tighe et al., 2010). The multilevel aspect of this approach makes it possible to model the clustering of respondents within surveys and within geographical locations. The Bayesian inference feature enables the adjustment of many factors, which facilitates the inclusion of more information in the inferential procedure (Gelman, 2014). Post-stratification allows us to correct for non-response bias (Park et al., 2004). The limitations of the model, reflected in the literature review, lie in the complexity of model building and the risk of misspecifications (Gelman, 2014).

Next is the description of some studies that employed the BMRP approach. Shirley and Gelman (2012) applied it to estimate state and demographic trends in U.S. public opinion on the death penalty. The authors used data from General Social Survey and Gallop polls conducted between 1974 and 2000 that asked a question of the type "Do you favor or oppose the death penalty for persons convicted of murder?" With the BMRP approach, they were able to estimate trends by geographical location, gender, and race.

Park et al. (2004) validated the BMRP method by applying it to pre-election poll data and compared the estimates with state election outcomes. The researchers concluded that the model outperformed other models commonly used in political science. Gelman (2014), in turn, presents the results of a project that succeeded for the first time in building a model that connects income and voting patterns (red vs. blue states). The model was fitted to pre-election polls from 2000, 2004, and 2008 with about twenty to forty thousand respondents per year. The multilevel aspect of the model allowed for different patterns of income and voting in different states.

For the last ten years, researchers at Brandeis University's Steinhardt Social Research Institute have been studying methods to estimate the size and demographic characteristics of the adult Jewish population in the US (Tighe et al., 2010; Tighe et al., 2012). They apply the BMRP approach to model the likelihood that a survey respondent will identify as Jewish and to examine the geographic and demographic distribution of the Jewish population in the US. The model draws on data from repeated independent samples of US adults who self-identify as Jewish (Tighe, Saxe, Kramer & Parmer, 2013). Hierarchical Bayesian analysis accounts for the clustering of respondents within surveys. In this way, researchers are able both to estimate the population and to account for the different sampling variances across data sources. When using this approach, weighting factors associated with the assessment of the Jewish population (such as geographical location and variation by age, education, and the interaction of the two) are estimated directly (Tighe et al., 2013).

Comparison of Cross-Survey Methods.

In this section I present the results of a study that compared the cross-survey methods that are assessed in this dissertation.

Figueiredo and Campos (2013) conducted a comparative study of three statistical approaches to combining survey data. The three approaches, applied to the estimation of the party share of the vote in the Portuguese legislative elections, were (a) meta-analysis type approach (the authors refer to it as poll aggregation), (b) Bayesian multilevel regression with post-stratification (BMRP) (the authors refer to it as multilevel regression method), and (c) small area estimation method (EBLUP). The third method is not relevant to this study, for in addition to combining data from surveys, it uses administrative data. They used the MAE (mean absolute error) and the CV (coefficient of variation) indicators to compare the methods. Results indicate that estimates obtained with the BMRP method were closer to exit poll results than estimates generated with the meta-analysis method. In addition, estimates generated with the BMRP approach were of better quality (lower CV) than the ones produced by the other two methods. Judging from the published article, it is not clear whether the authors took into consideration the weighting schema of each study or just the sample size.

Summary

This section of the literature review introduced the cross-survey methodology. First it provided an overview of this methodology, including general theoretical purposes and considerations for implementation. Then it examined the three cross-survey methods used in this dissertation. The review of studies that apply cross-survey methodology revealed a host of different approaches, especially in terms of the treatment of survey variances. As previously noted, it was not easy to find research that applied the meta-analysis methodology described by Fox (2011). A large number of studies that use meta-analysis approaches fail to consider the sampling design of the different surveys. In regard to the use of pooled design-based cross-survey methods modeled on Khorn and Graubard (1999), a significant number of studies were found in the health sciences. Many of these combine different cycles of the same study. The review of the literature showed various model-based approaches to combining data from different surveys that apply hierarchical modeling. Numerous examples were found of the use of Bayesian multilevel estimation with post-stratification, especially in the social sciences.

This dissertation adds to the small body of research that applies the meta-analysis and pooled design-based cross-survey methodologies to estimate populations. In addition, it contributes to the comparison of cross-survey methodologies.

Jewish Population in the US and Canada

Demographic data on the Jewish population is important for understanding contemporary Jewish life and for effective policy planning, and has implications on the global and local levels (DellaPergola, 2011; Rebhun, 2014; Tighe, Saxe & Livert, 2010). This section begins with a brief description of who is considered a Jew both in the context of the present study and in relation to broader sociodemographic definitions. Subsequently, it briefly discusses sociodemographic research on the Jewish population in the US and Canada. The intent is to illustrate the importance the Jewish community places on population studies. As Heilman (2103) points out,

> Jews have been counting themselves since the exodus from Egypt, and just as long there have been questions about the accuracy and completeness of those counts. In that sense, not a lot has changed in the long course of Jewish history. But we have always understood that while the question of what constitutes Jewish identity and activity is essential and critical, the question of how many there are for whom it is such, is no less important. (p 1).

Jewish by Religion

Central to demographic studies of the Jewish population is the definition of “who is a Jew.” Normative and operational definitions used to describe who is Jewish lack both coherence and uniformity (DellaPergola, 2014). As Saxe et al. (2013) argue, “who is a Jew” is a social classification that depends on diverse factors including Jewish law (*halacha*), religious identity, ethnicity, culture, language, and/or descent.

Kosmin et al. (1991) introduced the concept of core Jewish population. This Jewish population includes (a) respondents who identify themselves as Jews when asked about their religion; and (b) individuals who consider themselves to be Jewish (who are not interested in religion but see themselves as Jews by ethnicity or by other cultural criteria), have a Jewish background, and are not affiliated with other organized religions (Cooperman, Smith, Hackett, & Kuriakose, 2013; DellaPergola, 2014). The sector of the Jewish population that self-identifies as Jewish when asked about religious affiliation is frequently referred to as “Jewish by religion.” This is the sector of the Jewish population that was estimated in this study by means of cross-survey methods. The proportion of the Jewish by religion population within the core Jewish population is estimated to be about 78% in the US (Cooperman et al., 2013) and about 84% in Canada (Shahar, 2014).

The format of self-identifying religion questions might differ across surveys. Most of these questions in US surveys ask: “What is your religious preference? Is it Protestant, Catholic, Jewish, or other?” or “What is your religious preference, if any?” The format might influence how a respondent interprets the question. Previous multilevel research that modeled the likelihood of a respondent being Jewish in the US found very small survey variance (e.g., ICC of .004) after level one variables were included in the model, which led to the conclusion that the wording of the question was not a significant predictor (Tighe et al., 2010; Tighe et al., 2011).

Sociodemographic Research on the Jewish Population in Canada

Canada has included religion-related questions in its census since 1901 (Koffman & Weinfeld, 2011). For many decades the Canadian census has served as an important tool for researchers of the Canadian Jewish population. It provides an opportunity to obtain a snapshot of the demographics of this population as well as to study demographic trends over time (Shahar, 2003). The Canadian census has also been used to detect additional features of the Jewish population such as intermarriage rate and enrollment of children in Jewish day schools, and even to identify at-risk segments of this population (Shahar, 2013).

The Canadian census includes a religion question every ten years and an ethnic ancestry question every five years (Statistics Canada, 2005; Weinfeld, Schnoor & Koffman, 2013). It has counted Jews as both a religious and an ethnic group for many decades (Weinfeld, 2013). Starting in 1971, the Jewish Federations of Canada adopted what is known as the "Jewish standard definition" to estimate the size of the Jewish population. This definition derives from a combination of responses to both the religion and the ancestry question from the census. It establishes that a Jew is someone who identifies as Jewish either by religion or by ethnic ancestry, provided he or she does not identify with any other religious affiliation (Shahar, 2004; Weinfeld et al., 2013).

Based only on the religion question, the Jewish population for 2011 was estimated at 329,495. In 2011 a revised definition was adopted that includes additional census questions such as place of birth, five-year mobility, and knowledge of non-official languages (Shahar, 2014; Weinfeld et al., 2013). Based on the revised definition, the

Jewish population of Canada in 2011 was 391,665, representing approximately 1.2% of the total Canadian population (Shahar, 2014).

Prior to 2011 the religion and ethnicity questions were part of the mandatory long-form questionnaire, which was administered to 20% of the population. In 2011 the long form became the voluntary National Household Survey (NHS). The NHS was conducted on the same day as the 2011 census and distributed to a third of the households in Canada. It had a 68% response rate across the country (Weinfeld, 2015). Concerned about accurate and maximal census counts, the Jewish Federation urged Jews to answer the 2011 NHS (Weinfeld, 2015). It is not clear how the voluntary nature of the survey and the community's efforts to have a high Jewish participation rate might have affected the non-response bias (Shahar, 2014). For the purpose of this study, the estimation of the Jewish population from the Canadian census was calculated using only the religion question.

Sociodemographic Research on the Jewish Population in the US

The Jewish community is strongly interested in determining and understanding its population size and demographic distribution (Tighe et al., 2010). To this end, in the last decades the Jewish community in the US has invested more resources in sociodemographic studies than in any other type of systematic research (Saxe, Tighe & Boxer, 2014). In the absence of a census, researchers have used two main strategies to produce estimates of this population, namely, national population studies and a systematic combination of local population estimates (Saxe et al., 2014).

In 1957 the U.S. Census Bureau carried out a Current Population Survey (CPS) that included information on religious affiliation. This survey was highly reliable, but it was administered only once. The central Jewish federation, in turn, conducted National Jewish Population Studies (NJPS) on the size and characteristics of the U.S. Jewish population in 1970, 1990, and 2000 (DellaPergola, 2013). The 1990 survey is highly regarded. It included a random digit dialed (RDD) sample with oversamples of low-density Jewish population areas. It screened for Jews in the framework of an omnibus survey over the course of a year (DellaPergola, 2005; Kadushin, Phillips & Saxe, 2005). The 2000 survey, by contrast, has been the focus of substantial controversy and misinterpretation (Kadushin et al., 2005).

In addition to the NJPS, Mayer, Egon, Kosmin, and Keysar (2001) conducted an American Jewish Identification Survey (AJIS) with private funding in 2001 (DellaPergola, 2013) and in 2008 (Kosmin,2009). One of their purposes was to replicate the methodology of the 1990 survey. Furthermore, in 2013 the Pew Research Center's Religion & Public Life Project carried out a national representative survey of the Jewish population. Researchers performed more than 70,000 screening interviews to identify Jewish respondents in the US. Longer interviews were completed with 3,475 Jews (Cooperman et al., 2013).

National surveys of the Jewish population are very expensive and difficult to conduct. The main problem is that Jews only constitute approximately 2% of the total population of the US, which makes the use of random digit dialing (RDD) surveys extremely expensive (Saxe et al., 2014). As an example, in an effort to reduce costs and increase the size of the sample, Pew survey researchers did not carry out interviews in those areas of the country where previous studies indicated that there were very few "Jews by religion." The survey covered geographical areas that are home to roughly 90%

of the US adult population, which means that the results were less than ideal for calculating population estimates (Cooperman, 2013). Additional limitations of a national survey are associated with low response rates and the potential bias introduced by sampling error and weighting (Saxe et al., 2014).

Due to all these disadvantages, researchers at Brandeis University's Steinhardt Social Research Institute (SSRI) have resorted to a different approach based on cross-survey methods. This approach combines individual-level data across studies through statistical techniques that take into account differing designs of the studies. It is thus possible to obtain overall estimates as well as distributions by age, sex, education, and geographical area (Saxe et al., 2013; Tighe et al., 2012; Tighe et al., 2013).

A fourth strategy follows the major tenets of demography. DellaPergola (2013) examines the demographic trends of the Jewish population in the US. He compares and reassesses the various national Jewish population estimates from a demographic perspective, looking at international migration, fertility, and conversions trends.

An ongoing debate exists among scholars and researchers regarding the size and features of the Jewish population in the US. As Saxe and DellaPergola (2013) assert,

> Assessing the size and characteristics of the Jewish population in the United States is probably not the central question that needs to be addressed by American Jewry, but is surely one of the most intriguing, debated, and at times antagonizing tasks—not only in demographic studies but more generally in the social scientific study of Jewry (e.g., Contemporary Jewry 2005). Competing narrative and empirical approaches have generated diverging estimates, with a significant high-low gap of about one million, and opposite interpretations of current and expected

trends, ranging between rapid growth and slow decline (see DellaPergola 2013; Saxe and Tighe 2013). (DellaPergola, Saxe, and Tighe, 2013, p.4)

Relevance of Jewish Population Sociodemographic Research

Demographic data on the Jewish population have played an important role in both scholarly analyses and Jewish communal planning (DellaPergola, 2011; Rebhun, 2014; Saxe & Tighe, 2013). Scholars have been using sociodemographic research not only to study the demographic patterns of the Jewish community, but also to explore a wide variety of economic and social issues concerning the Jewish population. They have relied on sociodemogaphic information to study the economic status of the Jewish community (Wilder, 2014) and the effects of the economic crisis on Jews (Kotler-Berkowitz, 2014), examine the relationship between community features and Jewish identity (Hartman, & Sheskin, 2011), and look at attitudes toward and attachment to Israel (Sasson, Kadushin, & Saxe, 2010), to name a few.

Sociodemographic studies of the Jewish population in the US have played and continue to play a central role in the discourse on the continuity of the Jewish community and in long-term policy planning. For instance, "the findings of the 1990 National Jewish Population survey were a catalyst for academic and communal debate about intermarriage, the role of religion/ethnicity, and Jewish continuity (Goldstein 1992; Mayer, 1991)" (Saxe & Tighe, 2013, p. 44). A more recent example is the extended debate sparked by the 2013 Pew survey of Jewish Americans (Cooperman et al., 2013) on the implications of its findings, and thus on long-term planning (e.g., Cohen, 2015; Kosmin, 2015; Saxe, Sasson, & Aronson, 2015).

Jewish population estimates have also been used to adjust the samples of target studies on Jewish populations through post-stratification weighting. For example, researchers have used population estimates from the 2000–01 NJPS, or, more recently, from the Pew Religious Landscape Survey and the General Social Survey as frames to which to weight target studies (Boxer, Aronson & Saxe, 2013). Furthermore, Jewish population estimates from cross-survey analysis conducted at the SSRI have helped generate post-stratification weights for community studies (e.g., Boxer, Aronson, Brown & Saxe, 2015) and for surveys based on online panels (e.g., Boxer et al., 2013).

Summary

The first purpose of this section of the literature review was to provide a brief description of the operational definition of Jew in the context of the present study. The second was to illustrate the importance of sociodemographic research on the Jewish population in the US and Canada. The third was to describe current methodologies for estimating the Jewish population in these countries. Lastly, this section aimed to convey the complex and competing narratives of sociodemographic research in the US.

Chapter 3: Methodology

The goal of this study was to assess the feasibility of using cross-survey methods to estimate the proportions of low-incidence populations. This chapter discusses the overall research design and methodology; describes the samples and data preparation; explains the procedures used to obtain population estimates through cross-survey analysis; and details the statistical procedures followed to evaluate the cross-survey methods and answer the research questions. The study consisted of three stages: data preparation, population estimation using cross-survey methods, and evaluation of these methods.

Overall Research Design and Methods

Overall Research Design

This dissertation aimed to assess the operational characteristics of three cross-survey methods and to ascertain their suitability to estimate the proportions of low-incidence populations. Specifically, I evaluated the use of these methods to estimate the proportions of the adult population that self-identifies as Jewish in the largest Canadian and American metropolitan areas and in the Canadian provinces expected to have the largest proportions of Jewish adults based on the Canadian census. For this purpose, evaluation and cross-validation studies were performed with data from nationally representative probability samples of the adult population in Canada and the US. Three main sets of such surveys were included: Canadian data from 1997-2004, Canadian data from 2006-2014, and U.S. data from 2008-2014. The 1997 to 2004 time period includes surveys conducted before and after the 2001 Canadian census, which, when combined, could be used to provide a population estimate that could be compared to the 2001

Canadian census benchmark. Multiple years of data are required because there are too few surveys conducted in a single year. Similarly, data are combined from 2006 to 2014 for comparison to mid-year of 2011 Canadian Census.

The three cross-survey methods evaluated in this study are meta-analysis of complex surveys (MACS), pooled design-based cross-survey (PDCS), and Bayesian multilevel regression with post-stratification (BMRP). Chapter 2 discussed the theories underlying these methods as well as examples of their application. The Cross-Survey Analysis section below provides information concerning the estimation procedures carried out based on these methods.

Estimates of the proportions of the adult Jewish population by province and metropolitan area were generated for each of the batches in the study using each of the three cross-survey methods. These estimates, in turn, were compared and evaluated to answer the research questions.

To evaluate the accuracy of the methods, it would be optimal to compare the results to external benchmarks. In the United States census data do not include religious identification, so there are no external benchmarks with which results can be compared. In Canada, in contrast, religious identification is available through the census. Prior to 2011 the Canadian census included a question about religious identification every ten years as part of the mandatory long-form questionnaire, which was administered to 20% of the population. In 2011 the government introduced the National Household Survey (NHS) to replace sections of this questionnaire. The NHS was conducted on the same day as the 2011 census, but unlike the latter, it was voluntary.[1] [See Figure 3.1 for the

[1]The long-form questionnaire will be reinstated for the 2016 census.

religious identification question used in the 2011 NHS (Statistics Canada, 2011).] For the present study, the accuracy of the Canadian sample estimates was assessed using two time periods, namely, 2001, when the religious identification question was part of the long-form questionnaire, and 2011, when this question was voluntary.

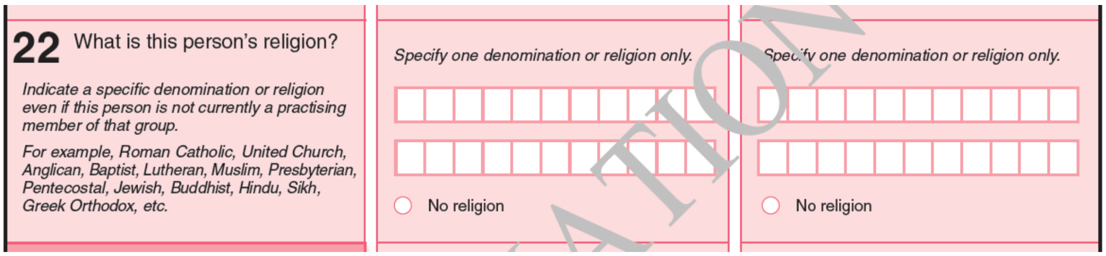

Figure 3.1. Religious identification question, 2011 NHS. Adapted from *2011 National Household questionnaire*, by Statistics Canada, 2011. Retrieved from http://www12.statcan.gc.ca/nhs-enm/2011/ref/about-apropos/ques_guide-eng.cfm

The precision and reliability of the three cross-survey methods was assessed by (a) systematically comparing the estimates generated by the different methods of the proportions of the adult Jewish population in the Canadian provinces with the largest Jewish population and in the largest metropolitan areas in Canada and the US (see the Data Analysis section below) and (b) cross-validating the estimates generated from random subsamples of the U.S. sample (see the Sample Description and Data Analysis sections below).

Target Population

The target population of the study is the adult population that self-identifies as Jewish living in (a) the four provinces with the largest Jewish population, (b) the three largest Canadian metropolitan areas, and (c) the 20 largest metropolitan areas in the US. As noted in the previous chapter, defining who is Jewish is a complicated process. In this study I am only considering adults (aged 25 years or older in Canada and aged 18 years or older in the US) who self-identified as Jewish when asked about their religion. For the sake of simplicity, Jewish adults are being defined as those who self-identify as Jewish when asked about their religion.

Dependent variable

The point estimate ($\widehat{\Theta}$) generated in this study with each of the three methods is the proportion of the adult Jewish population within a specific geographical location, for example, the proportion of the adult Jewish population in the Toronto metropolitan area. Proportions can be estimated either by dividing the counts of the adult Jewish population by the total adult population for the specific geographical location or by estimating the mean of a dichotomous variable in which Jewish adults are coded as 1 and other adults, as 0.

As will be described subsequently, the proportion of the adult Jewish population was estimated by metropolitan area in the US, by metropolitan area in Canada, and by province in Canada.

Geographical variables

In the US, metropolitan areas are defined as core urban areas with a population of 50,000 or more inhabitants. Metropolitan areas are uniquely defined by a five-digit number called Core-Based Statistical Area (CBSA) code (U.S. Census Bureau, 2015). Table 3.1 shows the 20 U.S. metropolitan areas included in the study. For each of these areas, the table displays the total adult population. The total adult population is based on census estimates for 2014.

Table 3.1

U.S. Largest Metropolitan Areas

CBSA	Name	Total Adults
35620	New York-Newark-Jersey City, NY-NJ-PA	15,650,031
31080	Los Angeles-Long Beach-Anaheim, CA	10,223,981
16980	Chicago-Naperville-Elgin, IL-IN-WI	6,616,152
19100	Dallas-Fort Worth-Arlington, TX	5,094,685
26420	Houston-The Woodlands-Sugar Land, TX	4,738,824
47900	Washington-Arlington-Alexandria, DC-VA-MD-WV	4,576,052
37980	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	4,187,805
33100	Miami-Fort Lauderdale-West Palm Beach, FL	4,705,592
12060	Atlanta-Sandy Springs-Roswell, GA	4,183,807
14460	Boston-Cambridge-Newton, MA-NH	3,411,280
41860	San Francisco-Oakland-Hayward, CA	3,655,838
38060	Phoenix-Mesa-Scottsdale, AZ	3,370,429
40140	Riverside-San Bernardino-Ontario, CA	3,250,224
19820	Detroit-Warren-Dearborn, MI	3,310,685
42660	Seattle-Tacoma-Bellevue, WA	2,861,543
33460	Minneapolis-St. Paul-Bloomington, MN-WI	2,649,105
41740	San Diego-Carlsbad, CA	2,531,831
45300	Tampa-St. Petersburg-Clearwater, FL	2,319,250
41180	St. Louis, MO-IL	2,167,262
19740	Denver-Aurora-Lakewood, CO	2,099,066

Canada is divided into ten provinces and three territories. This study only considers the population in the four provinces with the largest Jewish population: Quebec, Ontario, Manitoba, and British Columbia. The Canadian census metropolitan areas (CMAs) are composed of one or more adjacent municipalities situated around a population center (known as the core). These areas must include, at a minimum, 100,000 residents, of which 50,000 or more must live in the core (Statistics Canada, 2015). The study focuses on the three largest metropolitan areas, namely, Toronto, Montreal, and Vancouver.

Table 3.2 shows the provinces and metropolitan areas included in this study. For each geographical area, the table displays the total adult population and the adult Jewish population based on estimates from the 2001 census and the 2011 NHS public use microdata files (Statistics Canada, 2001; Statistics Canada, 2011). Counts represent adults aged 25 and older.

Table 3.2

Canadian Provinces and Metropolitan Areas included in the Study

Geographical area	Total Adult Population 2001	Adult Jewish Population 2001	Total Adult Population 2011	Adult Jewish Population 2011
Province				
Quebec	4,884,949	63,289	5,522,405	56,613
Ontario	7,573,083	130,037	8,758,569	127,927
Manitoba	718,771	9,516	788,157	8,211
British Columbia	2,651,638	16,101	3,106,909	14,794
Metropolitan Area				
Montreal, Quebec	2,317,112	62,179	2,637,381	56,117
Toronto, Ontario	3,122,378	110,806	3,796,746	113,028
Vancouver, British Columbia	1,358,010	13,294	1,628,592	12,836

Sample description

This study includes three sets of survey data. Each comprises existing survey samples for which raw data are available. I will further refer to the three sets as *batches* to distinguish between them and the survey samples that comprise the data. Next is an outline of these batches (a more detailed description is provided in the Data Preparation section, and in Chapter 4 as well).

2001 Canada batch (Ca2001): The Ca2001 batch comprises data from surveys of nationally representative random samples of the adult population in Canada. It includes 45 national surveys fielded between 1997 and 2004.

2011 Canada batch (Ca2011): The Ca2011 batch contains data from surveys of nationally representative random samples of the adult population in Canada. It includes 17 national surveys administered between 2006 and 2014.

2011-US batch (US2011): The US2011 batch consists of data from surveys of nationally representative random samples of the adult population in the continental US conducted between 2008 and 2014.

In addition to these batches, the following subsamples of the main US batch were generated:

US2011a-US batch (US2011a): A stratified randomly selected subsample of survey samples that represents half of the surveys in US2011 batch; and

UI2011b-US batch (US2011b): A stratified randomly selected subsample that represents the other half of the survey samples in US2011 (not included in US2011a).

US2011a and US2011b have been selected so that the resulting files are equivalent in terms of sample size, weighted proportion of Jewish population, and survey quality. To achieve matching samples, I stratified the US2011 batch based on sample size (three categories), weighted Jewish population (three categories), and survey sponsor. Surveys were then randomly assigned to US2011a and US2011b based on the stratification.

Data Preparation

This section includes information about the first stage of the study, that is, data preparation. It describes the preparation of the survey samples datasets and of the additional datasets that were needed to conduct the study. As previously described, several collections of datasets were needed: (a) US batches, US2011, US2011a and US2011b, (b) Canada batches Ca2001 and Ca2011. In addition, several auxiliary datasets were needed for analysis and benchmarking. These include: (a) U.S. and

Canadian census and NHS datasets, and (b) 2001 Canadian census and 2011 NHS Jewish population counts datasets.

Survey Samples Searches

Searches for the surveys for the Canadian batches were conducted in major Canadian data repositories, both governmental and private. Surveys were primarily identified through ODESI (Ontario Data Documentation, Extraction Service and Infrastructure). ODESI is a digital repository for the social sciences that contains information about an extensive number of collections, such as Statistics Canada's public-use survey data. In addition to ODESI, surveys were identified through the Canadian Opinion Research Archive (CORA), Canadian Gallup, Ipsos Reid, the Interuniversity Consortium for Political and Social Research (ICPSR), the Canadian Policy Research Network, and the Institute for Social Research (ISR).

Surveys for the U.S. batch were drawn from a dataset collected by researchers at Brandeis University's Steinhardt Social Research Institute (SSRI) (Tighe et al., 2013). Over the past ten years, SSRI researchers have stored more than 400 independent samples of surveys of the adult population in the US. This dataset includes surveys identified in major data repositories, such as ICPSR and the American Religion Data Archive (ARDA), as well as in poll archives at the Roper Center, Gallup, and Pew Research Center (Tighe et al., 2012; Tighe et al., 2013). Chapter 4 includes a detailed description of the batches.

As stated in Chapter 2, a number of features must be taken into consideration when combining data from different surveys. The choice of surveys for the samples was guided by this information. Surveys had to meet the following conditions:

- include the target population;
- refer to the correct time period;
- contain data that can be statistically combined;
- include nationally representative samples of the adult population in Canada or the US;
- offer baseline demographical information for each respondent: gender, race, educational attainment, and age;
- provide a measure of respondent's current religious identification;
- contain sufficient geographical information to identify respondent's metropolitan area; and
- offer sufficient (sampling) weight information to conduct the cross-survey analyses.

U.S. Dataset and Combined Canada Batches Dataset

The two main datasets for this study are composed of all the individual data from the surveys in the Canadian and U.S. batches, respectively. Each record in a dataset represents a respondent in a survey. The record for each respondent includes both survey-related and respondent-related information. Survey-related information consists of survey identification, sampling method, sample size, response rate, and year the survey was conducted. Respondent-related variables are religious identification, demographic and geographical information, and survey weights. Each survey and each respondent are uniquely identified.

Demographic data recoded for each respondent and included in the U.S. dataset consists of the following variables: gender, age, educational attainment, and race. For each respondent in the Canadian dataset, the demographic variables include gender, age, educational attainment, and income. The geographical information recoded for each respondent in the U.S. dataset consists of state, county, and metropolitan area. In the Canadian dataset, geographical variables consist of province and metropolitan area.

Sampling-related variables were also recoded for each respondent. This information includes the total weight for each record drawn from the survey, as well as any other available information regarding the provided weight.

Variables in the datasets are standardized, in the sense that they represent the same measure across surveys. Religious identification is coded 1 for Jewish and 0 for non-Jewish. Provinces and metropolitan areas in Canada and metropolitan areas in the US are coded according to their province or state, CMA, and CBSA codes, respectively. For demographics, in the U.S. dataset, age is coded for all respondents according to six categories: 1 for ages 18-24, 2 for ages 25-34, 3 for ages 35-44, 4 for ages 45-54, 5 for ages 55-64, and 6 for ages 65+. Education is coded into two levels: non-college-educated, coded 0, and college-educated, coded as 1. Race/ethnicity is coded into four categories: non-Hispanic White is coded 1, non-Hispanic Black is coded as, 2, Hispanic as, 3, and all others, 4. Gender is coded 0 for males and 1 for females. In the Canadian datasets, in turn, age is coded for all respondents according to five categories: 1 for ages 25-34, 2 for ages 35-44, 3 for ages 45-54, 4 for ages 55-64, and 5 for ages 65+. Education is coded into five levels: no certificate is coded 1, high school certificate is coded, 2, trade or college certificate is, 3, university certificate is, 4, and post-college certificate, is 5. Income is coded into four categories: 1 for annual incomes between 0 and $30,000, 2 for annual incomes between $30,000 and just below $60,000, 3 for annual incomes between $60,000 and just below $80,000, and 4 for annual incomes of $80,000 or more.

U.S. and Canadian Auxiliary Census Datasets

The U.S. auxiliary census population dataset includes total adult population counts for the year 2014 for metropolitan areas by, age, race/ethnicity, gender and education attainment. These data were adapted from the U.S. Census Current Population Estimates Program (PEP) and the U.S. American Community Survey (ACS). PEP produces yearly estimates of the population of the United States and its states and counties by age, sex, race, and Hispanic origin. Data from the ACS were used for distributions by educational attainment. Table 3.3 provides an example of a record in the dataset. This record indicates that the number of non-Hispanic White college-educated females aged 18-24 in the “Los Angeles-Long Beach-Anaheim, CA” metropolitan area (CBSA 31080) for age is 51,638.

Table 3.3

U.S. Census Population dataset record

CBSA	Age	Gender	Race	College Graduate	Count
31080	1	1	1	1	51,638

The auxiliary 2001 and 2011 Canadian census and NHS datasets includes population counts by demographic and geographical categories. The files contain total adult population counts for the years 2001 and 2011 for province and metropolitan area by age, gender, income, and educational attainment. The data included in the 2001 auxiliary census file were based on analysis of the 2001 Census of Population Public Use

Microdata File (2001 PUMF) produced by Statistics Canada (Statistics Canada, 2001). The data contained in the 2011 auxiliary NHS file were based on analysis of the 2011 National Household Survey Public Use Microdata File (2011 PUMF) obtained also from Statistics Canada (Statistics Canada, 2011). PUMFs files provide access to non-aggregated and anonymous data for a 2.7% sample of the Canadian population (Statistics Canada, 2014).

Table 3.4 offers an example of a record in these Census datasets. This record indicates that in 2001 the number of female, college-graduate adults aged 25-34 with an annual income over $80,000 in the Toronto metropolitan area was 136,938.

Table 3.4

Auxiliary 2001 Canadian Census Dataset Record

Year	Area	Age	Gender	Education	Income	Count
2001	Toronto	1 (25-34)	1 (Female)	4 (University Certificate)	3(over 80K)	136,938

2001 Canadian Census and 2011 NHS Adult Jewish Population Counts

As described below, census counts of the adult Jewish population by geographical location (province and metropolitan area) were needed to conduct the analyses. The 2001 Canadian census and 2011 NHS adult Jewish population counts datasets include this information for the years 2001 and 2011. Information was adapted from the 2001 Census of Population Public Use Microdata File (Statistics Canada, 2001) and 2011 National Household Public Use Microdata File (Statistics Canada, 2011) provided by Statistics Canada.

Table 3.5 offers an example of a record in the 2011 NHS adult Jewish population dataset. The example indicates that the female Jewish population aged 25-34 with a university certificate in the Toronto, Ontario metropolitan area (CMA 535) was 6,104 for the year 2001.

Table 3.5

2001 Census and 2011NHS Adult Jewish Population Counts Record

Area	Age	Gender	Education	Count
Toronto	1 (25-34)	1 (Female)	4 (University Certificate)	6,104

Population Estimation Using Cross-Survey Analysis

In this section I describe the types of population estimates that were generated with the three cross-survey methods, and the procedures followed to generate these estimates.

Description of Population Estimates

For each batch, I generated a series of population estimates using the three methods of cross-survey analysis. As stated previously, the population estimates are represented as proportions of the overall population. Below is a description of the population estimates generated in this study with each of the methods:

- proportions of the adult population that is Jewish by geographical location in the US. I provide estimates for the 20 largest metropolitan areas in the US for batch US2011, and for batches US2011a and US2011b;
- proportions of the total adult population that is Jewish in the four Canadian provinces with the largest proportions of Jewish population for batches Ca2001 and Ca2011; and
- proportions of the total adult population that is Jewish in the three largest Canadian metropolitan areas for batches Ca2001 and Ca2011.

Procedures to Generate Population Estimates for each Cross-Survey Method

What follows is a description of the steps involved in the application of each method to generate the population estimates discussed above.

Meta-Analysis of Complex Surveys (MACS). The first step in generating population estimates using the MACS method consists of generating the population estimates described in the previous section, $\widehat{\theta}_j$, and related variances, $v\ (\widehat{\theta}_j)$, for each survey *j* using the weights provided by the survey.

The second step requires computing relative weights w_j for each survey j. In the context of the MACS approach, weights are used to attach greater significance to the more precise estimates when pooling. Relative weights are a function of both the variance (v_j) and the design effect ($deff_j$) of the survey. The design effect adjusts for excess variability if there is clustering in the survey:

$$w_j = \frac{1}{\frac{deff_j}{v_j}}$$

The third step involves generating population estimates $\hat{\theta}_N$ and related variances across surveys for each batch. The estimate of the population for a given batch is the weighted average of the estimates across all survey samples in the batch:

$$\hat{\theta}_N = \frac{\sum_{j=1}^{k} w_j\, \hat{\theta}_j}{\sum_{j=1}^{k} w_j}$$

The variance of the population estimate $v(\hat{\theta})$ is calculated as the weighted average of the variances of the different surveys in the batch:

$$v(\hat{\theta}) = 1 \Big/ \sum_{j=1}^{k} w_j$$

Pooled Design-Based Cross-Survey (PDCS). The first step in generating population estimates using the pooled design-based cross-survey approach involves calculating a new relative weight (w_{ij}^{*}) for each respondent *i* in each survey *j*. This step is needed to take into account survey-specific sample designs and sample sizes.

$$w_{ij}^{*} = w_{ij} * k_j$$

where *k*j is defined as:

$$k_j = \frac{\sum_{i=1}^{n_1} w_{i1} + \cdots + \sum_{i=1}^{n_L} w_{iL}}{L\sum_{i=1}^{n_j} w_{ij}}(1 - \frac{(CV_j^2 + 1)/n_j}{(CV_1^2 + 1)/n_1\ + \cdots +\ (CV_L^2 + 1)/n_L})$$

where CV_j is the coefficient of variation of the sample weights in survey sample *j*.

The relative weight for each respondent is calculated independently for each batch. For instance, the relative weight for the Ca2011 batch is calculated using only the weights of the respondents included in this batch.

The second step consists of obtaining the estimate of the population parameters using the relative weights calculated in the first step:

$$\hat{\theta} = \frac{\sum_{j=1}^{L}\ \sum_{i=1}^{n_j} w^{*}{}_{ij}\, y_{ij}}{\sum_{j=1}^{L}\ \sum_{i=1}^{n_j} w^{*}{}_{ij}}$$

where L is the number of survey samples with $n_1, n_2 \ldots n_L$ sample sizes, $\{y_{ij} \mid i = 1, \ldots, n_j\ , j = 1, \ldots, L\}$.

In the last step, the variance of the target population parameters is calculated using the jackknife replication method. A separate jackknife analysis is conducted for each estimate generated. The jackknife is run on the same data used to generate the estimate. The PSU for each analysis is the individual respondent (case level). For example, the proportion of the adult Jewish population in the San Diego metropolitan area is estimated using data for the San Diego metropolitan area from the US2011 batch. The variance for this estimate is calculated by applying the jackknife method to the same data.

Bayesian Multilevel Regression with Post-Stratification (BMRP). The first stage of the BMRP approach involves building a Bayesian multilevel logistic regression model for each of the batches. The logistic regression model predicts the probability that a person will be identified as Jewish given that person's demographic characteristics (e.g., age, and education), geographical location and survey sample information. “Bayesian inference starts with a prior distribution on the unknown parameters and updates this with the likelihood of the data, yielding a posterior distribution which is used for inferences and predictions” (Gelman & Hill, 2007, p.143).

Bayesian data analysis is an iterative process that involves the following main steps: (a) fitting a Bayesian probability model to the data and specifying the prior distribution, (b) calculating the posterior probability of unknown quantities conditioned on observed quantities, (c) checking the fit of the model, and (d) improving/expanding and extending the model (Gelman, Carlin, Stern & Rubin, 2004; Krushke, 2014). Specifications for the Bayesian probability model include both developing a meaningful descriptive model and establishing prior distributions of its parameters (Krushke, 2014). Prior distributions represent researchers' initial beliefs regarding parameter distribution.

Logistic hierarchical linear modeling (glm) is used to explore the data, decide what initial parameters to include in the model, and obtain values for the priors. Likelihood models run faster than Bayesian models and offer accurate point estimates of low incidence populations. Nevertheless, they provide inaccurate variance estimates. The Bayesian hierarchical logistical model is developed and the prior distributions of the parameters are specified based on the logistic hierarchical linear model analysis. Initially, prior distributions are specified as non-informative. Non-informative prior distributions play a minimal role in determining the posterior distribution, letting the data guide the Bayesian inference (Gelman, Carlin, Stern & Rubin, 2004). Non-informative priors are appropriate for parameters about which little is known beyond the data (Gelman and Hill, 2007).

The next step in Bayesian analysis is assessing the fit of the model, which involves determining whether or not it fits the data and whether or not it has converged. Based on these assessments, the model is adapted and the analysis rerun using semi-informative priors. These steps are repeated until the model properly fits the data and the estimates converge.

Bayesian analysis [when run in R with STAN package] generates random samples from the posterior distribution (Krushchke, 2014). The output of the analysis also contains summary statistics of the coefficients of the posterior distribution including mean estimates, standard errors, and credible intervals. One of the suggested methods to evaluate the quality of the model is comparing the random samples from the posterior distribution under the estimated model with the observed data using summary statistics or graphical analysis (Gelman, et al., 2004; Gill, 2008).

An additional method to assess the quality of the model is testing different variations of the model to the same dataset (Gelman & Hill, 2007). Models can then be compared by (a) observing the implications of the model change for posterior distribution

(Gill, 2008) and (b) examining summary measures of fit such as *Deviance* and *DIC*. The Bayesian analysis software also calculates diagnostic indicators [e.g., potential scale reduction factor ($\hat{r}$) and effective sample size (*neff*)] that can be used to assess whether or not the model estimates have converged.

Once the Bayesian model is finalized, replications of the posterior distribution are used to generate population estimates and credible intervals. The latter are obtained from the variance of the estimates across the replications of posterior distributions.

In the second stage, population counts obtained from the census for each combination of demographics and geographical locations included in the model are used to estimate the corresponding Jewish population parameters. The estimated probability that a person is Jewish for each combination of the parameters in the model can then be calculated by evaluating the regression equation with the coefficients obtained from the Bayesian analysis and the population counts from the census. Summing the results for the different variables across the simulations, in turn, enables the generation of mean estimates of the Jewish population. For example, if the age, education, and race variables are predictors in the model, the estimates for a metropolitan area *cbsa1* are obtained by summing over the predicted values of all categories of race, age, and education within that metropolitan area.

Researchers at the SSRI center have been using the BMRP method to generate estimates of the adult Jewish population in the US by state, metropolitan area, and county groups (Tighe et al., 2010; Tighe et al., 2013). The hierarchical logistic regression model employed to estimate the proportion of respondents identifying as Jewish has respondents nested into surveys and geographical locations. The significant respondent-level predictors were found to be age, gender, education, race, the interaction of age and education, county, and metropolitan indicator (metro, micro or rural area). The process of fitting the Bayesian probability model to the U.S. sample is based on SSRI findings. The process of fitting the logistic hierarchical regression model to generate estimates for the Canadian batches required additional exploration to examine whether factors associated with sampling and the likelihood of identifying as Jewish were the same in Canada as they were in the US. I used data from the 2001 PUMF census file (Statistics Canada, 2001) and 2011 PUMF NHS file (Statistics Canada, 2011) to explore which demographic variables were related to Jewish self-identification.

Data Analysis - Evaluation of the Cross-Survey Methods

This section details the statistical techniques used to assess the operational characteristics of the cross-survey methods and address the specific research questions first presented in Chapter 1. The section begins with a description of the preliminary analyses conducted, continues with a brief presentation of the different indicators used to answer the research questions, and ends with the specific assessments needed to produce these answers

Preliminary Analyses

Prior to conducting the analyses required to answer the research questions, a series of descriptive statistical analyses were performed to understand the sample data and provide additional information to interpret the results of the main analyses. Additional information provided consists of the following: the total sample size as well as the sample size by geographical area for each batch; comparison between the demographic distribution of the samples included in the Ca2001 and Ca2011 batches and the corresponding Canadian census and NHS demographic distribution; an estimate of the weighted and unweighted proportion of the adult Jewish population for each of the samples that compose the three main batches; and the weighted and unweighted proportion of the adult Jewish population for each sample for each of the 20 metropolitan areas in the US (for the U.S. batch) and for the provinces and metropolitan areas included in the study (for the Canadian batches).

Statistical Indicators

A number of indicators were employed to make two types of comparisons. The estimates of the population parameters obtained with the cross-survey methods for batches Ca2001 and Ca2011 were compared with the estimates from the Canadian census and NHS, and the estimates obtained with the different methods were compared with each other. The first set of indicators includes standard errors (SE), confidence intervals (CI), credible intervals (CI_b), and coefficients of variation (CV). These indicators were calculated for each population estimate outlined above. What follows is a brief description of each indicator.

- Standard Error (SE) measures the variability of an estimate due to sampling.
- Confidence Interval (CI) is a measure of the precision of survey estimates. It is an interval, usually centered on the estimate obtained from a sample, which contains the actual corresponding population value with a specified level of probability.

- Credible Interval (CI_b) is the Bayesian analogue of the confidence interval (Gill, 2008) and is calculated from the posterior distributions of the parameters of interest.
- Coefficient of Variation (CV) measures the relative amount of sampling error associated with the sample estimate. It is defined as the ratio of the standard error to the value being estimated (U.S. Census Bureau, 2008). Following the guidelines from the National Center for Education Statistics (NCES) and the American Community Survey (ACS), CVs lower than 15% are considered good, CVs between 15 and 30% are considered, fair, and CVs larger than 30% should be interpreted with caution (National Center for Education Statistics, 2012; Washington State Office of financial Management, 2010). In the Bayesian analysis the CV is calculated from the posterior distribution as the ratio between the standard deviation to the estimate.

In addition, based on the work of Yowell and Devine (2013), I calculated three indicators that aim to identify differences between estimates: mean absolute percent error (MAPE), mean algebraic percent error (MALPE), and root mean squared error (RMSE). In their research into alternative methods to produce county population estimates for 2010, Yowell and Devine used these indicators to identify differences between estimates and census counts. I used them as an additional measure to assess the accuracy of the estimates generated from Ca2001 and Ca2011. They also served to determine the stability of the estimates generated from batches US2011a and US2011b with the three cross-survey methods. Next is a brief explanation of these indicators.

- Mean absolute percent error (MAPE). MAPE is a measure of accuracy (Bryan, 2004). The purpose of this indicator is to provide a relative measure of error. To calculate MAPE, we take the absolute value of the difference between the estimate and the census value for each estimated value, divide that by each respective census value, sum the results, divide this by the number of evaluations included in the analyses (N), and multiply the result by 100:

$$MAPE = \frac{\sum((|Estimate - census\ |)/census))}{N} * 100$$

- Mean algebraic percent error (MALPE). The purpose of this indicator is to identify systematic bias and provide an alternative for a relative measure of error. To obtain it, we take the difference between the estimate and the census value for each evaluation, divide that by each respective census value, sum the results, divide by the number of evaluations N, and multiply the result by 100.

$$MALPE = \frac{\sum((Estimate - census\ )/census))}{N} * 100$$

- Root mean squared error (RMSE). This indicator presents an alternative measure of error that places greater emphasis on large numeric errors than on mean absolute errors. To calculate it, we square the difference between the estimate and the census number for each evaluation, sum these values across evaluations, divide by the number of evaluations, and find the square root of the result.

$$RMSE = \frac{\sqrt{(\sum(estimate - census)^2)}}{N}$$

Research Question Analysis

Below are the data analyses that were conducted to answer the research questions.

Research Question 1: How do cross-survey estimates of the proportions of the total adult Jewish population in Canadian provinces and in metropolitan areas compare to the estimates from the Canadian census and national household survey (NHS)?

Two main sets of analyses were conducted to answer this question. The purpose of these analyses was to determine the accuracy and precision of the population estimates generated by each cross-survey method.

- Assessment of the differences between the estimates of the Jewish population from the Ca2001 batch that were generated by the three cross-survey methods and the corresponding population estimates from the 2001 Canadian census. The accuracy and precision of each method were evaluated separately. In each case I calculated the MAPE, MALPE, and RMSE indicators and examined the confidence intervals of each estimate in relation to the population parameter (taking into account that census estimates are also measured with error).
- Assessment of the differences between the estimates of the Jewish population from the Ca2011 batch that were developed with each cross-survey method, and the corresponding estimates from the 2011 National Household Survey conducted by Statistics Canada.

The accuracy and precision of the estimates of the three cross-survey methods were assessed across four main scenarios represented in Table 3.6 below. For each batch of surveys (Ca2001 and Ca2011) and geographic level (Province and Metropolitan area), estimates derived from each of the cross-survey methods were compared to the corresponding Census benchmark estimates.

Table 3.6

Research question 1-Assessments

	Batch:	Geographical Category	Benchmark
Scenario 1	Ca2001	Provinces	2001 Census
Scenario 2	Ca2001	Metropolitan Areas	2001 Census
Scenario 3	Ca2011	Provinces	2011 NHS
Scenario 4	Ca2011	Metropolitan Areas	2011 NHS

Research Question 2: How do the results of the cross-survey approaches compare in terms of their accuracy and precision in estimating the proportions of the total adult Jewish population in the Canadian provinces and metropolitan areas included in the study?

The purpose of the analyses conducted to answer this question is to assess how the cross-survey methods compare with each other. To this end, I compare the statistical indicators calculated for the population estimates generated by the three methods based on batches Ca2001 and Ca2011. The analyses consist of (a) comparing the MAPE, MAPLE, and RMSE indicators calculated for each batch with each method; (b) comparing the coefficients of variation calculated with the different methods; and (c) examining the confidence and credible intervals of the estimates generated by the three methods.

Research Question 3: How do the three cross-survey approaches compare in terms of their stability and precision for estimating the proportions of the total adult Jewish population in metropolitan areas in the continental US?

Two main analyses were carried out to answer this question. Their purpose was to further assess the stability and precision of the population estimates generated by each cross-survey method.

- A comparison of the population estimates from batch US201I developed with the three methods. Estimates, confidence intervals, credible intervals, and coefficient of variations are compared so as to identify differences in precision among the approaches.
- A comparison of the target population estimates with data from batches US2011a and US2011b that were obtained with the three methods. The purpose of these comparisons is to assess the stability of each cross-survey approach by comparing the results of the two subsamples.

The systematic review of the findings of the three research questions provides indications as to the suitable conditions under which the different cross-survey methods can be used to estimate low-incidence populations. The review also provides information concerning the advantages and limitations of each method.

Summary

This chapter offered a detailed description of the research design used to examine the feasibility of applying cross-survey methods to estimate the proportions of low-incidence religious groups. It included a description of the data, of the application of the three cross-survey methods to generate estimates, and of the methods used to assess the generated estimates in order to answer the research questions.

Chapter 4: Results

The goal of the present study was to assess the feasibility of cross-survey methods to estimate the proportion of low-incidence populations. This chapter details the results of the statistical analyses conducted to determine the operational characteristics of the three cross-survey methods examined in the study. First, it provides a thorough description of the batches and the results of preliminary analyses. Then it presents the estimates generated by each of the methods, along with information on the parameters and processes used to obtain these estimates. Finally, it discusses the results of the analyses in order to answer each research question.

Batch Description and Preliminary Analyses

This section starts with a detailed description of the data contained in the three batches used in the study, namely, 2011 US batch (US2011), 2001 Canada batch (Ca2001), and 2011 Canada batch (Ca2011). Included in the description are weighted and unweighted population distributions by geographical location and main demographics, as well as parallel census-based population distributions. Then it offers the weighted and unweighted distributions of the proportion of the adult Jewish population in the samples making up the three batches.

Descriptions of the 2001 Canada Batch and 2011 Canada Batch

The Ca2001 and Ca2011 batches contain data from representative national surveys of the Canadian population. The Ca2001 batch is composed of data from 45 surveys conducted between 1997 and 2004. The total number of respondents across the surveys is 99,373. The Ca2011 batch comprises data obtained from 68,343 respondents in 17 surveys conducted between 2006 and 2014. (Appendix A1 provides the list of surveys included in each batch. Table 4.1.1 presents descriptive information on the size of the surveys for the two Canadian batches.

Table 4.1.1

Canadian Batches: Sample Size

Batch	Number of Surveys	Number of Respondents	Smallest Sample Size	Largest Sample Size	Mean Sample Size	Median Sample Size
Ca2001	45	99,373	436	33,290	2,208	1,730
Ca2011	17	68,343	672	23,794	4,020	1,424

The surveys were conducted by governmental institutions, research centers, and polling companies. Out of the 45 surveys that form the Ca2001 batch, one was administered by Statistics Canada, ten by research centers, and 34 by polling organizations. As for the 17 surveys making up the Ca2011 batch, two were conducted by Statistics Canada, eight by research institutes or universities, and seven by polling organizations.

Overall, response to the religious identification question was high in both the Ca2001 and the Ca2011 batches. Almost 98% of Ca2001 batch respondents and more than 96% of Ca2011 batch respondents answered a religious identification question.

All surveys in both batches provided a measure of geographical information that allowed for the identification of respondents by province, largest metropolitan area (Toronto, Montreal, and Vancouver), or both. Province of residence was included in all but one of the Ca2001 surveys and in all of the Ca2011 surveys. Identification of the three largest metropolitan areas was possible for 43 of the Ca2001 surveys and for ten of the Ca2011 surveys. Table 4.1.2 shows the number of respondents by province and metropolitan area for each batch (the table comprises only the metropolitan areas and provinces included in the analyses).

Table 4.1.2

Number of Surveys and Respondents by Geographical Area Included in the Analyses

Geographical Area	Ca2001 Number of Surveys	Ca2001 Number of Respondents	Ca2011 Number of Surveys	Ca2011 Number of Respondents
Metropolitan Area				
Montreal	43	9,537	10	1,638
Toronto	43	12,613	10	1,886
Vancouver	43	6,042	10	876
Province				
Quebec	44	16,325	17	13,890
Ontario	44	19,148	17	19,230
Manitoba	44	3,577	17	3,700
British Columbia	44	7,764	17	8,004

The following tables display the census population distribution by gender, age, and educational attainment, as well as the unweighted and weighted samples that compose the two Canadian batches. Weighted distributions of the samples that comprise the two Canadian batches were calculated using survey-specific weights provided by the original researchers. Compared to the general population measured by the census, the sampled population is slightly older and more educated. The distribution of the 2001 batch better resembles the distribution of the Canadian population as estimated by the corresponding census. Table 4.1.3 presents the unweighted and weighted demographic distribution of the samples included in the Ca2001 batch compared with the demographic distribution of the Canadian adult population based on the 2001 census. Table 4.1.4, displays the samples included in the Ca2011 batch compared with 2011 NHS data.

Table 4.1.3

Demographic Distribution Based on the 2001 Census and the Ca2001 Batch

	2001 Census	Ca2001 Unweighted Distribution	Ca2001 Weighted Distribution
Gender			
Male	0.48	0.47	0.48
Female	0.52	0.53	0.52
Education			
No Certificate	0.31	0.20	0.21
High School Certificate	0.21	0.22	0.22
Trade/College Certificate	0.31	0.33	0.33
University Certificate	0.13	0.19	0.18
Post-College Certificate	0.04	0.07	0.07
Age			
25-34 years	0.20	0.21	0.20
35-44 years	0.25	0.26	0.30
45-54 years	0.22	0.22	0.20
55-64 years	0.14	0.15	0.10
65+ years	0.18	0.17	0.20

Table 4.1.4

Demographic Distribution Based on the 2011 NHS and the Ca2011 Batch

	2011 NHS	Ca2011 Unweighted Distribution	Ca2011 Weighted Distribution
Gender			
Male	0.48	0.45	0.45
Female	0.52	0.55	0.55
Education			
No Certificate	0.17	0.14	0.21
High School Certificate	0.23	0.24	0.22
Trade/College Certificate	0.36	0.32	0.33
University Certificate	0.17	0.20	0.18
Post-College Certificate	0.06	0.10	0.07
Age			
25-34 years	0.20	0.13	0.17
35-44 years	0.25	0.18	0.18
45-54 years	0.22	0.21	0.22
55-64 years	0.14	0.22	0.23
65+ years	0.18	0.26	0.22

Description of the US2011 Batch Samples

The US2011 batch includes data from 139 representative national surveys of the continental U.S. adult population that were conducted between 2008 and 2014. The total number of respondents is 215,122. Appendix A2 provides the list of surveys making up this batch. The batch was divided randomly (stratifying by sponsor and sample size) into two sub-batches. The first sub-batch, US2001a, comprises 69 samples with a sample size of 103,831. The second sub-batch, US2001b, is composed of 70 samples with a total sample size of 111,291. Table 4.1.5 presents descriptive information for the entire batch as well as for the two sub-batches.

Table 4.1.5

US Batches: Sample Size

	Number of Surveys	Number of Respondents	Smallest Sample Size	Largest Sample Size	Mean Sample Size	Median Sample Size
US2011a	69	103,831	596	4,966	1,505	1,475
US2011b	70	111,291	476	4,953	1,590	1,482
US2011	139	215,122	476	4,966	1,548	1,475

These surveys were administered by governmental institutions, research centers, and poll companies. Three of them are part of the American National Election Study (ANES) series, two are part of the General Social Surveys (GSS) series, 16 were conducted by ABC News / *The Washington Post*, eight were part of the CBS monthly polls, and the rest were administered by the Pew Research Center. Overall, there was a high response rate for the religious identification question – over 96% for all the surveys included in the batch.

All surveys were drawn from a pool gathered by researchers at the SSRI institute. The SSRI dataset comprises surveys identified in major data repositories such as ICPSR and the American Religion Data Archive (ARDA), as well as in poll archives at the Roper Center, Gallup, and Pew Research Center (Tighe et al., 2012; Tighe et al., 2013). Although it contains more than 400 surveys conducted between 2008 and 2014, only those that provided both survey-specific weights and a measure of geographical information that allowed the identification of respondents by metropolitan area were selected. Table 4.1.6 shows the number of respondents included in the analyses by metropolitan area for each of the two US2011 sub-batches (only the 20 metropolitan areas considered in the current analyses were included).

Table 4.1.6

Number of Surveys and Respondents by Metropolitan Area Included in the Analyses

	US2011a Number of Surveys	US2011a Number of Respondents	US2011b Number of Surveys	US2011b Number of Respondents
metro1	69	5,054	70	5,581
metro2	69	2,737	70	3,306
metro3	69	2,215	70	2,474
metro4	69	1,514	70	1,721
metro5	69	1,385	70	1,465
metro6	69	1,935	70	2,216
metro7	69	1,782	70	1,833
metro8	69	1,195	70	1,525
metro9	69	1,569	70	1,718
metro10	69	1,212	70	1,272
metro11	69	1,024	70	1,296
metro12	69	1,194	70	1,342
metro13	69	1,009	70	1,173
metro14	69	1,256	70	1,284
metro15	69	1,030	70	1,073
metro16	69	1,250	70	1,329
metro17	69	747	70	935
metro18	69	857	70	1,006
metro19	69	1,097	70	1,063
metro20	69	800	70	847

The unweighted demographic distribution of respondents differs from the demographic distribution of the U.S. population as estimated by the 2015 U.S. Census. Compared to the U.S. adult population, US2011 survey respondents are older and more educated. Table 4.1.7 displays the percentage of respondents in certain demographic categories across all surveys compared to the 2014 census.

Table 4.1.7

Demographic Distribution Based on the 2015 Census and the US2011 Batch

	2015 Census	US2011 Unweighted Distribution	US2011 Weighted Distribution
Gender			
Male	0.49	0.47	0.47
Female	0.51	0.53	0.53
Education			
Non-College	0.75	0.62	0.62
College Grad	0.25	0.38	0.38
Age			
18-24 years	0.13	0.09	0.08
25-34 years	0.18	0.12	0.12
35-44 years	0.17	0.14	0.14
45-54 years	0.18	0.19	0.19
55-64 years	0.16	0.21	0.21
65+ years	0.19	0.26	0.26
Race			
Non-Hispanic White	0.65	0.75	0.75
Non-Hispanic Black	0.12	0.11	0.11
Hispanic	0.15	0.08	0.08
Other	0.07	0.06	0.06

Estimates of the Proportion of the Jewish Population by Survey – Canadian Batches

In this section, I present the ranges of the estimates of the proportions of the adult Jewish population found in the surveys forming the two Canadian batches. Estimates of the weighted proportions of the adult Jewish population were calculated for each survey independently using survey-specific weights provided by the original researchers.

The weighted estimates of the proportions of the total adult Jewish population in the surveys that comprise the Ca2001 batch ranged from .3% to 2.1% (the unweighted estimates, from .23% to 1.78%). The distribution of weighted estimates of the Jewish population across the Ca2011 batch surveys ranged from .27% to 1.98% (unweighted estimates vary between .24% and 1.78%). Estimates did not vary by the year the survey was conducted. Correlation between Ca2001 estimates (both weighted and unweighted estimates) of the adult Jewish population and the year the survey was conducted was lower than .28 for Ca2001 and lower than .14 for Ca2011. The differences between weighted and unweighted estimates also changes across surveys. Figures 4.1.1 and 4.1.2 display the weighted and unweighted distributions of the adult Jewish population for each of the surveys comprising the Ca2001 and Ca2011 batches. Data are organized based on weighted estimates.

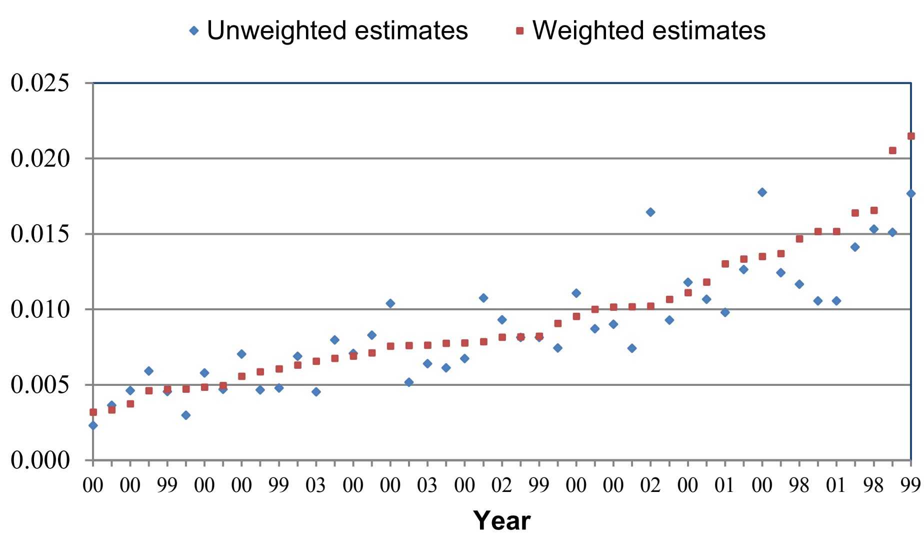

Figure 4.1.1. Weighted and unweighted distributions of the adult Jewish population for each of the surveys comprising the Ca2001 batch.

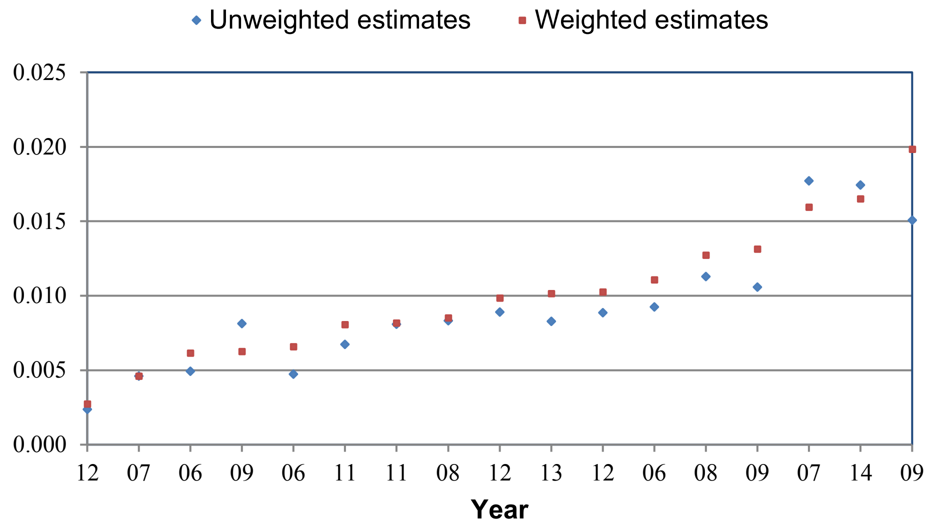

Figure 4.1.2. Weighted and unweighted distributions of the adult Jewish population for each of the surveys comprising the Ca2011 batch.

As expected, province and metropolitan area level estimates vary from survey to survey. For example, weighted estimates of the adult Jewish population in the metropolitan area of Montreal measured by the surveys included in the Ca2001 batch range from .3% to 5.9%.

Estimates of the Proportion of the Jewish Population by Survey – U.S. Batch

This section provides the ranges of the estimates of the proportions of the adult Jewish population found in the surveys comprising the US2011 batch. Estimates of the weighted proportion of this population were calculated for each survey independently using survey specific weights supplied by the original researchers.

The estimates of the Jewish population in the largest metropolitan areas measured with data from these surveys are diverse. They range from 1.01 % to 8.4% for the total sample (unweighted estimates range from 1.7% to 8.4%). Figures 4.1.3 and 4.1.4 present the weighted and unweighted distributions of the adult Jewish population for each of the surveys in the US2011a and US2011b batches. Information is organized by weighted estimates. The horizontal axis displays the year each survey was conducted.

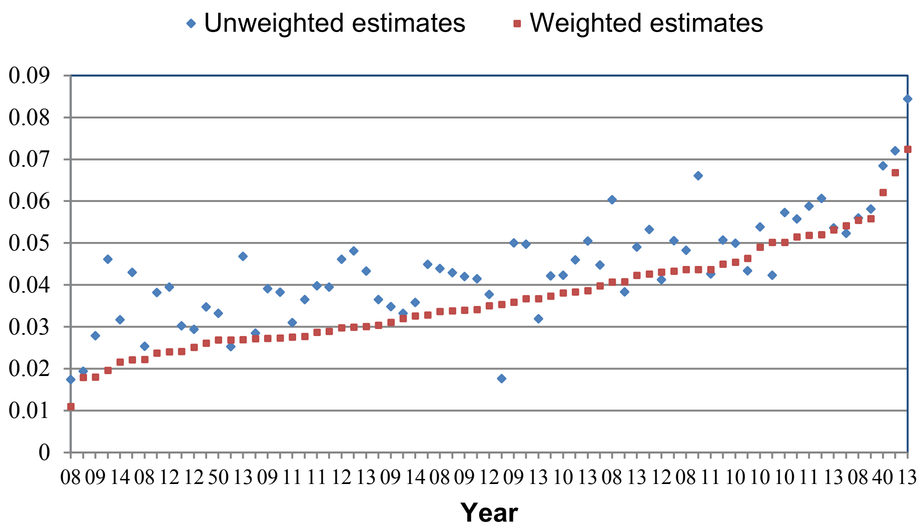

Figure 4.1.3. Weighted and unweighted distributions of the adult Jewish population for each of the surveys comprising the US2011a batch.

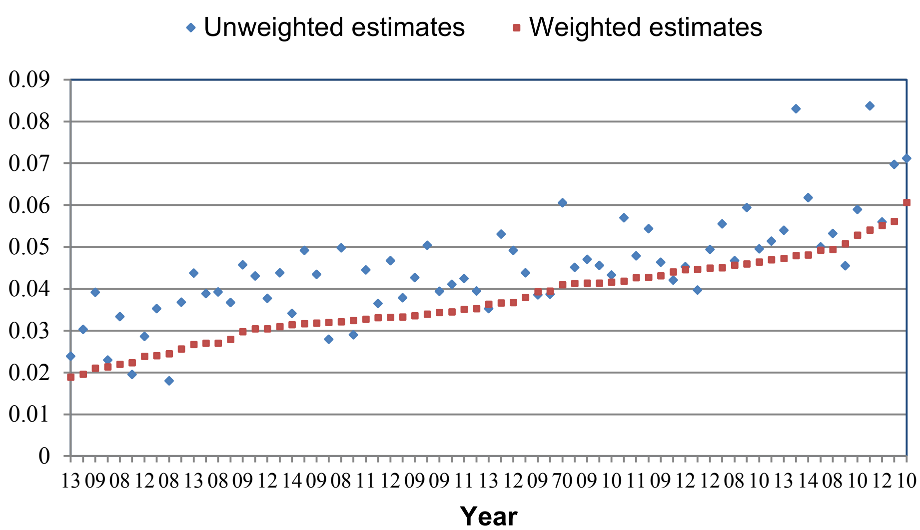

Figure 4.1.4. Weighted and unweighted distributions of the adult Jewish population for each of the surveys comprising the US2011b batch.

As was expected, metropolitan area level estimates measured with data from the different surveys also vary. For example, weighted estimates of the adult Jewish population in the metropolitan area of Los Angeles-Long Beach-Anaheim, CA measured with data from the surveys making up the US2011 batch range from .3% to 5.9%.

Estimates of the Adult Jewish Population in Canada Generated with Cross-Survey Methods

This section describes the process whereby adult Jewish population estimates were generated with each cross-survey method. Seven different estimates were calculated for each batch:

- $\widehat{\theta_M}$ estimates of the proportion of the adult Jewish population for the Montreal Metropolitan Area;
- $\widehat{\theta_T}$ estimates of the proportion of the adult Jewish population for the Toronto Metropolitan Area;
- $\widehat{\theta_V}$ estimates of the proportion of the adult Jewish population for the Vancouver Metropolitan Area;
- $\widehat{\theta_O}$ estimates of the proportion of the adult Jewish population for the Canadian province of Ontario;
- $\widehat{\theta_Q}$ estimates of the proportion of the adult Jewish population for the Canadian province of Quebec;
- $\widehat{\theta_N}$ estimates of the proportion of the adult Jewish population for the Canadian province of Manitoba; and
- $\widehat{\theta_{BC}}$ estimates of the proportion of the adult Jewish for the Canadian province of British Columbia.

All surveys, except for three that were conducted by Statistics Canada, have a total weight that resembles the sample size. Statistics Canada surveys are weighted to the total population. Prior to this analysis, I created a new weight for each of the three surveys that sums to the sample size. The new weight (*nwt*) for each respondent *i* in survey *j* was calculated as follows:

$nwt_{ij}=wt_{ij}*n_j/\ sum(wt)$

where nwt_{ij} is the new calculated weight for respondent *i* in survey *j*, wt_{ij} is the original weight of respondent *i* in survey *j*, n_j is the sample size of survey *j*, and *sum (wt)* is the sum of the weight of survey *j* respondents.

Estimates Generated with Method 1 – Meta-Analysis of Complex Surveys (MACS)

The first cross-survey method reviewed in this study generates population estimates by combining individual survey estimates. The estimates obtained represent the weighted average of the estimates across all survey samples. As previously noted, weights are used to take into consideration the unique characteristics of each survey.

This process involves three main steps. The first step consists of generating individual estimates of the adult Jewish population by geographical area for each survey *j* using the final survey weights provided by the researchers. In addition, related population variances, $v\ (\widehat{\theta}_J)$, and design effects, $deff_J$, are also calculated for individual surveys.

In the second step survey level weight variables (*wj*) are calculated for each survey. These variables are calculated separately for each of the seven population estimates generated for each survey. As was indicated earlier, survey level weight variables are calculated as a function of the variance estimates and the design effect of the survey. For example, the weight variable for survey j that was used to combine the estimates of the Jewish population in Toronto was calculated as.

wTj=1/v($\widehat{\theta_{TJ}}$) /deff$_j$

where $v(\widehat{\theta_{TJ}})$ is the variance of the estimate of the Jewish population in the Toronto metropolitan area measured with survey *j* data only.

Table 4.2a.1 displays an example of the variables that are calculated to estimate the adult Jewish population in Toronto for the Ca2001 batch. (Appendix B provides this information for all the geographical areas included in the Canadian batches.)

Table 4.2a.1

Survey Level Variables Calculated with the MACS Method to Estimate the Adult Jewish Population in Toronto

Survey ID	Sample Size	Proportion Jewish Adult Population	Var	SE	95% LCI	95% UCI	DEFF	CV
12440402	6426	0.031	0.0002	0.002	0.027	0.036	1.00	7.66
12450103	156	0.013	0.0065	0.009	0.003	0.051	1.94	70.28
12451003	151	0.019	0.0062	0.011	0.006	0.057	1.81	57.34
12454100	151	0.019	0.0064	0.011	0.006	0.059	1.83	57.33
12454101	164	0.037	0.0059	0.015	0.017	0.080	1.80	40.23
12454198	152	0.059	0.0062	0.019	0.031	0.110	1.45	32.34
12454199	149	0.062	0.0066	0.020	0.032	0.116	1.88	32.57
12454200	156	0.006	0.0064	0.006	0.001	0.044	1.83	99.70
12454201	154	0.044	0.0061	0.016	0.021	0.090	1.74	37.01
12454298	153	0.059	0.0062	0.019	0.031	0.109	1.46	32.35
12454299	152	0.032	0.0063	0.014	0.013	0.075	1.75	44.58
12454300	157	0.027	0.0087	0.015	0.009	0.080	1.73	57.18
12454301	154	0.019	0.0063	0.011	0.006	0.058	1.77	57.42
12454398	155	0.032	0.0062	0.014	0.013	0.075	1.45	44.01
12454399	147	0.012	0.0060	0.009	0.003	0.047	1.67	70.37
12454400	148	0.019	0.0063	0.011	0.006	0.058	1.80	57.30
12454401	149	0.066	0.0063	0.020	0.036	0.119	1.79	30.82
12454498	161	0.050	0.0059	0.017	0.025	0.096	1.40	34.48
12454499	178	0.008	0.0040	0.006	0.002	0.031	1.14	70.61
12470300	244	0.025	0.0040	0.010	0.011	0.054	1.77	40.42
124910504	393	0.000					0.00	
124911000	289	0.032	0.0038	0.011	0.016	0.062	1.48	34.26
124930100	114	0.004	0.0043	0.004	0.001	0.030	0.53	100.21
124930200	116	0.010	0.0052	0.007	0.002	0.040	0.56	71.36

Table 4.2a.1

Survey Level Variables Calculated with the MACS Method to Estimate the Adult Jewish Population in Toronto- Continuation

Survey ID	Sample Size	Proportion Jewish Adult Population	Var	SE	95% LCI	95% UCI	DEFF	CV
12440402	6426	0.031	0.0002	0.002	0.027	0.036	1.00	7.66
12450103	156	0.013	0.0065	0.009	0.003	0.051	1.94	70.28
12451003	151	0.019	0.0062	0.011	0.006	0.057	1.81	57.34
12454100	151	0.019	0.0064	0.011	0.006	0.059	1.83	57.33
12454101	164	0.037	0.0059	0.015	0.017	0.080	1.80	40.23
12454198	152	0.059	0.0062	0.019	0.031	0.110	1.45	32.34
12454199	149	0.062	0.0066	0.020	0.032	0.116	1.88	32.57
12454200	156	0.006	0.0064	0.006	0.001	0.044	1.83	99.70
12454201	154	0.044	0.0061	0.016	0.021	0.090	1.74	37.01
12454298	153	0.059	0.0062	0.019	0.031	0.109	1.46	32.35
12454299	152	0.032	0.0063	0.014	0.013	0.075	1.75	44.58
12454300	157	0.027	0.0087	0.015	0.009	0.080	1.73	57.18
12454301	154	0.019	0.0063	0.011	0.006	0.058	1.77	57.42
12454398	155	0.032	0.0062	0.014	0.013	0.075	1.45	44.01
12454399	147	0.012	0.0060	0.009	0.003	0.047	1.67	70.37
12454400	148	0.019	0.0063	0.011	0.006	0.058	1.80	57.30
12454401	149	0.066	0.0063	0.020	0.036	0.119	1.79	30.82
12454498	161	0.050	0.0059	0.017	0.025	0.096	1.40	34.48
12454499	178	0.008	0.0040	0.006	0.002	0.031	1.14	70.61
12470300	244	0.025	0.0040	0.010	0.011	0.054	1.77	40.42
124910504	393	0.000					0.00	
124911000	289	0.032	0.0038	0.011	0.016	0.062	1.48	34.26
124930100	114	0.004	0.0043	0.004	0.001	0.030	0.53	100.21
124930200	116	0.010	0.0052	0.007	0.002	0.040	0.56	71.36
124930300	114	0.019	0.0075	0.012	0.006	0.064	0.91	62.71
124930400	119	0.019	0.0093	0.013	0.005	0.072	1.18	70.71
124930499	126	0.009	0.0093	0.009	0.001	0.064	1.23	99.60
124930500	108	0.037	0.0096	0.019	0.013	0.098	0.99	51.08
124930599	118	0.015	0.0092	0.012	0.003	0.068	1.18	77.88
124930600	126	0.028	0.0066	0.014	0.011	0.072	0.87	48.45
124930700	107	0.006	0.0056	0.006	0.001	0.039	0.62	100.10
124930799	116	0.045	0.0094	0.020	0.018	0.107	1.17	45.78
124930800	111	0.016	0.0070	0.011	0.004	0.057	0.79	66.37
124930899	112	0.015	0.0049	0.008	0.005	0.045	0.58	58.07
124930900	108	0.026	0.0087	0.015	0.008	0.079	1.09	57.31
124931000	115	0.023	0.0074	0.013	0.007	0.068	0.88	57.26
124931099	108	0.035	0.0114	0.020	0.011	0.104	1.31	56.68
124931100	114	0.036	0.0122	0.021	0.011	0.109	1.48	58.42
124931199	116	0.018	0.0090	0.013	0.004	0.070	1.12	70.41
124931200	117	0.044	0.0100	0.021	0.017	0.110	1.25	47.57
124950702	70	0.038	0.0123	0.022	0.012	0.112	0.90	56.88
124970197	243	0.038	0.0054	0.014	0.018	0.078	1.99	37.66
124990300	96	0.000					0.00	

Not all surveys included Jewish respondents for all geographical areas. For this reason, the survey level weight (*w*) for the surveys in the areas with no Jewish respondents cannot be defined. The calculation of population estimates for these areas might hence be inflated. Since a review of the literature did not provide any guidance for this type of situation, I did the calculations using a slightly alternative method (alt-meth). The alternative calculation consisted in replacing the values of the variance and design effects for the survey-geographical area with no Jewish population with the variance and design effects of the survey-geographical area with the non-zero smallest Jewish population estimate.

Tables 4.2a.2 and 4.2a.3 present the estimated proportions of the adult Jewish population by geographical area calculated for the two Canadian batches with both the original and the alternative methods (zero population estimates replaced). The tables include the proportions of surveys with no adult Jewish respondents for the different geographical areas.

Table 4.2a.2

Ca2001 Estimates of the Adult Jewish Population Calculated with the Two Variations of MACS

Geographical Area	2001 MACS	2001 MACS (alt-meth) [a]	Proportion of Surveys with no Jewish Respondents
Metropolitan Area			
Montreal	0.017	0.010	0.18
Toronto	0.027	0.026	0.05
Vancouver	0.011	0.005	0.34
Province			
Quebec	0.008	0.005	0.19
Ontario	0.012	0.011	0.02
Manitoba	0.019	0.011	0.51
British Columbia	0.009	0.007	0.30

Note: [a] refers to the alternative mode used to calculate MACS estimates.

Table 4.2a.3

Ca2011 Estimates of the Adult Jewish Population Calculated with the Two Variations of MACS

Geographical Area	2011 MACS	2011 MACS (alt-meth) [a]	Proportion of Surveys with no Jewish Respondents
Metropolitan Area			
Montreal	0.017	0.015	0.10
Toronto	0.024	0.024	0.00
Vancouver	0.024	0.021	0.60
Province			
Quebec	0.008	0.008	0.24
Ontario	0.016	0.016	0.00
Manitoba	0.009	0.006	0.41
British Columbia	0.008	0.007	0.35

Note: [a] refers to the alternative mode used to calculate MACS estimates.

As expected, the estimates calculated with the alternative method, generated lower estimates. The differences between the estimates obtained in the two ways is more pronounced for less populated areas with small-incidence Jewish populations. It is worth noting the large proportion of surveys with no adult Jewish respondents in Vancouver in the Ca2011 batch (60% compared to 34% in Ca2001). The results provided in the next sections correspond to estimates calculated with the alternative method – replacement of values for survey-geographical areas with zero population.

Estimates were obtained by means of the STATA statistical software. Appendix C provides additional information related to the software and syntaxes used to generate these estimates.

Estimates Generated with Method 2 - Pooled Design-Based Cross-Survey (PDCS)

The second cross-survey method reviewed in this study generates population estimates from a pooled sample comprising individual respondents from the different surveys contained in a batch. Each record in the pooled sample is assigned a revised weight, which is a function of the weight of record *i* in survey *j* (w_{ij}) and a survey coefficient k_j:

$$w_{ij}^{*} = w_{ij} * k_j$$

where *k*j is defined as a function of the survey's sample size and coefficient of variation (CV):

$$k_j = \frac{\sum_{i=1}^{n_1} w_{i1} + \cdots + \sum_{i=1}^{n_L} w_{iL}}{L \sum_{i=1}^{n_j} w_{ij}} (1 - \frac{(CV_j^2 + 1)/n_j}{(CV_1^2 + 1)/n_1 \ + \cdots + \ (CV_L^2 + 1)/n_L})$$

The first step to obtain population estimates is to calculate these survey level coefficients (k_j). Seven coefficients were calculated for each survey, one for each of the seven population estimates (estimates for the three largest metro areas and for the four provinces with the largest proportions of adult Jewish population).

In the second step, the records of the individual surveys that compose each batch are assigned the new weight. In the third step estimates are generated for both pooled samples (one for each batch) using the new assigned weights.

As noted before, some surveys had no Jewish respondents in certain geographical areas. For this reason, the coefficient of variation (CV) cannot be defined for these areas in those surveys. As was the case with MACS, I found no possible solutions for this problem in the literature. Therefore, I calculated the estimates in two different ways. The first (mn1- miss) consists of setting the coefficient k for the given survey and geographical area as missing. The second (mn2-avg) involves replacing the coefficient of variation with the value of the averaged coefficient k for the given geographical area across surveys with valid CVs. Tables 4.2b.1 and 4.2b.2 show the estimates generated with both methods side by side, as well as the proportion of surveys for each geographical area that did not include Jewish respondents. Table 4.2b.1 displays the results for the Ca2001 batch, and Table 4.2b.2 for the Ca2011 batch.

Table 4.2b.1

Ca2001 Estimates of the Adult Jewish Population Generated with the Two Variations of PDCS

Geographical Area	2001 PDCS (mn1-miss)	2001 PDCS (mn2-avg)	Proportion of Surveys with no Jewish Respondents
Metropolitan Area			
Montreal	0.020	0.018	0.18
Toronto	0.030	0.029	0.05
Vancouver	0.012	0.010	0.34
Province			
Quebec	0.009	0.008	0.19
Ontario	0.016	0.015	0.02
Manitoba	0.017	0.015	0.51
British Columbia	0.010	0.008	0.30

Table 4.2b.2

Ca2011 Estimates of the Adult Jewish Population Generated with the Two Variants of PDCS

Geographical Area	2011 PDCS (mn1-miss)	2011 PDCS (mn2-avg)	Proportion of Surveys with no Jewish Respondents
Metropolitan Area			
Montreal	0.019	0.018	0.10
Toronto	0.030	0.030	0.00
Vancouver	0.025	0.010	0.60
Province			
Quebec	0.008	0.008	0.24
Ontario	0.015	0.016	0.00
Manitoba	0.015	0.009	0.41
British Columbia	0.008	0.007	0.35

As can be expected, the first method results in higher estimates for the geographical areas, as estimates of 0 for the adult Jewish population are not included in the analyses. Analyses presented later are based on estimates calculated in the second manner.

STATA was also used to generate estimates with the pooled design-based approach. Appendix C provides additional information regarding the software and syntaxes used to obtain these estimates.

Estimates Generated with Method 3 - Bayesian Multilevel Regression with Post-Stratification (BMRP)

The third cross-survey method reviewed in this study generates population estimates by combining data through a model-based approach. Two main steps were followed. In the first step Bayesian multilevel logistic regression models were built to model the probability that an individual survey respondent is Jewish as a function of demographic and geographical predictors. In the second step estimates for each demographic-geographical response were post-stratified by the percentages of each actual demographic-geographical combination (Kastellec, Lax, & Phillips, 2010).

For this study, I built four separate models: a metropolitan-area-based model and a province-based model for each Canadian batch. Basic demographic variables (age, educational attainment, gender, and income) associated with the distribution of the Jewish population in Canada were tested to build the different models. Age was represented as five categories: 25-34, 35-44, 45-54, 55-64 and over 65; education, as five categories: no certificate, high school certificate, trade/college certificate, university certificate and post-college certificate; and income, as three categories: under $60,000,

$60,000 to $80,000 and over $80,000. Appendix C provides additional information concerning the software and syntaxes used to generate the estimates.

The first one, Mod2001met, estimates the probability of being Jewish in the largest Canadian metropolitan areas (Montreal, Toronto, Vancouver) using data from the Ca2001 batch. Data from 43 surveys were used to build this model. Included in the model were the individual predictors of age, education, age by education interaction, metropolitan area (Montreal, Toronto, Vancouver, other), and metropolitan area by education interaction, as well as a survey-level identification.

The second one, Mod2001prv, estimates the probability of being Jewish in the four provinces with the largest proportions of adult Jewish population using data from the Ca2001 batch. Data from 44 surveys were used to build this model. Included in the model were the individual predictors age, education, age by education interaction, income, and province (Quebec, Ontario, Manitoba, British Columbia), as well as a survey-level identification.

The third one, *Mod2011met*, estimates the probability of being Jewish in the largest Canadian metropolitan areas with data from the Ca2011 batch. Data from ten surveys were used to build this model. Included in the model were the individual predictors of age, education, age by education interaction, metropolitan area, and metropolitan area by education interaction as well as a survey-level identification.

The fourth one, Mod2011prv, estimates the probability of being Jewish in the four provinces with the largest proportions of the adult Jewish population with data from the Ca2011 batch. Data from 17 surveys were used to build this model. Included in the model were the individual predictors: age, education, age by education interaction, income, and province (Quebec, Toronto, Manitoba, British Columbia) as well as a survey-level identification.

In the second step I generated estimates for each of the seven geographical areas using the data from the models built in the first step and census population counts. As described in the methodological section, these counts were obtained from the publicly available 2001 Census of Population Public Use Microdata File and 2011 National Household Public Use Microdata File distributed by Statistics Canada (Statistics Canada, 2001; Statistics Canada, 2011).

Table 4.2c.1 and 4.2c.2 present the estimates calculated for the seven geographical areas. Table 4.2c.1 shows the estimates generated with data from the Ca2001 batch using Mod2001met and Mod2001pv, and Table 4.2c.2 shows the estimates produced with data from the CA2011 batch using Mod2011met and Mod2011pv.

Table 4.2c.1

Ca2001 Estimates of the Adult Jewish Population Calculated with BMRP

Geographical Area	Ca2001 BMRP
Metropolitan Area	
Montreal	0.027
Toronto	0.036
Vancouver	0.011
Province	
Quebec	0.007
Ontario	0.015
Manitoba	0.015
British Columbia	0.009

Table 4.2c.2

Ca2011 Estimates of the Adult Jewish Population Calculated with BMRP

Metropolitan Area	
Montreal	0.014
Toronto	0.026
Vancouver	0.006
Province	
Quebec	0.006
Ontario	0.012
Manitoba	0.011
British Columbia	0.012

Ca2001 estimates are higher than Ca2011 estimates for all geographical areas except for British Columbia. As discussed below, in addition to the actual lower proportion of Jewish population in 2011, lower estimates could be a result of the smaller sample sizes included in the Ca2011 analyses.

I relied on R language, along with STAN and LME packages, to build the models and generate the estimates.

Estimates of the Adult Jewish Population in the US Generated with the Three Cross-Survey Methods

This section describes the process whereby adult Jewish population estimates were generated for the 20 metropolitan areas with the three cross-survey methods based on data from the US2011a and US2011b sub-batches. Each method is discussed separately. Twenty different estimates were obtained for each sub-batch using each method. For the sake of simplicity, I assigned a sequential number to the 20 metropolitan areas included in this study. For example, metro1 refers to the New-York-Newark-Jersey

City, NY-NJ-PA metropolitan area (CBSA 35620). Table 4.3.1 shows the 20 metropolitan areas.

Table 4.3.1

U.S. Metropolitan Areas Included in the Study

Number	CBSA	Name
metro1	35620	New York-Newark-Jersey City, NY-NJ-PA
metro2	31080	Los Angeles-Long Beach-Anaheim, CA
metro3	16980	Chicago-Naperville-Elgin, IL-IN-WI
metro4	19100	Dallas-Fort Worth-Arlington, TX
metro5	26420	Houston-The Woodlands-Sugar Land, TX
metro6	47900	Washington-Arlington-Alexandria, DC-VA-MD-WV
metro7	37980	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD
metro8	33100	Miami-Fort Lauderdale-West Palm Beach, FL
metro9	12060	Atlanta-Sandy Springs-Roswell, GA
metro10	14460	Boston-Cambridge-Newton, MA-NH
metro11	41860	San Francisco-Oakland-Hayward, CA
metro12	38060	Phoenix-Mesa-Scottsdale, AZ
metro13	40140	Riverside-San Bernardino-Ontario, CA
metro14	19820	Detroit-Warren-Dearborn, MI
metro15	42660	Seattle-Tacoma-Bellevue, WA
metro16	33460	Minneapolis-St. Paul-Bloomington, MN-WI
metro17	41740	San Diego-Carlsbad, CA
metro18	45300	Tampa-St. Petersburg-Clearwater, FL
metro19	41180	St. Louis, MO-IL
metro20	19740	Denver-Aurora-Lakewood, CO

Estimates generated using Method 1 - Meta-Analysis of Complex Surveys (MACS)

As previously noted, the MACS cross-survey method generates population estimates by combining individual survey estimates. The process involves three steps. In the first step estimates of the adult Jewish population in each metropolitan area are generated individually for each survey j using the final survey weights provided by the researchers. In the second, survey level weight variables (wj) are calculated separately for the 20 metropolitan areas within each survey. These variables are calculated as a function of the variance estimates and the design effect of the survey.

Table 4.3a.1 displays an example of the variables that were calculated to estimate the adult Jewish population in metropolitan area metro2 – Los Angeles-Long Beach-Anaheim, CA. (The table includes results from ten surveys in the US2001a sub-batch.)

Table 4.3a.1

Survey Level Variables Calculated with the MACS Method to Estimate the Adult Jewish Population in metro2 – Los Angeles-Long Beach-Anaheim, CA

Survey Id	Sample Size	Proportion Jewish Adult Population	Var	SE	95% LCI	95% UCI	DEFF	CV	Weighted Sample Size
192508	2	35.000	0.0405	0.020	0.029	0.010	0.15	1.00	71.0
194108	2	34.000	0.1258	0.026	0.057	0.050	0.28	1.15	45.0
194208	2	31.000	0.0227	0.023	0.023	0.003	0.15	0.92	100.5
194408	2	26.000	0.0521	0.026	0.037	0.012	0.19	0.70	70.8
1218121	2	80.000	0.0637	0.020	0.036	0.020	0.18	1.34	56.4
1218122	2	163.000	0.0103	0.005	0.007	0.003	0.04	0.71	67.0
1255081	2	90.000	0.0092	0.005	0.007	0.002	0.04	0.53	71.3
1255101	2	61.000	0.0421	0.009	0.020	0.016	0.10	0.79	47.2
1255121	2	93.000	0.0103	0.005	0.007	0.003	0.04	0.61	71.1
1375081	2	80.000	0.0372	0.025	0.031	0.007	0.17	2.69	82.7

In the last step, population estimates are generated separately for each of the 20 metropolitan areas in each US2011 sub-batch. These estimates are calculated as a weighted average of the estimates of individual survey j:

$$\hat{\theta}_N = \frac{\sum_{j=1}^{k} w_j \, \hat{\theta}_j}{\sum_{j=1}^{k} w_j}$$

Tables 4.3a.2 and 4.3a.3 display the estimated proportions of the adult Jewish population by metropolitan area for the two US2011 sub-batches calculated with both the original and the alternative methods (zero population estimates replaced).

Table 4.3a.2

US2011a Estimates of the Adult Jewish Population Calculated with the Two Variations of MACS

Metropolitan Area	US2011a MACS	US2011a MACS (alt-meth)
metro1	0.063	0.063
metro2	0.028	0.012
metro3	0.029	0.017
metro4	0.022	0.012
metro5	0.025	0.009
metro6	0.036	0.019
metro7	0.032	0.020
metro8	0.051	0.023
metro9	0.027	0.009
metro10	0.040	0.011
metro11	0.035	0.016
metro12	0.038	0.019
metro13	0.029	0.012
metro14	0.020	0.009
metro15	0.033	0.019
metro16	0.037	0.017
metro17	0.034	0.020
metro18	0.039	0.012
metro19	0.046	0.030
metro20	0.044	0.019

Table 4.3a.3

US2011b Estimates of the Adult Jewish Population Calculated with the Two Variations of MACS

Metropolitan Area	US2011b MACS	US2011b MACS (alt-meth)
metro1	0.074	0.072
metro2	0.027	0.020
metro3	0.018	0.005
metro4	0.025	0.010
metro5	0.018	0.009
metro6	0.044	0.021
metro7	0.033	0.012
metro8	0.081	0.069
metro9	0.009	0.002
metro10	0.046	0.023
metro11	0.045	0.021
metro12	0.035	0.013
metro13	0.021	0.010
metro14	0.021	0.005
metro15	0.036	0.017
metro16	0.028	0.016
metro17	0.018	0.008
metro18	0.015	0.004
metro19	0.026	0.007
metro20	0.041	0.022

Estimates Generated using Method 2 - Pooled Design-Based Cross-Survey

As noted earlier, the PDCS method generates population estimates of a pooled sample composed of the individual respondents of the different surveys forming a batch. Each record in the pooled sample is assigned a revised weight, which is a function of the weight of record i in survey j (w_{ij}), and a survey coefficient k_j:

$$w_{ij}^{*} = w_{ij} * k_j$$

where kj is defined as a function of the survey's sample size and coefficient of variation (CV):

$$k_j = \frac{\sum_{i=1}^{n_1} w_{i1} + \cdots + \sum_{i=1}^{n_L} w_{iL}}{L\sum_{i=1}^{n_j} w_{ij}}(1 - \frac{\left(CV_j^2 + 1\right)/n_j}{(CV_1^2 + 1)/n_1 \ + \cdots + \ (CV_L^2 + 1)/n_L}$$

Tables 4.3b.1 and 4.3b.2 show the estimates generated with this method for each metropolitan area. Table 4.3b.1 sets forth the results for the US2011a sub-batch, and Table 4.3b.2, for the US2011b sub-batch.

Table 4.3b.1

US2011a Estimates of the Adult Jewish Population Generated with the Two Variations of PDCS

Metropolitan Area	US2011a PDCS	US2011a PDCS (alt-avg)
metro1	0.0780	0.0780
metro2	0.0456	0.0405
metro3	0.0404	0.0317
metro4	0.0341	0.0073
metro5	0.0428	0.0098
metro6	0.0546	0.0397
metro7	0.0502	0.0394
metro8	0.0865	0.0777
metro9	0.0401	0.0156
metro10	0.0645	0.0342
metro11	0.0643	0.0398
metro12	0.0478	0.0144
metro13	0.0436	0.0120
metro14	0.0357	0.0097
metro15	0.0557	0.0153
metro16	0.0408	0.0118
metro17	0.0548	0.0205
metro18	0.0685	0.0264
metro19	0.0813	0.0183
metro20	0.0727	0.0267

Table 4.3b.2

US2011b Estimates of the Adult Jewish Population Generated with the Two Variations of PDCS

Metropolitan Area	US2011b PDCS	US2011b PDCS (alt-avg)
metro1	0.0915	0.0909
metro2	0.0391	0.0342
metro3	0.0364	0.0262
metro4	0.0460	0.0105
metro5	0.0262	0.0063
metro6	0.0574	0.0398
metro7	0.0534	0.0361
metro8	0.1073	0.0997
metro9	0.0414	0.0158
metro10	0.0680	0.0485
metro11	0.0799	0.0438
metro12	0.0387	0.0158
metro13	0.0404	0.0059
metro14	0.0337	0.0112
metro15	0.0484	0.0137
metro16	0.0290	0.0109
metro17	0.0493	0.0142
metro18	0.0401	0.0187
metro19	0.0658	0.0161
metro20	0.0570	0.0145

Estimates Generated using Method 3 - Bayesian Multilevel Regression with Post-Stratification

The BMRP cross-survey method generates population estimates by combining data through a model-based approach. As described earlier, estimates are produced in two steps. In the first step, Bayesian multilevel logistic regression models are built to model the probability that an individual survey responder will be Jewish as a function of demographic and geographical predictors. In the second step, estimates for each demographic-geographical response are post-stratified by the percentages of each actual demographic-geographical combination (Kastellec, Lax & Phillips, 2010).

Estimates for the two US2011 sub-batches are generated separately but with the same model. In other words, the variables included in the models used to obtain the US2011a and US2011b estimates are the same. In the second step, estimates were calculated for the 20 metropolitan areas in each sub-batch with demographic data from the 2014 census. Table 4.3c.1 presents the estimates generated with data from the US2011a sub-batch for the 20 metropolitan areas, and Table 4.3c.2 presents the estimates obtained with data from the US2011b sub-batch for the 20 metropolitan areas.

Table 4.3c.1

US2011a Estimates of the Jewish Population Generated with BMRP

	US2011a BMRP		
Metropolitan Area	Estimate	95% LCI	95% UCI
metro1	0.0643	0.0579	0.0709
metro2	0.0301	0.0254	0.0353
metro3	0.0286	0.0234	0.0344
metro4	0.0065	0.0045	0.0088
metro5	0.0063	0.0044	0.0089
metro6	0.0286	0.0238	0.0344
metro7	0.0337	0.0280	0.0403
metro8	0.0645	0.0545	0.0755
metro9	0.0095	0.0069	0.0128
metro10	0.0325	0.0246	0.0414
metro11	0.0283	0.0215	0.0359
metro12	0.0132	0.0096	0.0175
metro13	0.0074	0.0050	0.0105
metro14	0.0095	0.0067	0.0129
metro15	0.0105	0.0072	0.0145
metro16	0.0110	0.0078	0.0148
metro17	0.0165	0.0118	0.0224
metro18	0.0163	0.0116	0.0222
metro19	0.0102	0.0071	0.0141
metro20	0.0196	0.0138	0.0261

Table 4.3c.2

US2011b Estimates of the Jewish Population Generated with BMRP

Metropolitan Area	BMRP	US2011b BMRP 95% LCI	95% UCI
metro1	0.0770	0.0700	0.0842
metro2	0.0293	0.0251	0.0341
metro3	0.0270	0.0221	0.0327
metro4	0.0077	0.0053	0.0104
metro5	0.0062	0.0043	0.0087
metro6	0.0316	0.0265	0.0379
metro7	0.0369	0.0303	0.0436
metro8	0.0860	0.0743	0.0976
metro9	0.0099	0.0071	0.0132
metro10	0.0566	0.0456	0.0685
metro11	0.0283	0.0221	0.0350
metro12	0.0161	0.0117	0.0208
metro13	0.0061	0.0042	0.0085
metro14	0.0109	0.0075	0.0154
metro15	0.0104	0.0074	0.0145
metro16	0.0116	0.0085	0.0156
metro17	0.0136	0.0095	0.0183
metro18	0.0195	0.0140	0.0258
metro19	0.0103	0.0074	0.0141
metro20	0.0175	0.0123	0.0232

Research Question Analyses

This section presents the different indicators and the separate analyses that were used to answer each research question.

Research Question 1- How do cross-survey estimates of the proportions of the total adult Jewish population in Canadian provinces and in metropolitan areas compare to the estimates from the Canadian census and National Household Survey (NHS)?

To answer this question, I conducted two main sets of analyses. The first set identifies the differences between estimates of the adult Jewish population generated by the three cross-survey methods with data from the Ca2001 batch, and estimates produced with data from the 2001 census. The second compares the estimates of the adult Jewish population produced by the three cross-survey methods using data from the Ca2011 batch, with estimates from the 2011 national household survey (NHS).

The following sections present the results of both sets of analyses organized by cross-survey method. For each method, I provide results of four sets of assessments:

- Assessment of the differences between the proportions of the adult Jewish population in the three major Canadian metropolitan areas estimated by the cross-survey methods with data from the Ca2001 batch, and the proportions of the adult Jewish population in the same areas estimated with data from the 2001 Canadian census.
- Assessment of the differences between the proportions of the adult Jewish population in the four Canadian provinces with the largest Jewish population estimated by the cross-survey methods using data from the Ca2001 batch, and the proportions of the adult Jewish population in the same areas estimated with data from the 2001 Canadian census.
- Assessment of the differences between the proportions of the adult Jewish population in the three major Canadian metropolitan areas estimated by the cross-survey methods with data from the Ca2011 batch, and the proportions of the adult Jewish population in the same areas estimated with data from the 2011Canadian NHS.
- Assessment of the differences between the proportions of the adult Jewish population in the four Canadian provinces with the largest Jewish

population estimated by the cross-survey methods with data from the Ca2011 batch, and the proportions of the adult Jewish population in the same areas estimated with data from the 2011 Canadian NHS.

Results are presented separately for the different time periods and geographical areas so as to take into account the diverse number of surveys and sample sizes included in each analysis, as well as the source of the census estimates (2001 census estimates or 2011 NHS estimates). Results obtained with the cross-survey methods are compared with the census estimates in two ways: (a) by means of indicators that measure the proximity of the cross-survey estimates to the census estimates and (b) by way of precision measures.

Assessments of the Meta-Analysis of Complex Surveys (MACS) Method. This section presents the results of the comparisons between census estimates and MACS estimates of the adult Jewish population.

Tables 4.4.1 and 4.4.2 compare census estimates and MACS cross-survey estimates of the adult Jewish population for the three largest Canadian metropolitan areas and for the four provinces with the largest proportions of adult Jewish population. These tables display the proportions of the adult Jewish population calculated with census (and NHS) data along with the 95% lower confidence interval (95% LCI) and the 95% upper confidence interval (95% UCI), the MACS cross-survey estimates of the proportions and variances of the adult Jewish population, and the numeric differences (Err) and percentage errors (PE) between census and cross-survey estimates. The percentage error is calculated as the numeric difference between the cross-survey estimates and the census estimate, divided by the census estimate and multiplied by 100. The absolute percentage error (APE) is calculated as the absolute value of the percentage error. Table 4.4.1 shows data from the 2001 census and estimates based on the Ca2001 batch.

Table 4.4.1

2001 Census Estimates and Ca2001 MACS Estimates of the Proportions of the Adult Jewish Population

	2001 Census			Ca2001 MACS			Err	PE
	Proportion	95% LCI	95% UCI	Proportion	95% LCI	95% UCI		
Metropolitan Area								
Montreal	0.0268	0.0256	0.0281	0.0101	0.0079	0.0124	-0.017	-62
Toronto	0.0355	0.0343	0.0368	0.0257	0.0229	0.0284	-0.010	-28
Vancouver	0.0098	0.0088	0.0108	0.0048	0.0034	0.0061	-0.005	-51
Province								
Quebec	0.0130	0.0124	0.0136	0.0054	0.0034	0.0075	-0.008	-58
Ontario	0.0172	0.0166	0.0177	0.0114	0.0091	0.0137	-0.006	-34
Manitoba	0.0132	0.0117	0.0149	0.0105	0.0077	0.0134	-0.003	-20
BC	0.0061	0.0055	0.0067	0.0074	0.0046	0.0103	0.001	22

As can be seen in Table 4.4.1, the MACS estimates are lower than the census estimates for all geographical areas except for the province of British Columbia, for which the MACS estimate is slightly higher. Absolute percentage errors (APE) for the three metropolitan areas range from 28% for Vancouver to 62% for Montreal, and for the provinces, from 20% for British Columbia to 58% for Manitoba. Census estimates are outside the 95% CI for all geographical areas except for British Columbia. Given the low incidence of the Jewish population, survey estimates might be affected by the size of the sample in individual geographical areas, and this may explain why cross-survey estimates are generally lower than population estimates. Figure 4.4.1 presents this information graphically.

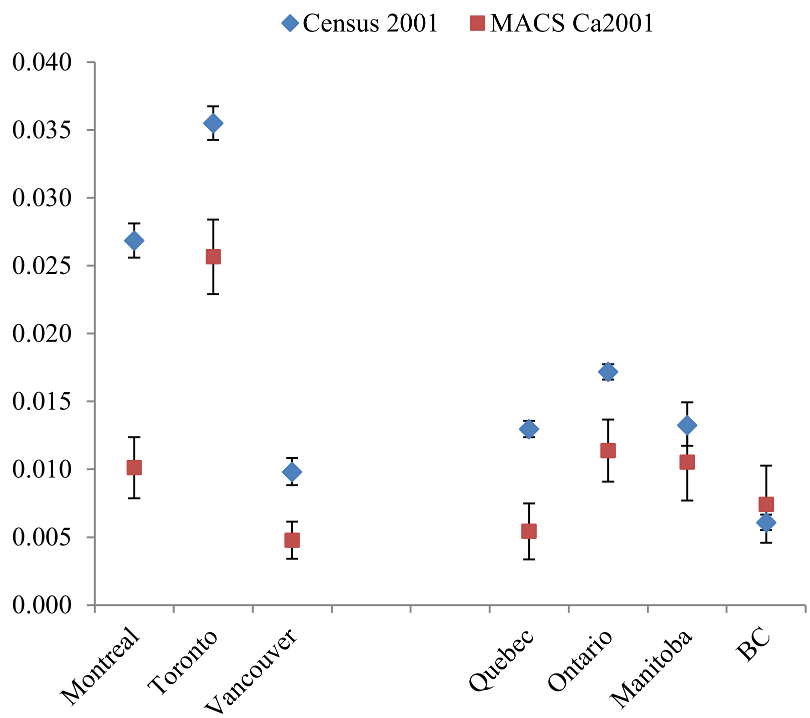

Figure 4.4.1. Comparisons between estimates of the proportions of the adult Jewish population generated with the MACS method and 2001 census estimates.

What follows is the comparison between the 2011 NHS estimates and the Ca2011 MACS cross-survey estimates of the adult Jewish population for the three largest Canadian metropolitan areas and for the four provinces with the largest proportions of adult Jewish population. Table 4.4.2 and Figure 4.4.2 present this information.

Table 4.4.2

2011 National Household Survey Estimates and Ca2011 MACS Estimates of the Proportions of the Adult Jewish Population

	2011 NHS			Ca2011 MACS			Err	PE
	Proportion	95% LCI	95% UCI	Proportion	95% LCI	95% UCI		
Metropolitan Area								
Montreal	0.0213	0.0202	0.0225	0.0146	0.0000	0.0294	-0.007	-32
Toronto	0.0298	0.0287	0.0309	0.0243	0.0094	0.0393	-0.005	-18
Vancouver	0.0079	0.0070	0.0089	0.0213	0.0000	0.0470	0.013	169
Province								
Quebec	0.0103	0.0098	0.0109	0.0078	0.0027	0.0129	-0.003	-24
Ontario	0.0147	0.0142	0.0152	0.0160	0.0115	0.0206	0.001	9
Manitoba	0.0106	0.0093	0.0121	0.0061	0.0004	0.0119	-0.004	-42
BC	0.0048	0.0043	0.0053	0.0071	0.0000	0.0142	0.002	47

As was the case with 2001 results, estimates of the adult Jewish population for Montreal, Toronto, Quebec, and Manitoba obtained with the MACS method are lower than 2011 NHS estimates. At the same time, Ca2011 MACS estimates are higher than 2011 NHS estimates for Vancouver and British Columbia, and slightly so for Ontario. Metropolitan area APEs range from 18% for Toronto to 169% for Vancouver, and province APEs, from 9% for Ontario to 47% for British Columbia. The higher-than-expected percentage error for the Vancouver area might be associated with the large number of surveys with no Jewish respondents. Unlike the cross-survey estimates obtained with the Ca2001 data, the 2011 NHS estimates for all geographical areas are within the 95% CI. This result is not surprising given the low precision of the estimates, as will be discussed next.

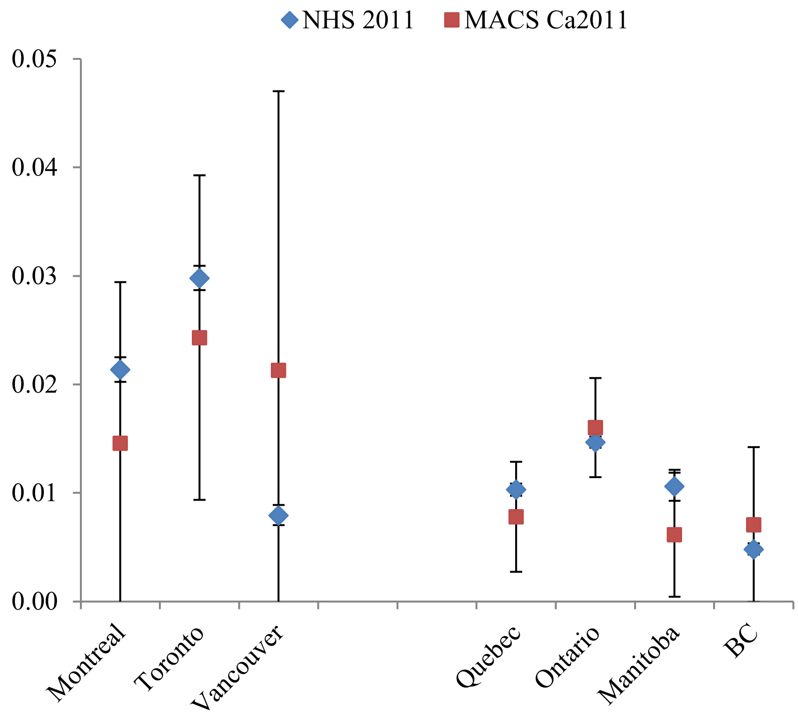

Figure 4.4.2. Comparisons between estimates of the proportions of the adult Jewish population generated with the MACS method and 2011 NHS estimates.

In terms of precision, as was expected given the number of surveys included in the different analyses, 2011 MACS estimates are less precise than 2001 MACS estimates. This difference is especially pronounced for the metropolitan area estimates. Figure 4.4.3 below compares the 95% confidence intervals for the four MACS analyses.

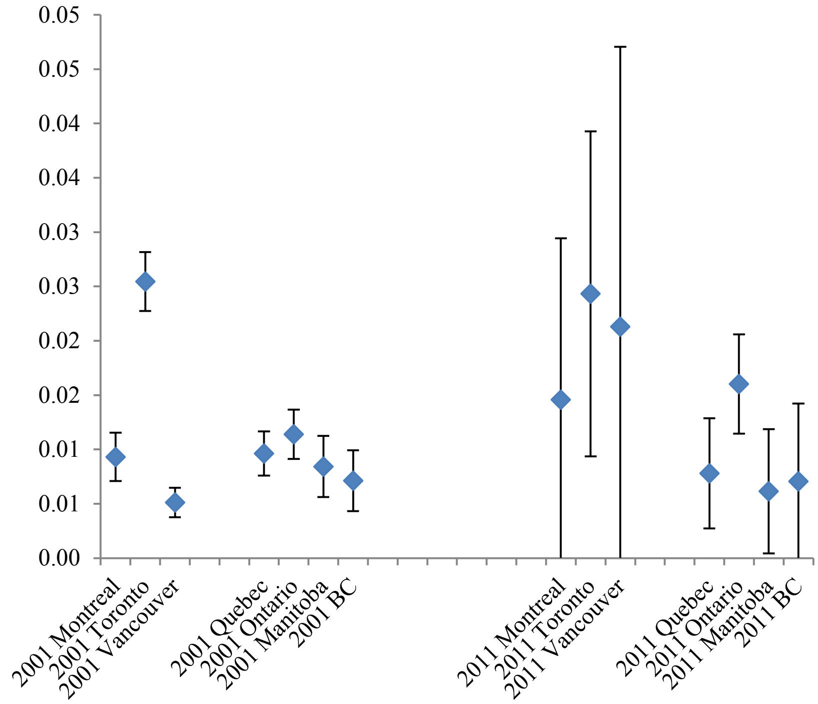

Figure 4.4.3. Comparisons between the 2001 and the 2011 estimates of the proportions of the adult Jewish population generated with the MACS method.

Assessment of the Pooled Design-Based Cross-Survey (PDCS) Method. This section presents the results of the comparisons between the proportions of the adult Jewish population measured by the 2001 Canadian census and 2011 NHS, and PDCS estimates of this population.

Table 4.4.3 compares the 2001 census estimates and Ca2001 PDCS cross-survey estimates of the adult Jewish population for the seven Canadian geographical areas included in this study.

Table 4.4.3

2001 Census Estimates and Ca2001 PDCS Estimates of the Proportions of the Adult Jewish Population

	2001 Census			Ca2001 PDCS			Err	PE
	Proportion	95% LCI	95% UCI	Proportion	95% LCI	95% UCI		
Metropolitan Area								
Montreal	0.0268	0.0256	0.0281	0.0184	0.0155	0.0212	-0.008	-31
Toronto	0.0355	0.0343	0.0368	0.0288	0.0256	0.0320	-0.007	-19
Vancouver	0.0098	0.0088	0.0108	0.0099	0.0073	0.0125	0.000	1
Province								
Quebec	0.0130	0.0124	0.0136	0.0084	0.0068	0.0100	-0.005	-35
Ontario	0.0172	0.0166	0.0177	0.0152	0.0133	0.0171	-0.002	-11
Manitoba	0.0132	0.0117	0.0149	0.0153	0.0110	0.0196	0.002	16
BC	0.0061	0.0055	0.0067	0.0081	0.0061	0.0101	0.002	34

As can be seen from the results presented in table 4.4.3, Ca2001 PDCS estimates are higher than census estimates for Vancouver, Manitoba, and British Columbia, and lower for Montreal, Toronto, Quebec, and Ontario. However, the comparison between the 2011 estimates obtained with the pooled design-based cross-survey analysis and the 2011 NHS estimates (Table 4.4.4) presents a slightly different pattern. PDCS estimates for Vancouver and British Columbia are indeed higher than NHS estimates, but so are the estimates for Ontario, if less so.

Ca2001 PDCS absolute percentage errors (APE) are similar in range to Ca2011 APEs. Ca2001 metropolitan area APEs range from 1 to 31% (for Vancouver and Montreal, respectively), and Ca2011 metropolitan area APEs, from 2 to 22% (for Toronto and Vancouver, respectively), while Ca2001 province APEs range from 11 to 35% (for

Ontario and Quebec, respectively), and Ca2011 province APEs range from 10 to 39% (for Ontario and British Columbia, respectively).

Table 4.4.4 shows the National Household Survey estimates of the proportions of the adult Jewish population, along with the corresponding Ca2011 PDCS estimates.

Table 4.4.4

2011 NHS Census Estimates and Ca2011 PDCS Estimates of the Proportions of the Adult Jewish Population

	2011 Census			Ca2011 PDCS			Err	PE
	Proportion	95% LCI	95% UCI	Proportion	95% LCI	95% UCI		
Metropolitan Area								
Montreal	0.0213	0.0202	0.0225	0.0175	0.0106	0.0245	-0.004	-18
Toronto	0.0298	0.0287	0.0309	0.0304	0.0220	0.0388	0.001	2
Vancouver	0.0079	0.0070	0.0089	0.0097	0.0025	0.0168	0.002	22
Province								
Quebec	0.0103	0.0098	0.0109	0.0079	0.0062	0.0096	-0.002	-23
Ontario	0.0147	0.0142	0.0152	0.0162	0.0142	0.0182	0.002	10
Manitoba	0.0106	0.0093	0.0121	0.0089	0.0050	0.0127	-0.002	-17
BC	0.0048	0.0043	0.0053	0.0067	0.0041	0.0093	0.002	39

Half of the 2001 census estimates (for Vancouver, Manitoba, and British Columbia) fall within the 95% confidence intervals of the corresponding PDCS estimates. All the 2011 census estimates fall within the 95% confidence intervals of the corresponding PDCS estimates, which points to the plausibility of these estimates (see Figure 4.4.4 and Figure 4.4.5 next).

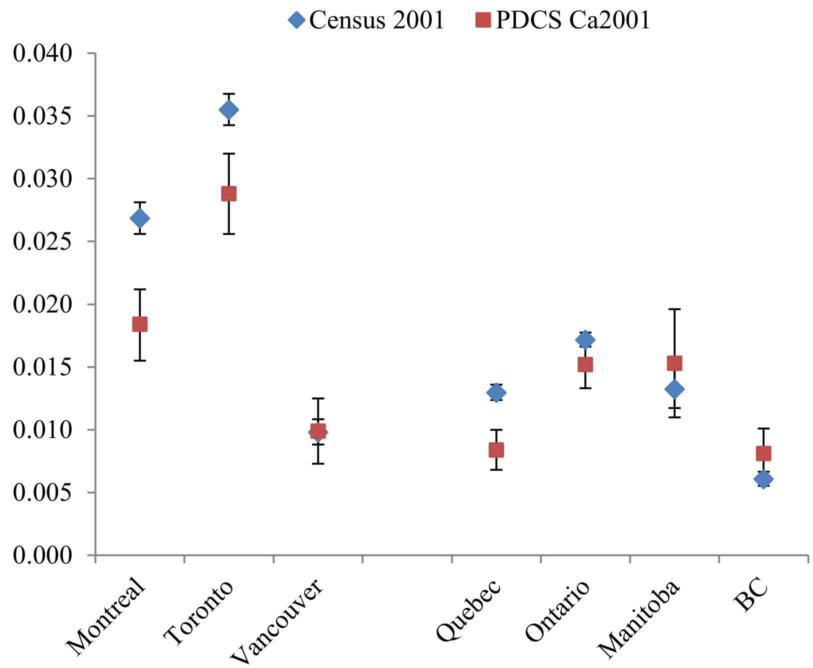

Figure 4.4.4. Comparisons between estimates of the proportions of the adult Jewish population generated with the PDCS method and 2001 census estimates.

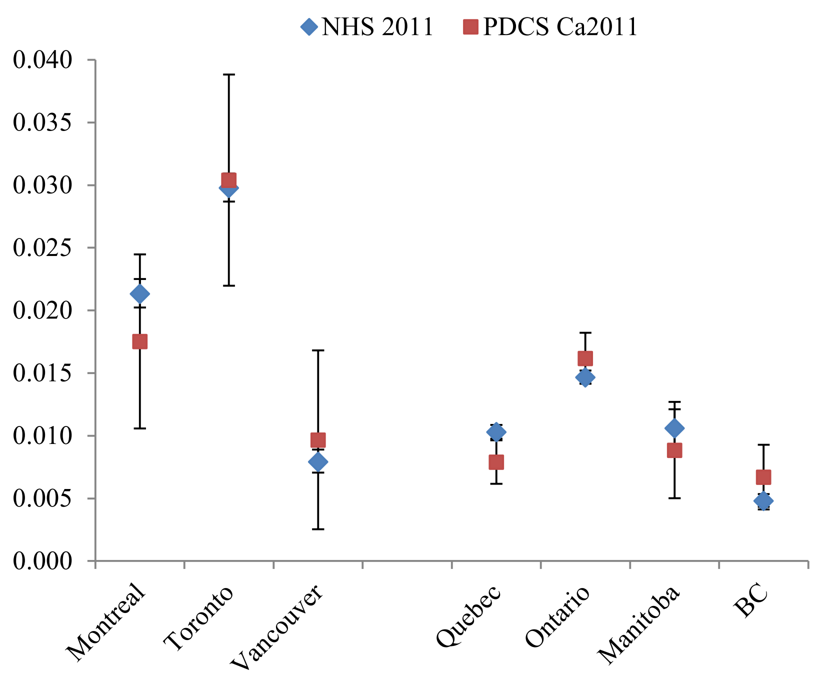

Figure 4.4.5. Comparisons between estimates of the proportions of the adult Jewish population generated with the PDCS method and 2011 NHS estimates.

The precision of the 2011 PDCS estimates for the metropolitan areas is lower than that of the 2001 PDCS estimates for the same areas. There is almost no difference between the precision of the 2001 and the 2011 PDCS estimates for the four provinces included in this study. Figure 4.4.6 below presents this information graphically.

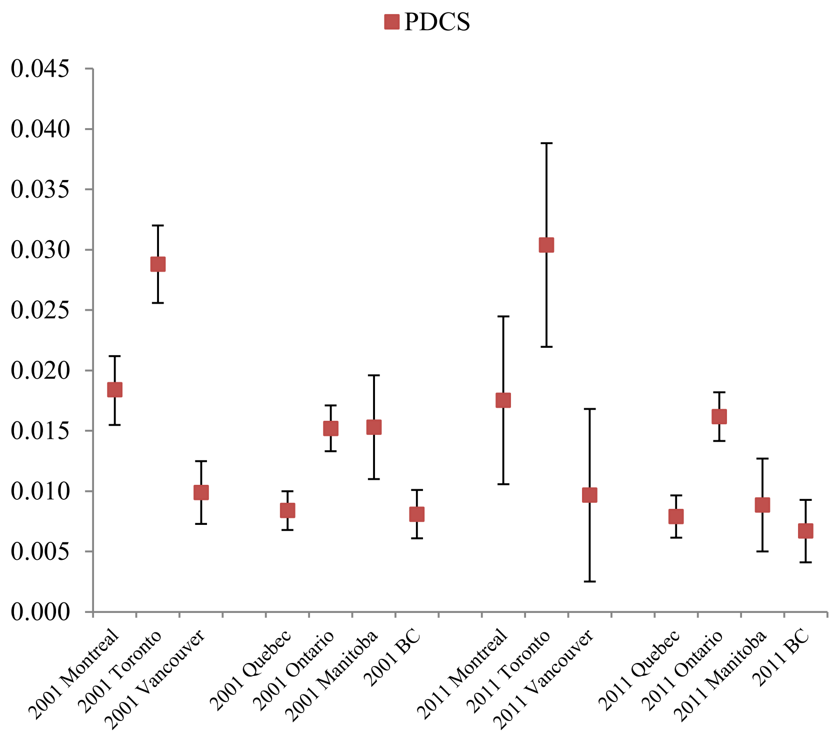

Figure 4.4.6. Comparisons between the 2001 and the 2011 estimates of the proportions of the adult Jewish population generated with the PDCS method.

Assessment of the Bayesian Multilevel Regression with Post-Stratification (BMRP) Method. This section provides results of the comparisons between census estimates and BMRP estimates of the adult Jewish population.

Similar to the previous sections, Tables 4.4.5 and 4.4.6 display the comparisons between BMRP cross-survey estimates and census estimates. Table 4.4.5 shows data from the 2001 census and estimates based on the Ca2001 batch, while Table 4.4.6 presents data from the 2011 NHS and estimates based on the Ca2011 batch. Along with the estimates of the proportions of the adult Jewish population generated with the BMRP method, the tables display the 95% credible intervals (CI_b). Credible intervals constitute the Bayesian analogue of confidence intervals (Gill, 2008) and are calculated from the posterior distribution.

Table 4.4.5

2001 Census Estimates and Ca2001 BMRP Estimates of the Proportions of the Adult Jewish Population

	2001 Census			Ca2001 BMRP			Err	PE
	Proportion	95% $^{a}LCI_b$	95% $^{a}UCI_b$	Proportion	95% $^{a}LCI_b$	95% $^{a}UCI_b$		
Metropolitan Area								
Montreal	0.0268	0.0256	0.0281	0.0268	0.0230	0.0310	< 0.001	0
Toronto	0.0355	0.0343	0.0368	0.0355	0.0320	0.0394	< 0.001	0
Vancouver	0.0098	0.0088	0.0108	0.0109	0.0083	0.0138	0.001	11
Province								
Quebec	0.0130	0.0124	0.0136	0.0074	0.0046	0.0110	-0.006	-43
Ontario	0.0172	0.0166	0.0177	0.0145	0.0095	0.0211	-0.003	-15
Manitoba	0.0132	0.0117	0.0149	0.0148	0.0087	0.0230	0.002	12
BC	0.0061	0.0055	0.0067	0.0087	0.0054	0.0134	0.003	44

Note: [a] CIb refers to credible intervals, the Bayesian analogue of confidence intervals.

As can be seen in Table 4.4.5, absolute percentage errors are low for all metropolitan areas (0% for Toronto, 1% for Montreal, and 11% for Vancouver). In addition, census estimates are within the 95% CI for all metropolitan areas. APEs for the provinces range from 9% for Manitoba, to 44% for British Columbia. Two out of the four provincial estimates are within the 95% CI of the Ca2001 BMRP estimates (Ontario and Manitoba). BMRP estimates are higher than census estimates for Vancouver, Manitoba, and British Columbia, which are comparatively less populated areas. Figure

4.4.7 below displays the Ca2001 BMRP estimates and credible intervals along with the census values.

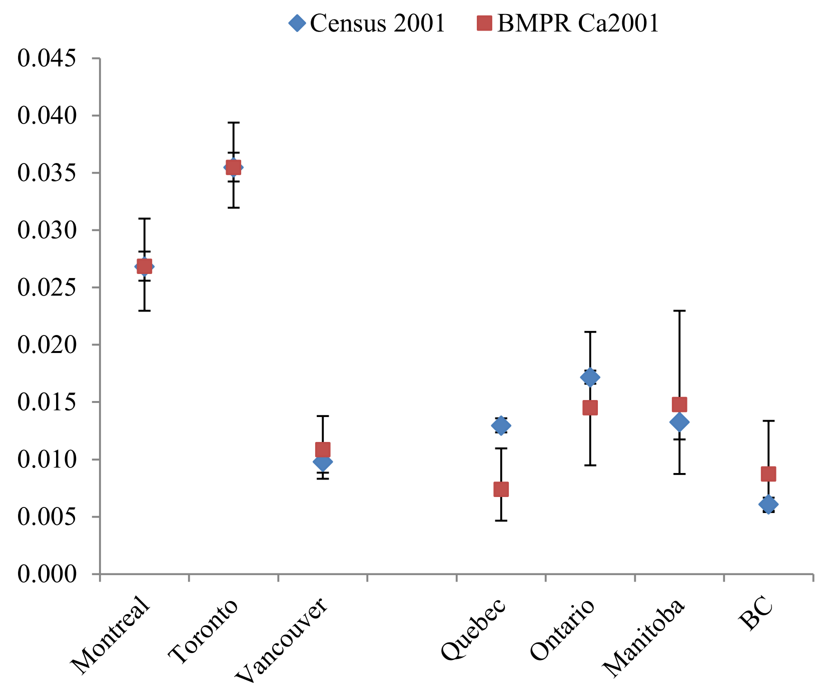

Figure 4.4.7. Comparisons between estimates of the proportions of the adult Jewish population generated with the BMRP method and 2001 census estimates.

Ca2011 BMRP estimates are higher than 2011 NHS estimates only for the Province of British Columbia. Metropolitan area APEs range from 13 to 35% (for Toronto and Montreal, respectively). The province APE range is similar (1% for Manitoba to 42% for Quebec), except for the British Columbia APE, which was very high: 153%. NHS estimates of the proportions of the adult Jewish population for Montreal, Toronto, Vancouver, Ontario and Manitoba are within the 95% credible intervals of the Ca2011 BMRP cross-survey estimates. These results are shown in Table 4.4.6 and Figure 4.4.8.

Table 4.4.6

2011 NHS Estimates and Ca2011 BMRP Estimates of the Proportions of the Adult Jewish Population

	2011 Census			Ca2011 BMRP			Err	PE
	Proportion	95% [a]LCI$_b$	95% [a]UCI$_b$	Proportion	95% [a]LCI$_b$	95% [a]UCI$_b$		
Metropolitan Area								
Montreal	0.0213	0.0202	0.0225	0.0139	0.0073	0.0222	-0.007	-35
Toronto	0.0298	0.0287	0.0309	0.0259	0.0141	0.0420	-0.004	-13
Vancouver	0.0079	0.0070	0.0089	0.0060	0.0019	0.0135	-0.002	-25
Province								
Quebec	0.0103	0.0098	0.0109	0.0060	0.0034	0.0092	-0.004	-42
Ontario	0.0147	0.0142	0.0152	0.0123	0.0073	0.0189	-0.002	-16
Manitoba	0.0106	0.0093	0.0121	0.0105	0.0063	0.0162	0.000	-1
BC	0.0048	0.0043	0.0053	0.0121	0.0072	0.0185	0.007	153

Note: [a] CI$_b$, refers to credible intervals, the Bayesian analogue of confidence intervals.

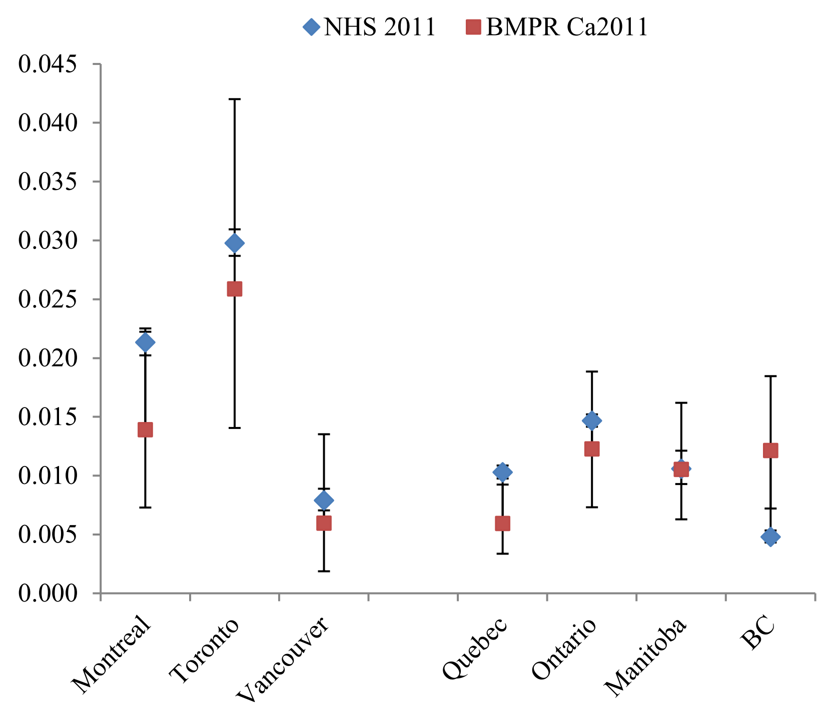

Figure 4.4.8. Comparisons between estimates of the proportions of the adult Jewish population generated with the BMRP method and 2011 NHS estimates.

Figure 4.4.9 below shows credible intervals for the four sets of estimates. Credible intervals are lower for the BMRP Ca2001 than for the BMRP Ca2011 estimates, and the differences are more pronounced for metropolitan area estimates.

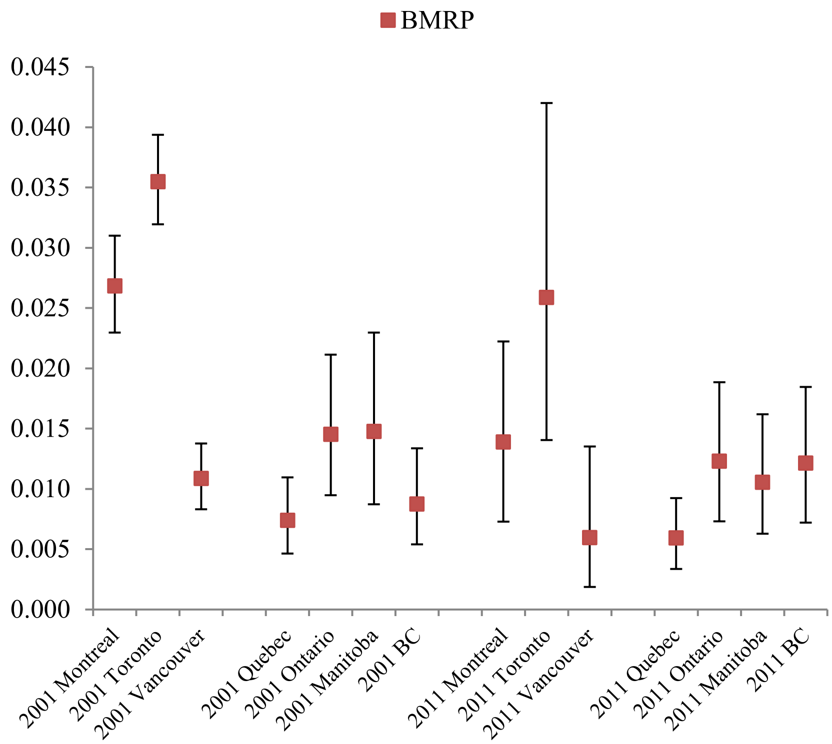

Figure 4.4.9. Comparisons between the 2001 and the 2011 estimates of the proportions of the adult Jewish population generated with the BMRP method.

Research Question 2: How do the results of the cross-survey approaches compare in terms of their accuracy and precision in estimating the proportions of the total adult Jewish population in the Canadian provinces and metropolitan areas included in the study?

The analyses I conducted to answer this question aimed to assess how the cross-survey methods compare with each other. To this end, I compared the statistical indicators calculated for the population estimates generated by the three methods based on data from both Canadian batches, Ca2001 and Ca2011. First, I compared Err, PE and CV as well as MAPE, MALPE, and RMSE values for the estimates obtained with each batch and method, and then I examined the confidence intervals and credible intervals of the estimates generated by the three methods. Results are presented separately for each time period by province and by metropolitan area. Comparing results for provinces and metropolitan areas separately makes it possible to take into account the number of surveys and sample sizes included in each analysis.

Assessment of Cross-Survey Estimates of the Adult Jewish Population in Metropolitan Areas Obtained with Data from Ca2001

Table 4.5.1 presents the comparisons between the 2001 census estimates and the cross-survey estimates of the adult Jewish population in the metropolitan areas of Montreal, Toronto, and Vancouver generated with data from the Ca2001 batch. This table shows the census estimates and the cross-survey estimates, as well as the number of surveys and sample sizes considered in each analysis.

Results presented in Table 4.5.1 show that the BMRP cross-survey method outperforms the MACS and PDCS methods in accuracy, although the difference is smaller in the second case. The MACS absolute percentage errors (APE) range from 28% in Toronto to 62% in Montreal; the PDCS APEs range from 1% in Vancouver to 31% in Montreal; and the BMRP APEs range from less than 1% in Montreal and Toronto to 11% in Vancouver. As the coefficients of variation (CV) indicate, the levels of dispersion of the estimates obtained with the three cross-survey estimates are low. The

CVs range from 5% (for the estimates of the adult Jewish population in Montreal) to 13% -15% (for the estimates of the adult Jewish population in Vancouver).

Table 4.5.1

Comparisons between Estimates of the Proportions of the Adult Jewish Population in Metropolitan Areas based on 2001 Census Estimates and Ca2001 Cross-Survey Estimates (MACS, PDCS, BMRP) – Err, PE, SE, and CV Indicators

	2001 Census Estimate	Surveys	Sample	Cross-survey Estimate	Err	PE	SE	CV[a]
MACS								
Montreal	0.0268	44	9537	0.0101	-0.017	-62	0.0011	11.3
Toronto	0.0355	44	12613	0.0257	-0.010	-28	0.0014	5.5
Vancouver	0.0098	44	6042	0.0048	-0.005	-51	0.0007	14.7
PDCS								
Montreal	0.0268	44	9537	0.0184	-0.008	-31	0.0015	7.9
Toronto	0.0355	44	12613	0.0288	-0.007	-19	0.0016	5.7
Vancouver	0.0098	44	6042	0.0099	0.000	1	0.0013	13.3
BMRP								
Montreal	0.0268	44	9537	0.0268	< 0.001	0	0.0001	7.7
Toronto	0.0355	44	12613	0.0355	< 0.001	0	0.0001	5.3
Vancouver	0.0098	44	6042	0.0109	-0.001	11	< 0.0001	13.2

Note: [a] CV, the coefficient of variation is calculated as the SE over the mean for estimates generated with the MACS and PDCS methods and as SD over the mean for estimates obtained with the BMRP method.

Both the BMRP and the PDCS methods produce accurate estimates for Vancouver, but only BMRP generates accurate estimates for all three metropolitan areas. MACS and PDCS underestimate the proportions of the adult Jewish population (as measured by the 2001 census). These results are reinforced by the analyses of precision (see Figure 4.5.1) and analyses of the aggregated indicators (see Table 4.5.2).

Table 4.5.2 displays the MAPE, MALPE, and RMSE aggregated indicators calculated for the three metropolitan areas with data from the Ca2001 batch. These indicators provide summary measures of the accuracy of each method (Yowell & Devine, 2013). As was the case with the results presented in Table 4.5.1, the aggregated measures show that estimates obtained with the BMRP method are far more accurate than the ones generated with PDCS and MACS. BMRP has a MAPE of 4%, compared to 17% for PDCS, and 47% for MACS.

Table 4.5.2

Comparison between Estimates of the Proportions of the Adult Jewish Population in Metropolitan Areas Based on 2001 Census Estimates and Ca2001 Cross-Survey Estimates (MACS, PDCS, BMRP) – MAPE, MALPE, and RMSE Indicators

	Ca2001 Metropolitan Area		
	MACS	PDCS	BMRP
MAPE	47.07	17.14	3.68
MALPE	-47.07	-16.41	3.66
RMSE	1.05	0.51	0.04

To further assess the precision and accuracy of the different methods, I compared the 95% confidence intervals of the estimates (95% credible intervals for the BMRP method). Figure 4.5.1 shows the results.

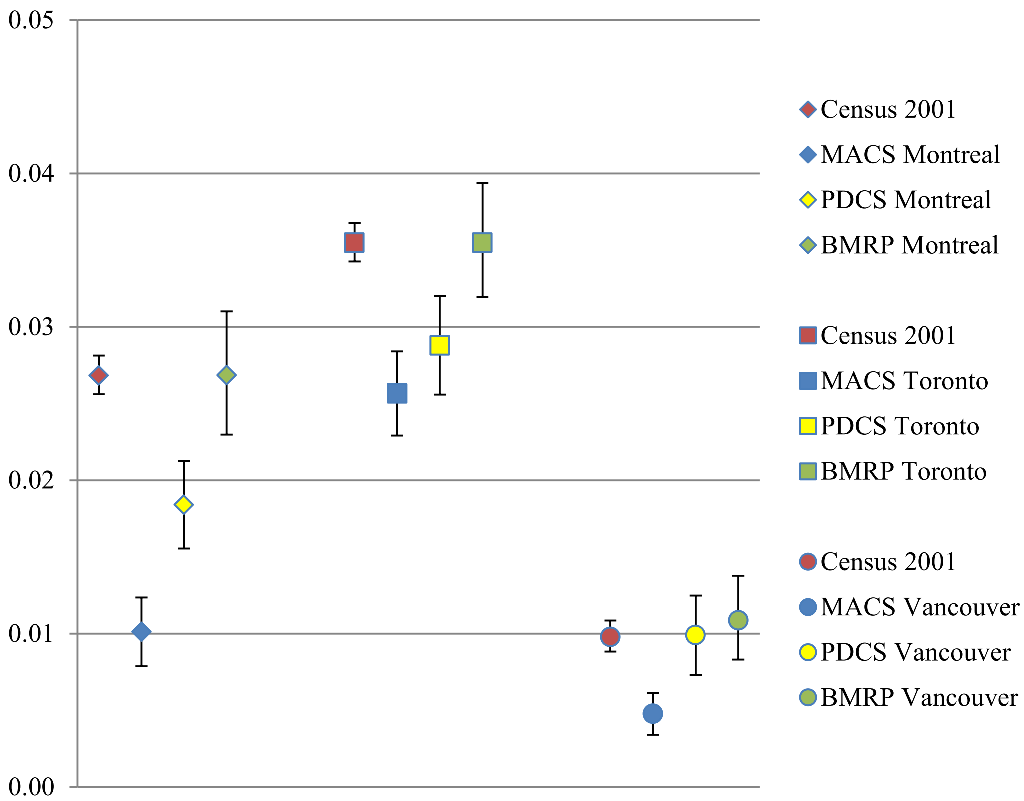

Figure 4.5.1. Comparisons between Ca2001 cross-survey estimates across methods and 2001 census estimates of the proportions of the adult Jewish population in metropolitan areas.

As noted earlier, the census estimates are within the 95% credible intervals for all three Ca2001 BMRP estimates of the adult Jewish population in Montreal Toronto and Vancouver. The only other census estimate within cross-survey confidence intervals is the census estimate for Vancouver which is within the 95% confidence intervals of the PDCS estimate. The 95% confidence and credible intervals are relatively small and of comparable size for all three metropolitan areas.

Assessment of Cross-Survey Estimates of the Adult Jewish Population in the Canadian Provinces with Data from Ca2001

This section presents comparisons of the estimates of the proportions of the adult Jewish population in the Canadian provinces generated with data from the Ca2001 batch.

Table 4.5.3 shows cross-survey estimates as well as individual statistical indicators; Table 4.5.3 displays aggregated indicators; and Figure 4.5.2 compares confidence intervals.

Table 4.5.3

Comparisons between Estimates of the Proportions of the Adult Jewish Population in the Four Provinces Based on 2001 Census Estimates and Ca2001 Cross-Survey Estimates (MACS, PDCS, BMRP) – Err, PE, SE, and CV Indicators

	2001 Census Estimate	Surveys	Sample	Cross-Survey Estimate	Err	PE	SE	CV[a]
MACS								
Quebec	0.0130	43	16325	0.0054	-0.008	-58	0.0010	19.3
Ontario	0.0172	43	19148	0.0114	-0.006	-34	0.0012	10.3
Manitoba	0.0132	43	3577	0.0105	-0.003	-20	0.0015	13.8
BC	0.0061	43	7764	0.0074	0.001	22	0.0014	19.5
PDCS								
Quebec	0.0130	43	16325	0.0084	-0.005	-35	0.0008	9.9
Ontario	0.0172	43	19148	0.0152	-0.002	-11	0.0010	6.4
Manitoba	0.0132	43	3577	0.0153	0.002	16	0.0022	14.3
BC	0.0061	43	7764	0.0081	0.002	34	0.0010	12.5
BMRP								
Quebec	0.0130	43	16325	0.0074	-0.006	-43	0.0001	21.9
Ontario	0.0172	43	19148	0.0145	-0.003	-15	0.0001	20.2
Manitoba	0.0132	43	3577	0.0148	0.002	12	0.0001	25.1
BC	0.0061	43	7764	0.0087	0.003	44	0.0001	23.2

Note: [a] CV, the coefficient of variation, is calculated as the SE over the mean for estimates generated with the MACS and PDCS methods and as SD over the mean for estimates obtained with the BMRP method.

Census estimates for the provinces of Ontario, Manitoba, and British Columbia are within the 95% confidence intervals of the PDCS estimates and within the 95% credible intervals of the BMRP estimates. Census estimates for the provinces of Manitoba and British Columbia are within the 95% confidence intervals of the MACS estimates. Yet the three methods failed to accurately estimate the proportions of the adult Jewish population in the province of Quebec. The three methods underestimated the proportions of the adult Jewish population in the provinces of Quebec and Ontario and overestimated the proportion of this population in the province of British Columbia.

Unlike the previous comparison, this one does not suggest a clear trend in the performances of the three methods in general, PDCS appears to outperform the others slightly in accuracy and precision (as indicated by the Err, APE and CV indicators, and by the aggregated indicators as well). The absolute percentage errors (APE) for all the estimates are larger than 10%. APEs range from 20% in Manitoba to 58% in Quebec for MACS; from 11% in Ontario to 35% in Quebec for PDCS; and from 12% in Manitoba to 44% in Quebec for BMRP. Although not as small as the ones found for the Ca2001 metropolitan area estimates, the coefficients of variation suggest low and moderate amount of variability. The variability of the PDCS method is slightly lower than that of the MACS and BMRP methods. PDCS CVs are good, ranging from 6% to 15%, while CVs for MACS and BMRP estimates are fair, ranging from 14% to 25%. The mean absolute percentage error (MAPE) associated with the PDCS method is the smallest (PDCS has a MAPE of 24%, compared to 27% for BMRP and 33% for MACS). (See Table 4.5.4).

Table 4.5.4

Comparisons between Estimates of the Proportions of the Adult Jewish Population in Provinces Based on 2001 Census Estimates and Ca2001 Cross-Survey Estimates (MACS, PDCS, BMRP) – MAPE, MALPE, and RMSE Indicators

	Ca2001 Province		
	MACS	PDCS	BMRP
MAPE	33.64	23.96	28.49
MALPE	-22.47	0.71	-0.70
RMSE	0.43	0.27	0.31

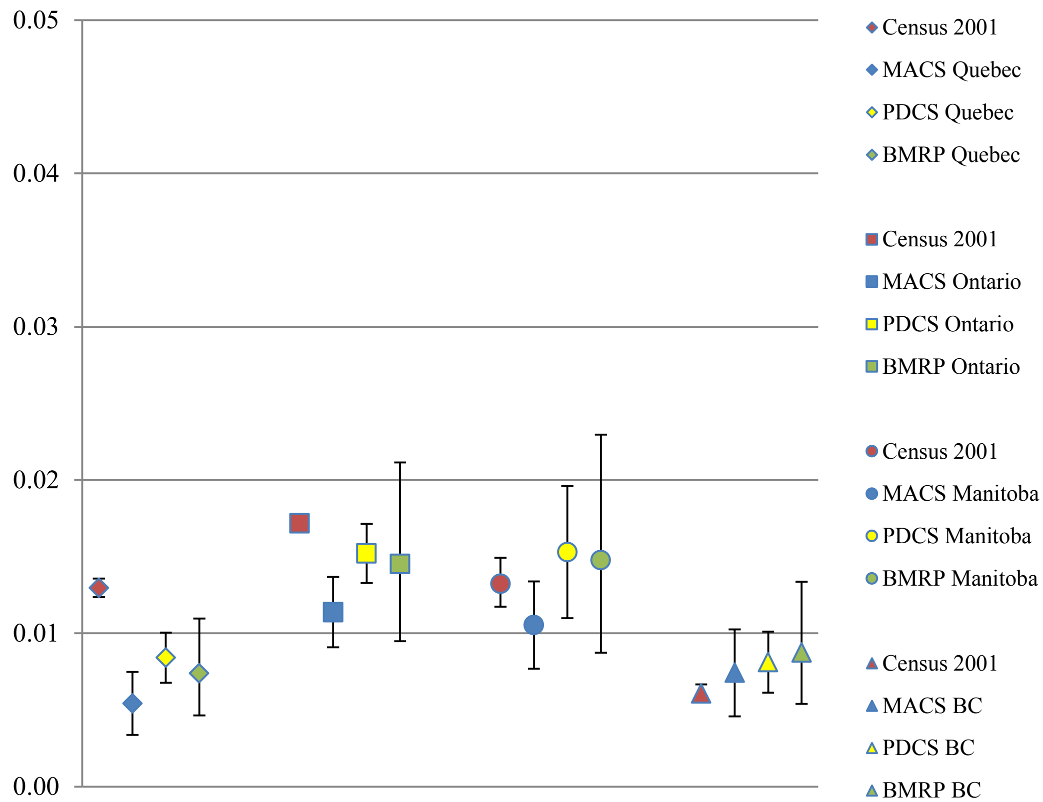

Figure 4.5.2 Comparisons between Ca2001 cross-survey estimates across methods and 2001 census estimates of the proportions of the adult Jewish population in the provinces.

Assessment of Cross-Survey Estimates of the Adult Jewish Population in the Canadian Metropolitan Areas with Data from Ca2011

This section presents the estimates of the proportions of the adult Jewish population in the Canadian metropolitan areas calculated with data from the Ca2011 batch as they compare to the 2011 NHS estimates. The following tables show cross-survey estimates (Table 4.5.5), individual statistical indicators (Table 4.5.5), and aggregated indicators (Table 4.5.6).

Table 4.5.5

Comparisons between Estimates of the Proportions of the Adult Jewish Population in Metropolitan Areas Based on 2011 NHS Data and Ca2011 Cross-Survey Estimates (MACS, PDCS, BMRP) – Err, PE, SE, and CV Indicators

	2011 NHS Estimate	Surveys	Sample	Cross-Survey Estimate	Err	PE	SE	CV[a]
MACS								
Montreal	0.0213	10	1638	0.0146	-0.007	-32	0.0076	52.0
Toronto	0.0298	10	1886	0.0243	-0.005	-18	0.0076	31.4
Vancouver	0.0079	10	876	0.0213	0.013	169	0.0131	61.6
PDCS								
Montreal	0.0213	10	1638	0.0175	-0.004	-18	0.0035	20.2
Toronto	0.0298	10	1886	0.0304	0.001	2	0.0043	14.1
Vancouver	0.0079	10	876	0.0097	0.002	22	0.0036	37.6
BMRP								
Montreal	0.0213	10	1638	0.0139	-0.007	-35	0.0001	29.2
Toronto	0.0298	10	1886	0.0259	-0.004	-13	0.0002	28.4
Vancouver	0.0079	10	876	0.0060	-0.002	-25	0.0001	50.7

Note: [a] CV, the coefficient of variation, is calculated as the SE over the mean for estimates generated with the MACS and PDCS methods and as SD over the mean for estimates obtained with the BMRP method.

Among the three methods, PDCS generated more accurate estimates for Montreal (MACS APE=32%, PDCS APE=18%, BMRP APE=35%), Toronto (MACS APE=18%, PDCS APE =2%, BMRP APE=13%), and Vancouver (MACS APE=169%, PDCS APE=22%, BMRP APE=25%). The particularly high MACS percentage error (PE) for Vancouver seems to indicate an outlier. As previously discussed, the relatively large number of surveys with no Jewish population (6 out of 10) might have influenced the overestimation observed here. The coefficients of variation suggest relatively unstable estimates, especially for the three MACS estimates (Montreal CV=52%, Toronto CV=31%, and Vancouver CV=32%) and for the Vancouver estimates generated with the MACS and BMRP methods (62% and 51% respectively).

Table 4.5.6

Comparisons between Estimates of the Proportions of the Adult Jewish Population in Metropolitan Areas Based on 2011 NHS Estimates and Ca2011 Cross-Survey Estimates (MACS, PDCS, BMRP) – MAPE, MALPE, and RMSE Indicators

	Ca2011 Metropolitan Area		
	MACS	PDCS	BMRP
MAPE	73.07	14.06	24.20
MALPE	39.66	2.11	-24.20
RMSE	0.85	0.21	0.44

The comparisons among MAPE indicators shows that the range of errors is smallest for the PDCS method. Still, none of the mean average percentage errors (MAPE) is smaller than 10%.

Figure 4.5.3 graphically compares the accuracy and precision of the metropolitan area NHS estimates with the accuracy and precision of the estimates obtained with the different cross-survey methods.

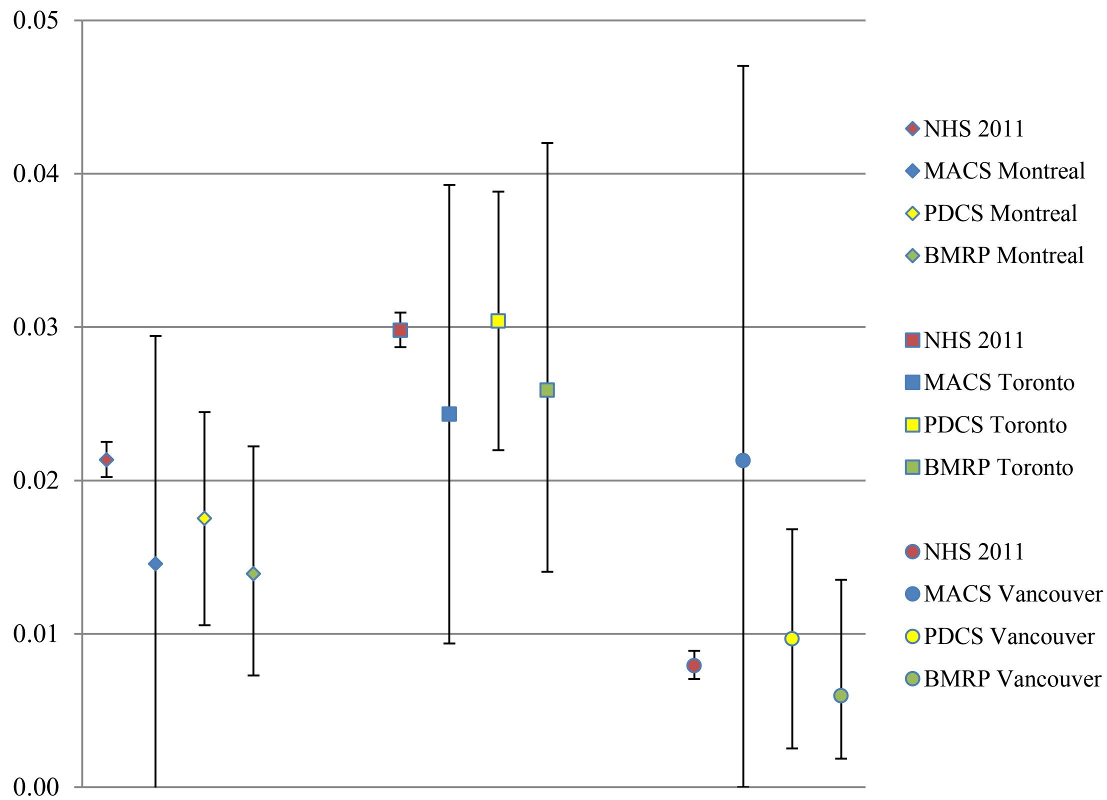

Figure 4.5.3. Comparisons between Ca2011 cross-survey estimates across methods and 2011 NHS estimates of the proportions of the adult Jewish population in metropolitan areas.

As Figure 4.5.3 shows, the 95% confidence intervals (credible intervals for the BMRP method) are large for all the estimates but especially so for the Vancouver MACS estimates. The NHS estimates of the proportions of the adult Jewish population in the three largest metropolitan areas – Montreal, Toronto, and Vancouver- are within the 95% confidence intervals of the cross-survey estimates generated with the three cross-survey methods. Nonetheless, given the wide CIs, interpretations of the accuracy of these estimates should be cautious.

Assessment of Cross-Survey Estimates of the Adult Jewish Population in the Canadian Provinces Obtained with Data from Ca2011

This section presents the comparisons of the estimates of the adult Jewish population in the four Canadian provinces generated with data from the Ca2011 batch. Table 4.5.7 shows the cross-survey estimates and individual statistical indicators, and Table 4.5.8 displays the aggregated indicators.

Table 4.5.7

Comparisons between Estimates of the Proportions of the Adult Jewish Population in the Provinces Based on 2011 NHS Estimates and Ca2011 Cross-Survey Estimates (MACS, PDCS, BMRP) – Err, PE, SE, and CV Indicators

	2011 NHS Estimate	Surveys	Sample	Cross-Survey Estimate	Err	PE	SE	CV[a]
MACS								
Quebec	0.0103	17	13890	0.0078	-0.003	-24	0.0026	33.2
Ontario	0.0147	17	19230	0.0160	0.001	9	0.0023	14.5
Manitoba	0.0106	17	3700	0.0061	-0.004	-42	0.0029	47.5
BC	0.0048	17	8004	0.0071	0.002	47	0.0036	51.6
PDCS								
Quebec	0.0103	17	13890	0.0079	-0.002	-23	0.0009	11.3
Ontario	0.0147	17	19230	0.0162	0.002	10	0.0010	6.4
Manitoba	0.0106	17	3700	0.0089	-0.002	-17	0.0020	22.2
BC	0.0048	17	8004	0.0067	0.002	39	0.0013	19.7
BMRP								
Quebec	0.0103	17	13890	0.0060	-0.004	-42	< 0.0001	25.5
Ontario	0.0147	17	19230	0.0123	-0.002	-16	0.0001	23.7
Manitoba	0.0106	17	3700	0.0105	0.000	-1	0.0001	23.8
BC	0.0048	17	8004	0.0121	0.007	153	0.0001	23.7

Note: [a] CV, the coefficient of variation, is calculated as the SE over the mean for estimates generated with the MACS and PDCS methods and as SD over the mean for estimates obtained with the BMRP method.

Among the three methods, PDCS and MACS methods generated slightly more accurate estimates for Quebec (MACS APE=24%, PDCS APE=23%, BMRP APE=42%) and Ontario (MACS APE=9%, PDCS APE =10%, BMRP APE=16%). The BMRP method generated the most accurate estimates for Manitoba (MACS APE=42%, PDCS APE=17%, BMRP APE=1%). All three cross-survey methods underestimated the proportion of the adult Jewish population in Quebec and underestimated the proportions of the Jewish population in British Columbia. MACS and PDCS slightly overestimated the proportion of this population in Ontario, while BMRP slightly underestimates it. The particularly high BMRP proportional error for British Columbia seems to indicate an outlier, and may be due to the influence of the small sample sizes and geographical variables included in the model. More research is needed to understand the factors behind this much-higher-than-expected estimate. The sampling error of the estimates, as indicated by the coefficients of variation, is smaller for the PDCS estimates (ranging from 6 to 22%). The CVs for the BMRP method are fair, ranging from 23 to 26%. The CVs for the MACS estimates fluctuates.

Table 4.5.8

Comparisons between Estimates of the Proportions of the Adult Jewish Population in the Provinces Based on 2011 NHS Estimates and Ca2011 Cross-Survey Estimates (MACS, PDCS, BMRP) – MAPE, MALPE, and RMSE Indicators

	Ca2011 Province		
	MACS	PDCS	BMRP
MAPE	30.70	22.39	52.96
MALPE	-2.55	2.45	23.44
RMSE	0.26	0.19	0.35

Figure 4.5.4 graphically shows the comparisons between the NHS estimates of the adult Jewish population in the Canadian provinces with the cross-survey estimates of the same population.

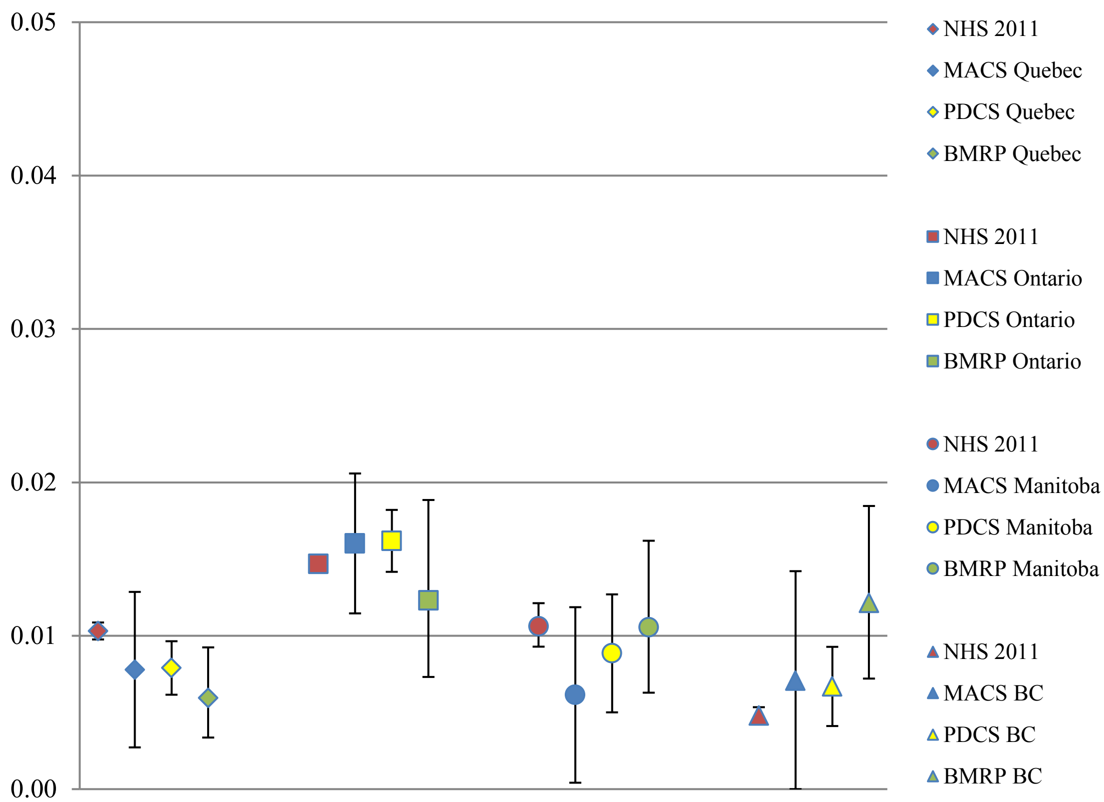

Figure 4.5.4. Comparisons between Ca2011 cross-survey estimates across cross-survey methods and 2011 NHS estimates of the proportions of the adult Jewish population in provinces.

As evidenced by the size of the confidence intervals in Figure 4.5.4, the Ca2011 provincial estimates generated with the PDCS method are the most precise. The NHS estimates of the proportions of the adult Jewish population are within the 95% confidence intervals of the cross-survey estimates generated for the four provinces with the three cross-survey methods.

Research Question 3: How do the three cross-survey approaches compare in terms of their stability and precision for estimating the proportions of the total adult Jewish population in metropolitan areas in the continental US?

To answer this question, I conducted two sets of analyses. In the first set, I compared the estimates of the proportions of the adult Jewish population calculated with the three cross-survey methods based on data from sub-batches US2011a and US2011b. This comparison was meant to assess the stability of the methods. First, I compared the Err, PE, and APE indicators calculated for each method, and then I examined the MAPE and MALPE measures across methods. Err corresponds to the differences between the estimates of the two US2011 sub-batches by metropolitan area (US2011b-US2011a), and PE, to the differences between the estimates of the two US2011 sub-batches by metropolitan area as a proportion of the US2011a sub-batch [(US2011b-US2011a)/ US2011a]. APE is the absolute value of PE. MAPE is calculated for each cross-survey method as the average APE across metropolitan areas, and MALPE, as the average PE. In the second set of analyses I compared the estimates of the proportions of the Jewish population generated by the three cross-survey methods based on data from the US2011 batch.

Meta-Analysis of Complex Surveys - Assessment of Stability. The comparisons between the estimates of the adult Jewish population in the 20 largest U.S. metropolitan areas generated with the MACS cross-survey method for the two US2011 sub-batches indicates relatively low stability for many of the estimates. The results of this analysis are presented in Table 4.6.1. In addition to the estimates of the adult Jewish population by metropolitan area for the two U.S. sub-batches, this table shows Err and PE indicators.

These indicators represent the differences between the estimates obtained with data from the two sub-batches. The percentage errors (PE) range from low (5%, for metro5) to very high (204%, for metro8). Forty-five percent of the PEs are larger than 50%, and only 30% are smaller than 15%. In 65% of the comparisons the US2011a estimates are larger than US2011b estimates.

Table 4.6.1

US2011a and US2011b MACS Estimates of the Proportions of the Adult Jewish Population

	US2011a			US2011b			Err	PE
Metro Area	Proportion	95% LCI	95% UCI	Proportion	95% LCI	95% UCI		
1	0.063	0.060	0.066	0.072	0.069	0.076	0.009	14
2	0.012	0.011	0.014	0.020	0.017	0.022	0.007	58
3	0.017	0.015	0.020	0.005	0.004	0.006	-0.013	-73
4	0.012	0.010	0.015	0.010	0.009	0.012	-0.002	-17
5	0.009	0.007	0.011	0.009	0.007	0.010	0.000	-5
6	0.019	0.017	0.022	0.021	0.019	0.024	0.002	11
7	0.020	0.017	0.023	0.012	0.010	0.014	-0.009	-43
8	0.023	0.020	0.025	0.069	0.065	0.073	0.046	204
9	0.009	0.007	0.011	0.002	0.002	0.003	-0.007	-75
10	0.011	0.009	0.013	0.023	0.020	0.026	0.012	105
11	0.016	0.013	0.018	0.021	0.018	0.025	0.006	37
12	0.019	0.016	0.021	0.013	0.010	0.015	-0.006	-33
13	0.012	0.011	0.014	0.010	0.008	0.011	-0.003	-21
14	0.009	0.007	0.011	0.005	0.004	0.005	-0.005	-50
15	0.019	0.017	0.021	0.017	0.014	0.020	-0.002	-11
16	0.017	0.014	0.020	0.015	0.013	0.018	-0.001	-8
17	0.020	0.017	0.023	0.008	0.007	0.009	-0.012	-61
18	0.012	0.010	0.014	0.004	0.003	0.005	-0.008	-68
19	0.030	0.025	0.034	0.007	0.006	0.009	-0.022	-75
20	0.019	0.016	0.023	0.022	0.020	0.024	0.003	14

Pooled Design-Based Cross-Survey - Assessment of Stability. The comparisons between the estimates of the adult Jewish population in the 20 largest U.S. metropolitan areas generated with the PDCS cross-survey method for the two U.S. sub-batches indicates relatively more stability than the comparison between the MACS estimates. The percentage error indicators range from no-error (0%) for metro6, to 51% for metro13. Twenty percent of the PEs are less than 10%, and an additional 40% are less than 20%. Moreover, only three metropolitan areas have a PE greater than 45%. The US2011a estimates are larger than the US2011b estimates in 55% of the metropolitan areas. The results of these comparisons are presented in Table 4.6.2.

Table 4.6.2

US2011a and US2011b PDCS Estimates of the Proportions of the Adult Jewish Population

Metro Area	US2011a			US2011b			Err	PE
	Proportion	95% LCI	95% UCI	Proportion	95% LCI	95% UCI		
metro1	0.078	0.070	0.087	0.091	0.082	0.100	0.013	16
metro2	0.040	0.033	0.049	0.034	0.028	0.042	-0.006	-15
metro3	0.031	0.025	0.040	0.026	0.020	0.034	-0.005	-16
metro4	0.007	0.003	0.016	0.011	0.006	0.019	0.003	46
metro5	0.010	0.005	0.018	0.006	0.003	0.012	-0.003	-35
metro6	0.040	0.031	0.050	0.040	0.032	0.050	0.000	0
metro7	0.039	0.030	0.050	0.036	0.027	0.047	-0.003	-7
metro8	0.077	0.063	0.094	0.100	0.084	0.118	0.022	29
metro9	0.016	0.010	0.024	0.016	0.010	0.025	0.000	2
metro10	0.034	0.024	0.049	0.049	0.038	0.062	0.014	41
metro11	0.040	0.028	0.056	0.044	0.032	0.061	0.004	11
metro12	0.015	0.009	0.024	0.016	0.010	0.024	0.001	6
metro13	0.012	0.007	0.021	0.006	0.003	0.012	-0.006	-51
metro14	0.010	0.006	0.016	0.011	0.007	0.019	0.002	17
metro15	0.015	0.009	0.026	0.014	0.008	0.023	-0.001	-10
metro16	0.012	0.008	0.020	0.011	0.007	0.017	-0.001	-11
metro17	0.020	0.011	0.036	0.014	0.008	0.025	-0.006	-30
metro18	0.026	0.017	0.041	0.019	0.012	0.028	-0.008	-29
metro19	0.018	0.010	0.033	0.016	0.008	0.032	-0.002	-11
metro20	0.027	0.017	0.042	0.014	0.008	0.025	-0.012	-45

Bayesian Multilevel Regression with Post-Stratification - Assessment of Stability.

Comparison between the estimates of the adult Jewish population in the 20 largest U.S.

metropolitan areas generated with the BMRP cross-survey method for the two U.S. sub-batches indicates relative stability between the estimates. Forty percent of the estimates have a percentage error of less than 10 percent, and an additional 45% have a percentage error of less than 20%. All percentage errors but one are less than 31%, the exception being the estimates for metro10, which exhibit an error of 71%. US2011a estimates are larger than US2011b estimates only in 30% of the metropolitan areas. Table 4.6.3 shows the results of the comparisons.

Table 4.6.3

US2011a and US2011b BMRP Estimates of the Proportions of the Adult Jewish Population

Metro Area	US2011a			US2011b			Err	PE
	Proportion	95% LCI	95% UCI	Proportion	95% LCI	95% UCI		
metro1	0.064	0.058	0.071	0.077	0.070	0.083	0.012	19
metro2	0.030	0.025	0.035	0.029	0.025	0.034	-0.001	-3
metro3	0.029	0.024	0.034	0.027	0.022	0.033	-0.002	-6
metro4	0.007	0.005	0.009	0.008	0.005	0.010	0.001	16
metro5	0.006	0.004	0.009	0.006	0.004	0.008	0.000	-2
metro6	0.029	0.024	0.034	0.032	0.028	0.039	0.004	13
metro7	0.034	0.028	0.041	0.037	0.030	0.045	0.004	11
metro8	0.066	0.055	0.077	0.086	0.075	0.098	0.020	31
metro9	0.009	0.007	0.012	0.010	0.007	0.013	0.001	6
metro10	0.033	0.024	0.043	0.056	0.045	0.068	0.023	72
metro11	0.029	0.022	0.036	0.028	0.022	0.034	-0.001	-3
metro12	0.013	0.009	0.017	0.016	0.012	0.021	0.003	27
metro13	0.007	0.005	0.011	0.006	0.004	0.009	-0.001	-17
metro14	0.009	0.007	0.013	0.011	0.008	0.015	0.002	16
metro15	0.011	0.007	0.015	0.011	0.007	0.015	0.000	1
metro16	0.011	0.008	0.015	0.012	0.008	0.016	0.001	10
metro17	0.016	0.011	0.022	0.013	0.009	0.018	-0.003	-18
metro18	0.016	0.012	0.022	0.019	0.014	0.025	0.003	19
metro19	0.010	0.007	0.014	0.010	0.007	0.014	0.000	0
metro20	0.020	0.014	0.026	0.017	0.012	0.024	-0.002	-12

Assessment of Stability across the Three Cross-Survey Methods. To further assess the stability of the cross-survey methods, aggregated MAPE, MALPE, and RMSE indicators were calculated for each method. Table 4.6.4 displays these indicators for the three methods.

Table 4.6.4

Difference between US2011a and US2011b Estimates of the Adult Jewish Population across Metropolitan Areas

	MACS	PDCS	BMRP
MAPE	49.17	21.44	15.12
MALPE	-4.89	-4.71	9.04
RMSE	0.009	0.006	0.004

The MAPE and RMSE measures indicate that, on average, absolute differences between the estimates generated for US2011a and US2011b sub-batches across metropolitan areas are smallest for the BMRP method. MAPE was 15% for this method compared to 21% for the PDCS method and 49% for the MACS method, suggesting that BMRP results are more stable across different subsamples. Similarly, RMSE for BMRP is .0042, compared to .0057 for PDCS and .0088 for MACS, while MACS and PDCS MALPEs are similar and smaller than the BMRP MALPE. The MALPE indicator takes into account the direction of the error, so that negative errors are canceled by positive errors. The difference between MAPE and MALPE indicators gives an idea of the direction of the errors across metropolitan areas. MAPE is smaller than MALPE, for the three methods which suggests that errors are in different directions. In other words, neither sub-batch produces higher estimates.

Assessment of the Estimation of the US2011 Batch

In this section I present the estimates for the Jewish population in the 20 largest U.S. metropolitan areas generated with data from the entire US2011 batch. The aim is to identify and compare the precision of the estimates produced with the different cross-survey methods. In addition, and taking into consideration that the true value of the estimands is not known, the differences between the estimates obtained with the three methods were also assessed.

The dispersion and variability of the estimates of the adult Jewish population in the 20 largest U.S. metropolitan areas generated with the three methods were found to be similar. The coefficients of variation are slightly lower for the MACS model (MACS CVs range from 2 to 10%, PDCS CVs from 4 to 25%, and BMRP CVs from 3 to 15%). Table 4.6.5 presents this information.

Table 4.6.5 presents comparisons of the estimates of the Jewish population in the 20 largest U.S. metropolitan areas generated with the three cross- survey methods. Along with these estimates, it displays their corresponding standard errors and coefficients of variation.

Table 4.6.5

Comparisons between Estimates of the Proportions of the Adult Jewish Population in the 20 Largest U.S. Metropolitan Areas (MACS, PDCS, BMRP) – SE and CV Indicators

Metro Area	MACS			PDCS			BMRP		
	Estimate	SE	CV[a]	Estimate	SE	CV[a]	Estimate	SD	CV[a]
metro1	0.0672	0.0011	2	0.0847	0.0031	4	0.0707	0.0025	3
metro2	0.0122	0.0007	6	0.0370	0.0026	7	0.0301	0.0018	6
metro3	0.0051	0.0004	8	0.0286	0.0025	9	0.0274	0.0021	8
metro4	0.0101	0.0006	6	0.0089	0.0022	25	0.0069	0.0009	14
metro5	0.0087	0.0006	6	0.0079	0.0018	22	0.0061	0.0009	14
metro6	0.0198	0.0009	4	0.0398	0.0033	8	0.0316	0.0021	7
metro7	0.0125	0.0008	6	0.0375	0.0036	10	0.0358	0.0026	7
metro8	0.0304	0.0010	3	0.0890	0.0059	7	0.0761	0.0041	5
metro9	0.0022	0.0002	10	0.0156	0.0025	16	0.0105	0.0013	13
metro10	0.0138	0.0008	6	0.0417	0.0044	10	0.0427	0.0035	8
metro11	0.0153	0.0008	6	0.0419	0.0051	12	0.0279	0.0025	9
metro12	0.0123	0.0008	7	0.0154	0.0024	16	0.0144	0.0017	12
metro13	0.0098	0.0005	5	0.0087	0.0019	22	0.0071	0.0011	15
metro14	0.0044	0.0003	7	0.0104	0.0020	19	0.0107	0.0014	13
metro15	0.0167	0.0011	7	0.0144	0.0027	19	0.0107	0.0014	14
metro16	0.0157	0.0010	6	0.0116	0.0020	17	0.0115	0.0015	13
metro17	0.0079	0.0004	5	0.0169	0.0035	21	0.0151	0.0019	13
metro18	0.0039	0.0003	8	0.0220	0.0035	16	0.0184	0.0022	12
metro19	0.0072	0.0005	7	0.0169	0.0040	24	0.0104	0.0015	14
metro20	0.0189	0.0011	6	0.0204	0.0037	18	0.0186	0.0023	12

Note: [a] CV, the coefficient of variation, is calculated as the SE over the mean for estimates generated with the MACS and PDCS methods and as SD over the mean for estimates obtained with the BMRP method.

To calculate indices reflecting the differences among the estimates generated by the three methods, I calculated mean average percent errors (MAPE) between two cross-survey methods at a time. The value of the mean percentage error is dependent on which estimate is used as a reference. Consequently, for each pair of cross-survey methods, I calculated two indicators using these two methods alternatively. This analysis indicates that the closest estimates are the ones obtained with PDCS and BMRP (MAPEs between

PDCS and BMRP estimates are 16% and 21%, while MAPEs between MACS and any of the other two methods are at least 46%). Table 4.6.6 presents these results.

Table 4.6.6

Mean Absolute Percent Errors (MAPE) between the Cross-Survey Methods.

Cross Survey Methods	MAPE 1	MAPE 2
MACS vs PDCS	156	47
MACS vs BMRP	123	48
PDCS vs BMRP	16	21

Figures 4.6.1 and 4.6.2 show the estimates of the adult Jewish population for the 20 largest metropolitan areas generated with the three cross-survey methods based on data from the US2011 batch. Figure 4.6.1 displays the estimates for the first ten metropolitan areas, and Figure 4.6.2, for the remaining ten. Along with these estimates, the figures show the 95% confidence intervals.

In 65% of the metropolitan areas (areas 1, 2, 3, 6, 7, 8, 9, 11, 12, 17, 18, 19, and 20) the PDCS method produced the highest estimates, followed by BMRP. MACS estimates, in turn, are highest for 25% of the metropolitan areas (areas 4, 5, 13, 15, and 16). These five metropolitan areas have some of the smallest Jewish population estimates. The BMRP estimate is the highest only for one metropolitan area (metro10). In terms of precision, confidence intervals are smallest for MACS estimates (for all metropolitan areas), followed by BMRP estimates.

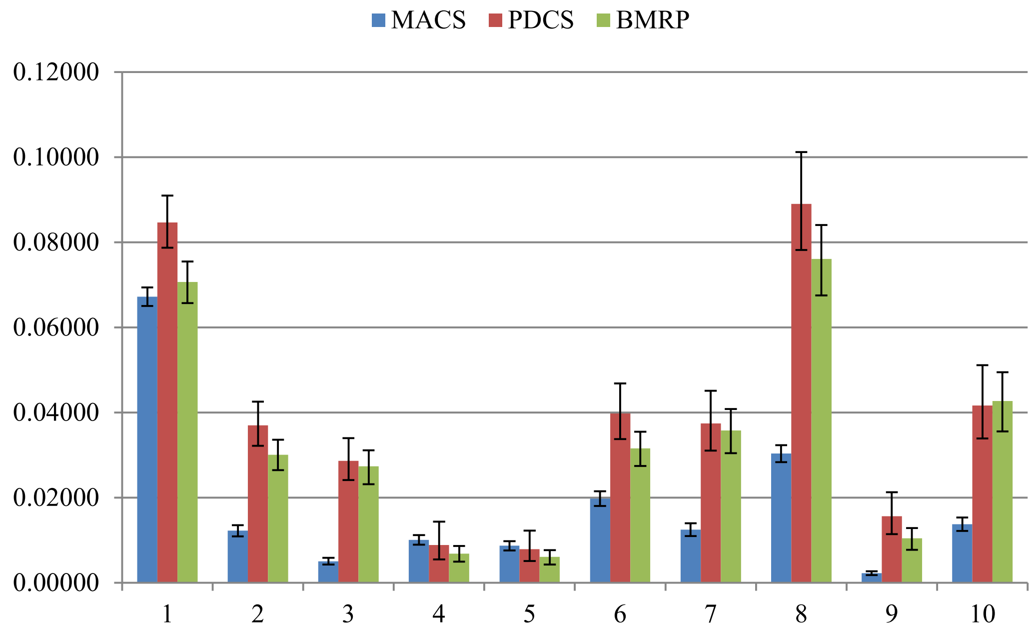

Figure 4.6.1. Estimates of the proportions of the adult Jewish population for the ten largest metropolitan areas generated with the three cross-survey methods based on data from the US2011 batch.

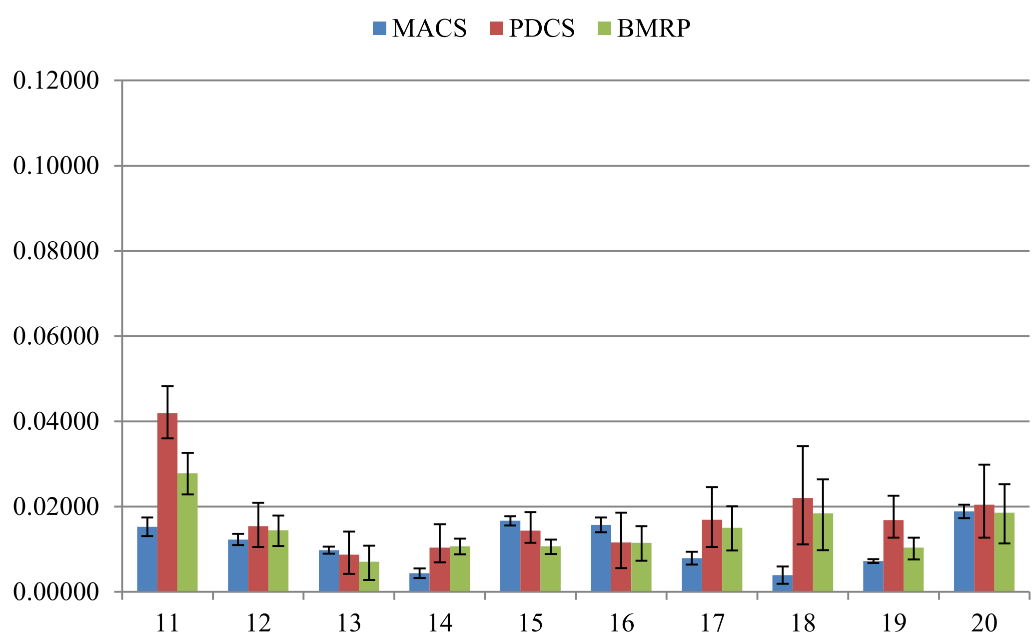

Figure 4.6.2. Estimates of the proportions of the adult Jewish population for metropolitan areas 11 to 20 generated with the three cross-survey methods based on data from the US2011 batch.

Chapter 5: Discussion

The goal of this study was to evaluate the operating characteristics of three different methods of cross-survey analysis in order to determine their suitability for estimating the proportions of low-incidence population groups. Cross-survey analysis offers an approach to generating low-incidence population estimates not readily available in today's census without conducting targeted, costly surveys to estimate group size directly. While the use of this methodology is growing (e.g., Gelman, 2009; Tighe et al., 2010), there is limited research on the accuracy of the different methods and on their relative strengths and weaknesses. This study presents an empirical example of their application.

The three cross-survey methods reviewed were meta-analysis of complex surveys (MACS), pooled design-based cross-survey (PDCS), and Bayesian multilevel regression with post-stratification (BMRP). I assessed their accuracy and precision through comparisons between the estimates of the proportions of low-incidence religious groups obtained with each of them, and benchmark estimates. Specifically, the proportions of the Canadian adult Jewish population generated with each cross-survey method (using data from nationally representative surveys of the Canadian population from two time periods) were compared with estimates of the same population generated by the Canadian Census and National Household Survey (NHS).

In addition, I determined the stability of the estimates produced with these methods by comparing two different estimates of the proportions of the adult Jewish population in the twenty largest U.S. metropolitan areas. These estimates were calculated based on data included in two equivalent subsamples of nationally representative surveys of the adult U.S. population. This final chapter reviews the findings, discusses the limitations of the study, and provides recommendations for future research.

Review of the findings

The findings indicate that cross-survey methods have the potential to produce accurate and precise estimates of low-incidence populations. Under the right conditions, cross-survey methods can generate accurate and precise estimates, as is the case with the BMRP estimates of the proportions of the adult Jewish population in Canadian metropolitan areas for 2001 [mean percent error (MAPE) of less than 4%]. Nonetheless, the level of accuracy and precision of the estimates varies depending on the conditions under which they are produced. This section offers some general remarks about the findings of the study and discusses each method.

Recall that the Canadian Jewish population estimates were obtained for four main scenarios in order to assess the accuracy and precision of the three cross-survey methods. The four scenarios differ in (a) the data batch used (number of surveys, number of respondents, and year of the survey) and (b) the level of the geographical areas for which estimates were generated. The first scenario contained data from forty-four surveys conducted between 1997 and 2004, and the geographical areas considered were the three largest Canadian metropolitan areas: Montreal, Toronto, and Vancouver. The second included data from forty-three surveys administered between 1997 and 2004, and the geographical areas examined were the four Canadian provinces with the largest proportions of this population, namely, Quebec, Ontario, Manitoba, and British Columbia. The third scenario comprised data from ten surveys conducted between 2006 and 2014, and the geographical areas analyzed were the above-mentioned metropolitan areas. Lastly, the fourth scenario contained data from seventeen surveys administered between 2006 and 2014, and the geographical areas considered were the four above-mentioned Canadian provinces.

As will be discussed later, the accuracy and precision of the estimates generated by the three cross-survey methods, BMRP in particular, are tied to the amount and type of data included in each analysis. Figure 5.1 presents the estimates of the proportions of the adult Jewish population calculated with the three methods for the four scenarios. It also includes the 2001 census estimates and the 2011 NHS estimates, which were used as benchmarks.

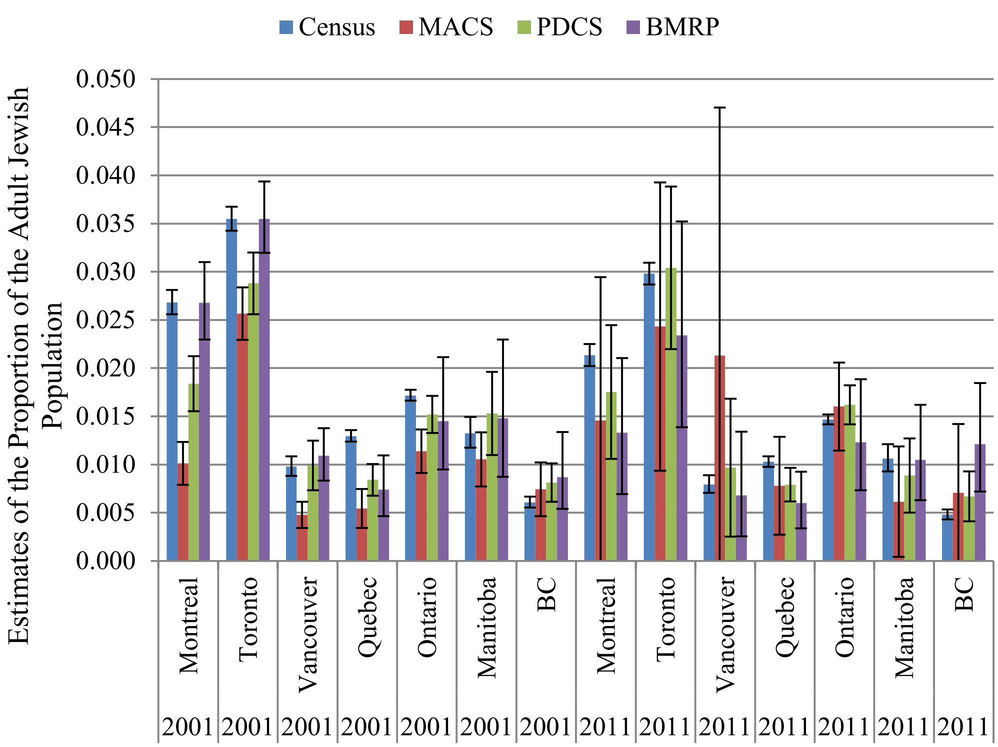

Figure 5.1 Comparisons between estimates of the proportions of the adult Jewish population (MACS, PDCS, BMRP) across the four scenarios of the study, and the 2001 census and 2011 NHS estimates.

The most accurate estimates of the adult Jewish population in Canada obtained in this study are the 2001 metropolitan area estimates produced by the BMRP method. [The MAPE for the BMRP estimates for 2001 survey data from the three metropolitan areas is less than 4%, compared to 14% to 73% for all other MAPEs]. No other cross-survey method under any conditions generated estimates with mean percent errors lower than 10%. Nonetheless, PDSC produced relatively accurate estimates (MAPE between 10% and 30%) in all four scenarios, as did BMRP for the adult Jewish population in the three metropolitan areas for 2011 (MAPE=25%) and in the four provinces for 2001 (MAPE=28%). Tables 5.1 to 5.4 display the comparisons between 2001 census and NHS estimates and cross-survey estimates of the adult Jewish population across the four scenarios of the study.

Table 5.1

Comparisons between Ca2001 Cross-Survey Estimates of the Proportions of the Adult Jewish Population in Metropolitan areas and 2001 Census Estimates

	2001 Census Estimates	MACS Est.	Err	PE	PDCS Est.	Err	PE	BMRP Est.	Err	PE
Montreal	0.027	0.010	-0.017	-62	0.018	-0.008	-31	0.027	< 0.000	0
Toronto	0.035	0.026	-0.010	-28	0.029	-0.007	-19	0.035	< 0.000	0
Vancouver	0.010	0.005	-0.005	-51	0.010	0.000	1	0.011	0.001	11

Table 5.2

Comparisons between Ca2011 Cross-Survey Estimates of the Proportions of the Adult Jewish Population in Metropolitan areas and 2011 NHS Estimates

	2011 NHS Estimates	MACS Est	Err	PE	PDCS Est	Err	PE	BMRP Est	Err	PE
Montreal	0.021	0.015	-0.007	-32	0.018	-0.004	-18	0.013	-0.008	-38
Toronto	0.030	0.024	-0.005	-18	0.030	0.001	2	0.023	-0.006	-21
Vancouver	0.008	0.021	0.013	169	0.010	0.002	22	0.007	-0.001	-14

Table 5.3

Comparisons between Ca2001 Cross-Survey Estimates of the Proportions of the Adult Jewish Population in the Provinces and 2001 Census Estimates

	2001 Census Estimates	MACS Est	Err	PE	PDCS Est.	Err	PE	BMRP Est	Err	PE
Quebec	0.013	0.005	-0.008	-58	0.008	-0.005	-35	0.007	-0.006	-43
Ontario	0.017	0.011	-0.006	-34	0.015	-0.002	-11	0.015	-0.003	-15
Manitoba	0.013	0.011	-0.003	-20	0.015	0.002	16	0.015	0.002	12
BC	0.006	0.007	0.001	22	0.008	0.002	34	0.009	0.003	44

Table 5.4

Comparisons between Ca2011 Cross-Survey Estimates of the Proportions of the Adult Jewish Population in the Provinces and 2011 NHS Estimates

	2011 Census Estimates	MACS Est	Err	PE	PDCS Est	Err	PE	BMRP Est	Err	PE
Quebec	0.010	0.008	-0.003	-24	0.008	-0.002	-23	0.006	-0.004	-42
Ontario	0.015	0.016	0.001	9	0.016	0.002	10	0.012	-0.002	-16
Manitoba	0.011	0.006	-0.004	-42	0.009	-0.002	-17	0.011	< 0.000	-1
BC	0.005	0.007	0.002	47	0.007	0.002	39	0.012	0.007	153

In general, estimates obtained with PDCS and BMRP were more accurate than the ones generated by MACS. Interestingly, the three methods underestimated the proportion of the adult Jewish population in Quebec and overestimated the proportion of this population in British Columbia for both time periods, although with different results. These results (underestimation and overestimation of the provincial estimates) may be linked to the size and distributions of the total population in the provinces or to the level of incidence of the Jewish population in these areas. As expected, due to the smaller number of surveys and of individual respondents, the confidence intervals are larger for the 2011 estimates. As will be described later, the relation between condition and precision differed across methods. As indicated by the coefficients of variation, the relative precision of all the cross-survey methods is higher for the metropolitan area estimates generated with data from the Ca2001 batch. In the other three scenarios, CVs are smaller for the PDCS method. Table 5.4 shows the coefficients of variation associated with the three cross-survey methods across the four scenarios of the study.

Table 5.5

Coefficients of Variation for the Three Cross-Survey Methods

	CV Ca2001 MACS	CV Ca2001 PDCS	CV[a] Ca2001 BMRP	CV Ca2011 MACS	CV Ca2011 PDCS	CV[a] Ca2011 BMRP
Metropolitan area						
Montreal	11.3	7.9	7.9	52.0	20.2	29.2
Toronto	5.5	5.7	5.7	31.4	14.1	28.4
Vancouver	14.7	13.3	13.3	61.6	37.6	50.7
Province						
Quebec	19.3	9.9	21.9	33.2	11.3	25.5
Ontario	10.3	6.4	20.2	14.5	6.4	23.7
Manitoba	13.8	14.3	25.1	47.5	22.2	23.8
BC	19.5	12.5	23.2	51.6	19.7	23.7

Note: [a] CV, the coefficient of variation, is calculated as the SE over the mean for estimates generated with the MACS and PDCS methods and as SD over the mean for estimates obtained with the BMRP method.

As noted earlier, the stability of the cross-survey methods was assessed by comparing the estimates produced with two parallel subsamples of the US adult population in the twenty largest metropolitan areas. The percentage differences between the estimates of the two subsamples were 15% for the BMRP method, 21% for the PDCS method, and 49% for the MACS method.

More research is needed to better understand the factors associated with the accuracy, precision, and stability of low-incidence population estimates generated with cross-survey methods. Still, this study sheds light on some of the strengths and limitations of these methods. What follows is a discussion of the findings.

Meta-Analysis of Complex Surveys

Results suggest that under the conditions of this study, the MACS method is not well suited for the estimation of low-incidence populations. The accuracy of the estimates of the adult Jewish population generated with this method was found to be relatively low for all four scenarios. The MAPE associated with the estimates of the adult Jewish population in the four scenarios of the study was larger than 30% (Ca2001 metropolitan area MAPE = 47%; Ca2001 province MAPE = 34%; Ca2011 metropolitan area MAPE = 73%; and Ca2011 province MAPE = 31%). Most of the estimates obtained with this method were lower than census and NHS values, the exception being the 2001 and 2011 British Columbia estimates and the 2011 Vancouver and Ontario estimates, which overestimated these populations. The 2001 census estimates were outside the 95% CI for all geographical areas except for British Columbia. 2011 NHS estimates, instead, were within the 95% CI in all cases, which is not surprising given the low precision of the 2011 estimates.

The precision of the estimates was relatively high for 2001, but not so for 2011. The differences in precision were especially pronounced for metropolitan area estimates. The variability of the estimates, measured by the coefficient of variation, differed within each scenario. It ranged from low (less than 15%) to very high (larger than 30%) (CVs for Ca2001 metropolitan areas ranged from 6 to 15%; for Ca2001 provinces, from 10 to

20%; for Ca2011 metropolitan areas, from 31 to 62%; and for Ca2011 provinces, from 15 to 52%).

In terms of stability, a large discrepancy was found between the estimates of the proportions of the adult Jewish population in the U.S. metropolitan areas obtained with the two parallel subsamples. The absolute percentage difference between the two samples varied between 5% and 200%, with half of the percentage differences being higher than 40%.

It is worth recalling here that the meta-analysis of complex surveys follows a separate approach; estimates generated for each survey are combined to create the cross-survey estimate. This method is thus more sensitive to the characteristics of individual surveys, especially when their number is small. In addition, given the low incidence of the Jewish population, this method might be more vulnerable to the sample sizes of the surveys from the individual geographical areas included in the study. These two factors might explain why the MACS method tended to underestimate this population. At the same time, its extreme overestimation of the 2011 adult Jewish population in Vancouver may be associated with the large number of surveys with zero Jewish respondents. Overall, further research is needed to determine the best way to combine surveys with zero low-incidence population. Although the alternative method suggested in this study does facilitate the inclusion of these surveys in the analysis, it may have altered results.

Given that results suggest that under certain conditions (such as the conditions in the four scenarios considered here), the MACS method is not well suited for the estimation of low-incidence religious populations, more research is needed to identify the factors associated with the accuracy and stability of the estimates generated by this method. Unlike PDCS and BMRP, MACS does not require individual records to produce estimates; these can be generated as long as there are survey level estimates (point estimates and variances) and design effects available for each survey. It is, therefore, worth investigating further how MACS-generated estimates relate to the different survey characteristics and under which conditions this method might produce more accurate results.

Pooled Design-Based Cross-Survey Method

The pooled design-based cross-survey method seems to be well suited to generate estimates of low-incidence populations with relatively good accuracy and precision. The method did produce estimates with absolute percentage errors (APE) lower than 15%, for example, the 2001 and 2011 estimates for Ontario. The MAPE associated with the estimates of the adult Jewish population in all four scenarios ranged from 14% to 24% (Ca2001 metropolitan area PDCS = 17%; Ca2001 province PDCS = 24%; Ca2011 metropolitan area PDCS = 14%; and Ca2011 province MAPE = 22%). Four of the estimates generated with this method where highly accurate; their APEs were lower than 15% (2001 Vancouver, 2001 and 2011 Ontario, and 2011 Toronto estimates). At the same time, four estimates had APEs larger than 30% (2001 Montreal, Quebec, British Columbia, and 2011 British Columbia).These results point to the fact that the method does have the potential to produce accurate estimates under the conditions of the four different scenarios. Nonetheless, it is not clear under which conditions this method produces accurate estimates and under which conditions it does not.

The 2001 estimates obtained with this method were lower than the 2001 census estimates for Montreal, Toronto, Quebec, and Ontario, and higher than these estimates for

Manitoba and British Columbia. The areas where the population was underestimated have relatively larger populations as well as larger proportions of adult Jewish population, which suggests that there might be a connection between these factors and the method's accuracy. Nonetheless, this pattern was not found among the 2011 estimates. PDCS estimates for Montreal and Quebec were indeed lower than NHS estimates, but the estimates for Ontario and Toronto were not; they were very close to NHS estimates (PE=2 and PE=10 respectively).

The precision of the 2001 PDCS estimates for the metropolitan areas is higher than that of the 2011 PDCS estimates for the same areas, but there is almost no difference between the precision of the 2001 and the 2011 PDCS estimates for the four provinces. The variability of the 2001 estimates, measured by the coefficient of variation, was low (less than 15%). The variability of the 2011 estimates, in turn, ranged from 6 to 22% for the provinces and from 14 to 38% for the metropolitan areas. Concerning stability, the percentage differences between the estimates generated with the two parallel subsamples ranged from 0 to 51%, with more than half of such differences being lower than 20%.

The precision and variability of the metropolitan area estimates seem to be related to the number of surveys and respondents included in the analysis. The standard error of 2001 metropolitan area estimates ranged from .0013 to .0016, and the coefficients of variation, from 6 to 13%. By contrast, the standard error of 2001 metropolitan area estimates ranges from .0035 to .0043, and the coefficient of variation, from 14 to 38%. The differences in precision and variability between the 2001 and the 2011 provincial estimates are much smaller. The results of the study suggest that the method has the potential for generating accurate and precise estimates of low-incidence populations. It is not clear, however, which factors affect their accuracy, precision and stability. More research is needed in this respect.

Bayesian Multilevel Regression with Post-Stratification

Among the cross-survey methods compared in this study, the Bayesian multilevel regression with post-stratification method overall generated the most accurate estimates of adult Jewish population in Canada. This method produced very accurate and precise estimates of the proportions of the adult Jewish population for the Canadian metropolitan areas with 2001 data (APE of less than .01% for Montreal and Toronto and of 11% for Vancouver). However, the 2011 estimates of this population in the three metropolitan areas, as well as the 2001 and 2011 estimates in the four Canadian provinces, varied more in accuracy and precision.

In terms of accuracy, the MAPE for metropolitan area estimates generated with data from the Ca2001 batch was less than 4%; for metropolitan area estimates produced with data from the Ca2011 batch, 24%, for provincial estimates obtained with data from the Ca2001 batch, 28%; and for provincial estimates generated with data from the Ca2011 batch, 53%. 2001 census estimates were within the credible intervals for all 2001 metropolitan area estimates and provincial estimates, except for Quebec. 2011 NHS estimates where within the credible intervals for all 2011 metropolitan area estimates and for the provinces of Toronto and Manitoba.

As was the case with PDCS estimates, the precision of the 2001 BMRP estimates for the metropolitan areas was higher than that of the 2011 BMRP estimates for the same areas. At the same time, the differences between the precisions of the 2001 and the 2011 BMRP estimates for the four provinces was smaller than for the metropolitan areas. The variability of the estimates, measured by the coefficients of variation, was low for the 2001 estimates (ranging from 5 to 13%), and higher for the 2011 metropolitan area estimates (from 29% to 51%). The difference between the variability of the 2001 and 2011 provincial estimates was small; CVs for the 2001 estimates ranged from 20 to 25%, and CVs for the 2011 estimates, from 24 to 26%.

With regard to stability, assessed by determining the difference between the estimates of the adult Jewish population in U.S. metropolitan areas obtained with two parallel subsamples, it varied between 0 and 20% for seventeen of the twenty metropolitan areas. Credible intervals of the estimates for the two subsamples overlap for all but one of the areas.

In view of the summary of results presented above, there are differences in accuracy, precision, and variability among the estimates produced in each of the four scenarios of the study. While all the Ca2001 BMRP estimates of the adult Jewish population in metropolitan areas were very accurate and precise and had low variability, results are not so clear for the other scenarios. These differences could be attributed to the conditions in each of the tested scenarios. The results of the study suggest (a) a relation between the number of surveys and respondents included in the analysis and the accuracy and precision of the estimates, and (b) a relation between the magnitude of the estimates and the level of geography considered.

Recall that estimates generated with BMRP are obtained through Bayesian hierarchical modeling. Data sets with a large number of surveys and respondents allow for more precise and accurate estimation of model parameters. This procedure may explain why estimates based on data from the Ca2001 batch were more accurate and precise than the ones produced with data from the Ca2011 batch. Estimates are also sensitive to the level of geographical area considered in the model. The distribution of the Jewish population in a province is far from homogeneous; it varies depending on the type of location (i.e. metropolitan, suburban, or rural) as well as on the size of the community. Unfortunately most of the surveys found did not include variables that allow the identification of type and size of location other than largest metropolitan areas (Montreal Toronto and Vancouver, and sometimes Calgary). This may be the reason behind the lower accuracy of the 2001 estimates for the provinces compared to the 2001 estimates for the metropolitan areas. Nonetheless, additional research should be conducted to identify the factors associated with the much-higher-than-expected estimates of the 2011 adult Jewish population in British Columbia.

Limitations of the study

There are a number of limitations to this study. What follows is a discussion of these limitations.

- Number of surveys: As is evident from the results of the study, the amount of data available to generate estimates affects the precision and variability of these estimates. Furthermore, given the nature of the MACS method, having additional surveys reduces the influence that the estimates of a single survey might have on cross-survey estimates. Finally, at least in the case of BMRP, results indicate that the accuracy of the estimates is associated with the amount of data used to generate them.

- Survey time-periods: Due to the relatively small number of surveys available to conduct the study, surveys administered in a period of seven years were collected in the same dataset. Although the preliminary analysis showed no correlations between the survey estimates of the adult Jewish population and the year the study was conducted, changes in the population analyzed may have occurred during this period.
- Survey sample sizes: The sample size of the surveys contained in the datasets were quite varied; some were as low as five hundred. In the estimation of low-incidence populations, the inclusion of surveys with small sample sizes may introduce bias into the estimation of less-populated areas, thus limiting the validity of the results, especially when applying MACS.
- Level of geographical information: It was difficult to find surveys that would allow for the identification of the size and location of respondents' residence beyond the largest metropolitan areas. The absence of these variables in the BMRP province models might have reduced the accuracy of the estimates.

Moreover, the voluntary nature of the 2011 NHS, which may have introduced bias to the 2011 estimates of the Jewish population, is another limitation of the study. The validity of the assessments of the accuracy and precision of 2011 estimates may have been affected by this bias.

Lastly, the generalizability of the study may be limited due to the nature of the low-incidence population being assessed. Questions regarding religious identification may be more susceptible to respondent interpretation than other types of questions used to identify low-incidence populations (e.g., questions about medical conditions).

Recommendations for future research

The results of this study have shown that cross-survey methods have the potential for generating estimates that are reasonably accurate and precise. Nonetheless, more research is needed to better understand the factors that might affect the quality of the estimates generated by each method. To this end, the current study may be expanded in four different ways: (a) by introducing simulated data for both surveys and respondents to facilitate a better understanding of sample-related factors that may affect the operational characteristics of the cross-survey methods; (b) by in-depth analysis of the distribution of estimates of the adult Jewish population within each survey and its impact on cross-survey estimates; (c) by estimating the proportions of other religious groups with the data used in this study; and (d) by exploring the possibility of combining and complementing the estimates generated by the various methods, taking advantage of their individual strengths to improve the estimation of low-incidence populations.

Future research can also benefit from further exploration into potential diagnostic tools that could be used to evaluate the sufficiency of available data for generating population estimates using cross-survey methods. These diagnostic tools can include those related to the assessment of "power", having sufficient sample sizes in terms of the number of independent surveys and the number of observations within surveys, as well as model-fitting in terms of having the appropriate sampling and other variables needed to generate accurate estimates.

Final conclusions

Cross-survey methods were found to be suitable for the estimation of low-incidence populations; more specifically, of a low-incidence religious group. The results of the study show that with sufficient data, in terms of both numbers of surveys and respondents, as well as auxiliary variables available for analysis, the Bayesian multilevel regression with post-stratification cross-survey method generates accurate and precise estimates of low-incidence religious groups. The study also shows that the pooled design-based cross-survey method generates relatively accurate and precise estimates. Additional research is needed, however, to better understand under what circumstances the method generates accurate and precise estimates and under what circumstances it does not. The meta-analysis of complex surveys method was found to be less suitable for estimating low-incidence populations under the conditions of this study. Although it generated precise estimates, these were less accurate than the ones produced by the other two methods.

The study was motivated by (a) the need to find alternative methodologies to estimate the size and distribution of low-incidence populations, especially those populations that are not measured directly by a national census; (b) the need to further understand the potential use and limitations of cross-survey methods for the estimation of low-incidence populations; and (c) the importance of the study of religious groups in the context of the scant availability of information on religious features in the US census. Although more research is needed to reach a greater understanding of the mechanisms that affect the accuracy and precision of estimates generated by cross-survey methods, the findings clearly show that these methods constitute a viable strategy for the estimation of low-incidence populations and, in particular, of low-incidence religious groups.

References

Baker, E. H., Rendall, M. S., & Weden, M. M. (2015). Epidemiological Paradox or Immigrant Vulnerability? Obesity Among Young Children of Immigrants. *Demography*, *52*(4), 1295-1320.

Baker, J. O. (2013). Acceptance of evolution and support for teaching creationism in public schools: The conditional impact of educational attainment. *Journal for the Scientific Study of Religion, 52*(1), 216-228. doi:10.1111/jssr.12007.

Bartolucci, A. A., & Hillegass, W. B. (2010). Overview, strengths, and limitations of systematic reviews and meta-analyses. In *Evidence-Based Practice: Toward Optimizing Clinical Outcomes* (pp. 17-33). Springer Berlin Heidelberg.

Binder, D. A., and Roberts, G. (2009), Design and model-based inference for model parameters. In Danny Pfeffermann and C.R. Rao (Eds.), *Handbook of Statistics 29: Sample Surveys: Inference and Analysis. (*pp. 33 – 54). Elsevier.

Binder, D. A. (2011). Pak. J. Statist. 2011 Vol. 27 (4), 371-390 Estimating model parameters from a complex survey under a model-design randomization framework. *Pak. J. Statist*, *27*(4), 371-390.

Boxer, M., Aronson, J. K., & Saxe, L. (2013). Using consumer panels to understand the characteristics of US Jewry. *Contemporary Jewry*, *33*(1-2), 63-82.

Boxer, M., Aronson, J., Brown, M. A., & Saxe, L. (2015). 2014 Greater Seattle Jewish Community Study.

Bryan, T. (2004). Population estimates. *The methods and materials of demography*, *2*, 523-560. In Swanson, D., Siegel, J. S., & Shryock, H. S. (2004). *The methods and materials of demography*. Elsevier.

Chiu, M., Austin, P. C., Manuel, D. G., & Tu, J. V. (2010). Comparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. *Canadian Medical Association Journal*, *182*(8), E301-E310.

Chipman, H., George, E. I., McCulloch, R. E., Clyde, M., Foster, D. P., & Stine, R. A. (2001). The practical implementation of Bayesian model selection. *Lecture Notes-Monograph Series*, 65-134.

Cohen, S. M. (2015). The Shrinking Jewish Middle. In *American Jewish Year Book 2014* (pp. 27-31). Springer International Publishing.

Cooperman, A., Smith, G. A., Hackett, C., & Kuriakose, N. (2013). A portrait of Jewish Americans: Findings from a Pew Research Center survey of US Jews. *Washington, DC: Pew Research Center*.

Cumming, G., & Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. *American Psychologist*, *60*(2), 170.

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

DellaPergola, S. (2005). Was it the demography?: A reassessment of US Jewish population estimates, 1945–2001. *Contemporary Jewry*, *25*(1), 85-131.

Della Pergola, S. (2011). *Jewish demographic policies: Population trends and options in Israel and in the Diaspora*. Jewish People Policy Institute.

DellaPergola, S. (2013). How many Jews in the United States? The demographic perspective. *Contemporary Jewry*, *33*(1-2), 15-42.

DellaPergola, S. (2014). World Jewish Population, 2013. In Dashefsky, A, & Sheskin, I. (Eds.), *American Jewish Year Book 2013: The Annual Record of the North American Jewish Communities.* (pp. 279-358). Springer International Publishing.

Figueiredo, J., & Campos, P. (2013). Estimation of Underrepresented Strata in Preelection Polls: A Comparative Study. In *Recent Developments in Modeling and Applications in Statistics* (pp. 47-57). Springer Berlin Heidelberg.

Fox, K. M. (2011). *A framework for the meta-analysis of survey data.* (Doctoral dissertation, Queen's University, 2011). Retrieved from http://hdl.handle.net/1974/6900

Gelman, A. (2009). *Red state, blue state, rich state, poor state: Why Americans vote the way they do (expanded edition).* Princeton University Press.

Gelman, A. (2014). How Bayesian analysis cracked the red-state, blue-state problem. *Statistical science*, *29*(1), 26-35.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., Lee, D., & Guo, J. (2015). Stan a probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics,* 40, 530-543.

Ghosh, M. (2009), Bayesian Developments in Survey Sampling. In Danny Pfeffermann and C.R. Rao (Eds.), *Handbook of Statistics 29: Sample Surveys: Inference and Analysis.* (pp. 153 -187). Elsevier.

Ghosh, M., & Rao, J. N. K. (1994). Small area estimation: an appraisal. *Statistical science*, 55-76.

Gill, J. (2014). Bayesian methods: A social and behavioral sciences approach (Vol. 20). CRC press.

Goldstein, S. (1992). Profile of American Jewry: Insights from the 1990 National Jewish Population Survey. In David Singer (Ed), *American Jewish Year Book*, (pp. 77–173). New York: The American Jewish Committee.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. New York: John Wiley & Sons.

Hartman, H., & Sheskin, I. (2011). The influence of community context and individual characteristics on Jewish identity: A 21-community study. *Storrs, CT: North American Jewish Data Bank*.

Heilman, S. (2013). Editor's Introduction for the Special Issue on Jewish Demography in the United States. *Contemporary Jewry*, *33*(1-2), 1-2.

Herrera, D. G., Schiefelbein, E. L., Smith, R., Rojas, R., Mirchandani, G. G., & McDonald, J. A. (2012). Cervical Cancer Screening in the US–Mexico Border Region: A Binational Analysis. *Maternal and child health journal*, *16*(2), 298-306.

IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.

Kadushin, C., Phillips, B. T., & Saxe, L. (2005). National Jewish Population Survey 2000–01: A guide for the perplexed. *Contemporary Jewry*, *25*(1), 1-32.

Kalton, G. (2009). Methods for oversampling rare subpopulations in social surveys. *Survey Methodology*, *35*(2), 125-141.

Kastellec, J. P., Lax, J. R., & Phillips, J. (2010). Estimating state public opinion with multi-level regression and poststratification using R. *Unpublished manuscript, Princeton University*.

Kastellec, J. P., Lax, J. R., & Phillips, J. H. (2010b). Public opinion and Senate confirmation of Supreme Court nominees. *The Journal of Politics*, *72*(03), 767-784.

Kim, J. J., Li, J., & Valliant, R. (2007). Cell collapsing in poststratification. *Survey Methodology, 33*(2), 139-150.

Kish, L. (1965). Survey sampling.

Kish, L. (1999). Cumulating/combining population surveys. *Survey Methodology*, *25*(2), 129-138.

Kish, L. (2002). Combining multipopulation statistics. *Journal of statistical planning and inference*, *102*(1), 109-118.

Koffman, David S., and Morton Weinfeld. "Recent Developments in the Social Scientific Study of Canadian Jews." *Contemporary Jewry* 31.3 (2011): 199-221. *ProQuest.* Web. 5 Feb. 2016.

Korn, E. L., & Graubard, B. I. (1999). *Analysis of health surveys*. John Wiley & Sons.

Kosmin, B. A. (2015). It's the Best of Times: It's the Worst of Times. In *American Jewish Year Book 2014* (pp. 61-65). Springer International Publishing.

Kosmin, B., Goldstein, S., Waksberg, J., Lerer, N., Keysar, A., & Scheckner, J. (1991). *Highlights of the CJF 1990 national Jewish population survey* (pp. 10-19). New York: Council of Jewish Federations.

Kosmin, B. A. (2009). The Changing Population Profile of American Jews, 1990-2008.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Lachan, R. (2008). Finite Population. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods.* (pp. 690-692). Thousand Oaks, CA: SAGE Publications, Inc.

Lansky, A., Finlayson, T., Johnson, C., Holtzman, D., Wejnert, C., Mitsch, A., Gust, D., Chen, R., Mizuno, Y. & Crepaz, N. (2014). Estimating the number of persons who inject drugs in the United States by meta-analysis to calculate national rates of HIV and hepatitis C virus infections.

Lepkowski, J. M. (2008). Population. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods.* (pp. 690-692). Thousand Oaks, CA: SAGE Publications, Inc.

Lim, C. (2013). Counting the faithful: Measuring local religious contexts in the United States. *Journal for the Scientific Study of Religion, 52*(2), 386-400. doi:10.1111/jssr.12027.

Little, R. J. (2004). To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association*, *99*(466).

Lohr, S. L. (2011). Alternative survey sample designs: Sampling with multiple overlapping frames. *Survey Methodology*, *37*(2), 197-213.

Lohr, S., & Rao, J. N. K. (2006). Estimation in multiple-frame surveys. *Journal of the American Statistical Association*, *101*(475).

Lohr, S. L., & Prasad, N. N. (2003). Small area estimation with auxiliary survey data. *Canadian Journal of Statistics*, *31*(4), 383-396.

Lumley, T (2004). Analysis of complex survey samples. Journal of Statistical Software 9(1): 1-19.

Lumley, T (2014). Survey: analysis of complex survey samples. R package version 3.30.

Malec, D. (2008). Superpopulation. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods.* (pp. 690-692). Thousand Oaks, CA: SAGE Publications, Inc.

Mattis, J. S. (2001). Religion and African American political life. *Political Psychology*, *22*(2), 263-278.

Mayer, E., Kosmin, B., & Keysar, A. (2001). American Jewish identity survey. *New York: Center for Cultural Judaism*.

National Center for Education Statistics. (2012). *2012 Revision of NCES Statistical Standards: Analysis of data production of estimates or projections*. Retrieved from https://nces.ed.gov/statprog/2012/

Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis, 12*(4), 375-385.

Perl, P., Greely, J. Z., & Gray, M. M. (2006). What proportion of adult Hispanics are catholic? A review of survey data and methodology. *Journal for the Scientific Study of Religion, 45*(3), 419-436. doi:10.1111/j.1468-5906.2006.00316.x

Rebhun, U. (Ed.). (2014). *The Social Scientific Study of Jewry: Sources, Approaches, Debates* (Vol. 27). Oxford University Press, USA.

Pfeffermann, D. (2002). Small Area Estimation: New Developments and Directions. *International Statistical Review / Revue Internationale De Statistique*, *70*(1), 125–143. http://doi.org/10.2307/1403729

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science, 28*(1), 40-68.

Porter, S. (2008). Rare populations. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods.* (pp. 690-692). Thousand Oaks, CA: SAGE Publications, Inc. doi: http://dx.doi.org/10.4135/9781412963947.n444.

Purcell, D. W., Johnson, C. H., Lansky, A., Prejean, J., Stein, R., Denning, P., Gau, Z., Weinstock, H., Su, J., Crepaz, N. (2012). Estimating the Population Size of Men Who Have Sex with Men in the United States to Obtain HIV and Syphilis Rates. *The Open AIDS Journal*, *6*, 98–107. http://doi.org/10.2174/1874613601206010098

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rao, J. N. (2003). *Small area estimation*. John Wiley & Sons, Inc.

Rao, S. R., Graubard, B. I., Schmid, C. H., Morton, S. C., Louis, T. A., Zaslavsky, A. M., Finkelstein, D. M.. (2008). Meta-analysis of survey data: Application to health services research. *Health Services and Outcomes Research Methodology, 8*(2), 98-114.

Roberts, G., & Binder, D. (2009). Analyses based on combining similar information from multiple surveys. In *JSM Proceedings* (pp. 2138-2147).

Sasson, T., Kadushin, C., & Saxe, L. (2010). Trends in American Jewish attachment to Israel: An assessment of the ‘‘distancing’’ hypothesis. *Contemporary Jewry 30:* 297–319.

Saxe, L., & DellaPergola, S. (2013). Introduction: Special issue on Jewish demography in the United States. *Contemporary Jewry, 33*(1-2), 3-8. doi:http://dx.doi.org.resources.library.brandeis.edu/10.1007/s12397-013-9096-4

Saxe, L., Sasson, T., & Aronson, J. K. (2015). Pew's portrait of American Jewry: A reassessment of the assimilation narrative. In *American Jewish Year Book 2014* (pp. 71-81). Springer International Publishing.

Saxe, L., & Tighe, E. (2013). Estimating and understanding the Jewish population in the United States: A program of research. *Contemporary Jewry*, *33*(1-2), 43-62.

Saxe, L., Tighe, E., & Boxer, M. (2014). Measuring the size and characteristics of American Jewry: A new paradigm to understand an ancient people. In Rebhun, U. (Ed), *The Social Scientific Study of Jewry: Sources, Approaches, Debates (Vol. 27).* (pp.37-54). Oxford University Press, USA.

Schenker, N., Gentleman, J. F., Rose, D., Hing, E., & Shimizu, I. M. (2002). Combining estimates from complementary surveys: a case study using prevalence estimates from national health surveys of households and nursing homes. *Public Health Reports*, *117*(4), 393.

Schenker, N., & Raghunathan, T. E. (2007). Combining information from multiple surveys to enhance estimation of measures of health. *Statistics in medicine*, *26*(8), 1802-1811.

Scheuren, F. (2008). Post-stratification. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods.* (pp. 595-598). Thousand Oaks, CA: SAGE Publications, Inc. doi: http://dx.doi.org/10.4135/9781412963947.n388.

Schieman, S. (2011). Education and the importance of religion in decision making: Do other dimensions of religiousness matter?. *Journal for the Scientific Study of Religion*, *50*(3), 570-587.

Shahar, C. (2004). 2001 Census Analysis Series: The Jewish Community of Canada. Part 1: Basic Demographics. UIA Federations Canada, Montreal.

Shahar, C. (2014). 2011 National Household Survey Analysis: The Jewish Population of Canada Part 1 Basic Demographics. UIA Federations Canada, Montreal.

Shahar, C. 2004. *The Jewish community of Canada*. Toronto: Jewish Federations of Canada.UIA.

Sherkat, D. E., & Ellison, C. G. (1999). Recent developments and current controversies in the sociology of religion. *Annual Review of Sociology*, 363-394.

Sheskin, I. M. (2013). Uses of local Jewish community study data for addressing national concerns. *Contemporary Jewry*, *33*(1-2), 83-101.

Sheskin, I., & Dashefsky, A. (2015). Jewish Population in the United States, 2014. In *American Jewish Year Book 2014* (pp. 215-283). Springer International Publishing.

Shirley, K. E., & Gelman, A. (2015). Hierarchical models for estimating state and demographic trends in US death penalty public opinion. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(1), 1-28.

Stan Development Team. (2016). RStan: the R interface to Stan, Version 2.10.1. http://mc-stan.org

StataCorp. (2011). *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP. Statistics Canada. (2005). History of the census in Canada. Retrieved from http://www12.statcan.ca/english/census01/info/history.cfm

Statistics Canada, (2001). Census of Canada, 2001, Individuals File (public-use microdata file). Statistics Canada (producer). Using LANDRU (distributor). All computations, use and interpretation of these data are entirely those of the author.

Statistics Canada. (2005). History of the census in Canada. Retrieved from http://www12.statcan.ca/english/census01/info/history.cfm

Statistics Canada, (2011). National Household Survey, 2011, Individuals File (public-use microdata file). Statistics Canada (producer). Using LANDRU (distributor). All computations, use and interpretation of these data are entirely those of the author.

Statistics Canada. (2011b). *2011 National Household (questionnaire)*. Retrieved from http://www12.statcan.gc.ca/nhs-enm/2011/ref/about-apropos/ques_guide-eng.cfm.

Steel, Z., Chey, T., Silove, D., Marnane, C., Bryant, R. A., & Van Ommeren, M. (2009). Association of torture and other potentially traumatic events with mental health outcomes among populations exposed to mass conflict and displacement: a systematic review and meta-analysis. *Jama*, *302*(5), 537-549.

Sterba, S. K. (2009). Alternative model-based and design-based frameworks for inference from samples to populations: From polarization to integration. *Multivariate behavioral research*, *44*(6), 711-740.

Tayman, J., & Swanson, D. A. (1999). On the validity of MAPE as a measure of population forecast accuracy. *Population Research and Policy Review*, *18*(4), 299-322.

Thomas, S., & Wannell, B. (2009). Combining cycles of the Canadian community health survey. *Health Rep*, *20*(1), 53-58.

Tighe, E., Livert, D., Barnett, M., & Saxe, L. (2010). Cross-survey analysis to estimate low-incidence religious groups. *Sociological Methods & Research, 39*(1), 56-82.

Tighe, E., Saxe, L., Magidin de Kramer, R., & Parmer, D. (2013*). American Jewish population estimates: 2012.* Waltham, MA: Steinhardt Social Research Institute at Brandeis University.

Tighe, E., Saxe, L., Kadushin, C., Magidin de Kramer, R., Nursahedov, B., Aronson, J., Cherney, L. (2012). *Estimating the Jewish population of the United States: 2000–2010*. Waltham, MA: Cohen Center for Modern Jewish Studies, Brandeis University.

U.S. Census Bureau (2008). A compass for understanding and using American Community Survey data: What general data users need to know. Washington DC: U.S. Government Printing Office.

U.S. Census Bureau (2014). *Does the Census Bureau have data for religion*. Retrieved from https://ask.census.gov/faq.php?id=5000&faqId=29 .

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York: Springer.

Yowell, T., & Devine, J. (2013, July). Evaluating current and alternative methods to produce 2010 County population estimates. Population Division Working Paper 100, U.S. Bureau of the Census, Washington, DC.

Washington State Office of financial Management. (2010). *American Community Survey Publication No. 1*: *User Guide*. Gardner, E., Kimpel, T., Zhao, Y. Retrieved from http://www.ofm.wa.gov/pop/acs/userguide/ofm_acs_user_guide.pdf

Weinfeld, M., & Schnoor, R. F. (2015). The Demography of Canadian Jewry, the "Census" of 2011: Challenges and Results. In *American Jewish Year Book 2014* (pp. 285-299). Springer International Publishing.

Weinfeld, M., Schnoor, R. F., & Koffman, D. S. (2013). Overview of Canadian Jewry. In *American Jewish Year Book 2012* (pp. 55-90). Springer Netherlands.

Wilder, E. I. (2014). Defining and Measuring the Socioeconomic Status of the Jews. In Rebhun, U. (Ed), *The Social Scientific Study of Jewry: Sources, Approaches, Debates (Vol. 27).* (pp.109-132). Oxford University Press, USA.

Appendix A1: Surveys Included In the Canadian Batches

America's Barometer	I thank the Latin American Public Opinion Project (LAPOP) and its major supporters (the United States Agency for International Development, the United Nations Development Program, the Inter-American Development Bank, and Vanderbilt University) for making the data available.
	The AmericasBarometer, Canada 2014. Latin American Public Opinion Project (LAPOP), www.LapopSurveys.org.
	The AmericasBarometer, Canada 2012. Latin American Public Opinion Project (LAPOP), www.LapopSurveys.org.
Canadian Election Study	Data from the Canadian Election Surveys were provided by the Institute for Social Research, York University. The survey was funded by the Social Sciences and Humanities Research Council of Canada (SSHRC), and was completed for the 2000 Canadian Election Team of André Blais (Université de Montréal), Elisabeth Gidengil (McGill University), Richard Nadeau (Université de Montréal) and Neil Nevitte (University of Toronto). Neither the Institute for Social Research, the SSHRC, nor the Canadian Election Survey Team are responsible for the analyses and interpretations presented here.
	André Blais, Elisabeth Gidengil, Richard Nadeau, Neil Nevitte. Canadian Election Study, 1997. Institute for Social Research, York University. Canadian Opinion Research Archive CES-E-1997. http://odesi1.scholarsportal.info/webview/index/en/Odesi/ODESI-Click-to-View-Categories-.d.6/Elections-and-Politics.d.18/CANADA.d.19/Canadian-Election-Study.d.189/1997.d.1124/Canadian-Election-Study-1997/fStudy/CES-E-1997
	André Blais, Elisabeth Gidengil, Richard Nadeau, Neil Nevitte. Canadian Election Study, 2000. Institute for Social Research, York University. Canadian Opinion Research Archive CES00. http://130.15.161.246:82/webview/index/en/CORA/Canadian-Opinion-Research-Archive-CORA-.c.CORA/Canadian-Election-Studies.d.107/Canadian-Election-Study-2000/fStudy/CES00
	André Blais, Elisabeth Gidengil, Neil Nevitte, Patrick Fournier, Joanna Everitt. Canadian Election Study, 2004. Institute for Social Research, York University. Canadian Opinion Research Archive CES-E-2004. http://odesi1.scholarsportal.info/webview/index/en/Odesi/ODESI-Click-to-View-Categories-.d.6/Elections-and-Politics.d.18/CANADA.d.19/Canadian-Election-

	Study.d.189/2004.d.1032/Canadian-Election-Study-2004/fStudy/CES-E-2004
	Elisabeth Gidengil, Joanna Everitt, Patrick Fournier, Neil Nevitte. Canadian Election Study, 2008. Institute for Social Research, York University. Canadian Opinion Research Archive CES08-E-2008. http://odesi2.scholarsportal.info/webview/index/en/Odesi/ODESI-Click-to-View-Categories-.d.6/Elections-and-Politics.d.18/CANADA.d.19/Canadian-Election-Study.d.189/2008.d.1027/Canadian-Election-Study-2008/fStudy/CES08-E-2008
	Patrick Fournier, Fred Cutler, Stuart Soroka, Deitland Stolle. Canadian Election Study, 2011. Institute for Social Research, York University. Canadian Opinion Research Archive CES2011. http://130.15.161.246:82/webview/index/en/CORA/Canadian-Opinion-Research-Archive-CORA-.c.CORA/Canadian-Election-Studies.d.107/Canadian-Election-Study-2011/fStudy/CES2011
Centre for Research and Information on Canada Surveys	CRIC Survey on Official Languages, 2003, (December, 2003) [cricsol03]. Centre for Research and Information on Canada, Montreal, QC [producer], Environics Research Group, Toronto, ON [producer], Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	CRIC Canada and World Affairs Survey, 2001-2002, [criccwa0102]. Centre for Research and Information on Canada, Montreal, QC [producer], Environics Research Group, Toronto, ON [producer], Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Centre for Research and Information on Canada, 2015, "Attitudes towards Internationalism & Federalism, 2003 [Canada]", hdl:10864/11071, Scholars Portal Dataverse, V3
Equality, Security and Community Survey	Data from the first wave of the Equality, Security and Community (ESC) survey were provided by the Institute for Social Research, York University. The ESC project was funded by the Social Sciences and Humanities Research Council of Canada (SSHRC), grant number 412-97-0003, Project Director Dr. Jonathan R. Kesselman, University of British Columbia. The survey component of the ESC study was completed under the direction of Dr. Richard Johnston, UBC. Neither the Institute for Social Research, SSHRC, nor the ESC Research Team are responsible for the analyses and interpretations presented here.
	Equality, Security and Community Survey - Wave 1, 1999 [Canada]. Canadian Opinion Research Archive escw1-99. http://130.15.161.246:82/webview/index/en/CORA/Canad

	ian-Opinion-Research-Archive-CORA-.c.CORA/Other-Academic-Surveys.d.88/Equality-Security-and-Community-Survey-Wave-1-1999-Canada-/fStudy/escw1-99
Ethnic Diversity Survey	Statistics Canada. Ethnic Diversity Survey, 2002. eds-89M0019-E-2002, V1. http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4508. The results or views expressed in this dissertation are those of the author and are not those of Statistics Canada.
Focus Canada Surveys	Environics Focus Canada 2011 (November, 2011) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 2006-3 (September, 2006) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 2006-4 (December, 2006) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 2001-1 (April, 2001) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 2001-2 (June, 2001) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 2001-3 (September, 2001) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 2001-4 (December, 2001) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 2000-1 (March, 2000) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 2000-2 (June, 2000) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 2000-3 (September, 2000) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].

	Environics Focus Canada 2000-4 (December, 2000) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 1999-1 (April, 1999) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 1999-2 (July, 1999) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 1999-3 (September, 1999) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 1999-4 (December, 1999) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 1998-1 (March, 1998) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 1998-2 (June, 1998) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 1998-3 (October, 1998) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
	Environics Focus Canada 1998-4 (December, 1998) [computer file]. Environics Research Group, Toronto, ON [producer],Canadian Opinion Research Archive, Queen's University, Kingston, ON [distributor].
Gallup Canada, Inc. Polls	Gallup Canada, Inc.. Gallup Poll # 1999-04, Apr, 1999 [dataset]. CNCIPO1999-04, Version 2. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-04, Apr, 2000 [dataset]. CNCIPO2000-04, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Canadian Gallup Poll # 1999-08: Politics/Health Care/Death Penalty, Aug, 1999 [dataset]. CNCIPO1999-08, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center

	for Public Opinion Research, RoperExpress[distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-08, Aug, 2000 [dataset]. CNCIPO2000-08, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll: Federal Election/Christmas, Dec, 2000 [dataset]. CNCIPO2000-12, Version 4. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-02, Feb, 2000 [dataset]. CNCIPO2000-02, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-01, Jan, 2000 [dataset]. CNCIPO2000-01, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 1999-07: Crime/Child Abuse, Jul, 1999 [dataset]. CNCIPO1999-07, Version 2. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-07, Jul, 2000 [dataset]. CNCIPO2000-07, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-06, Jun, 2000 [dataset]. CNCIPO2000-06, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-03, Mar, 2000 [dataset]. CNCIPO2000-03, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Poll # 1999-05, May, 1999 [dataset]. CNCIPO1999-05, Version 2. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.

	Gallup Canada, Inc.. Gallup Canada Poll # 2000-05, May, 2000 [dataset]. CNCIPO2000-05, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Canadian Gallup Poll: Politics/Abortion, Nov, 1999 [dataset]. CNCIPO1999-11, Version 2. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-11, Nov, 2000 [dataset]. CNCIPO2000-11, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 1999-10: Charity/Gun Control, Oct, 1999 [dataset]. CNCIPO1999-10, Version 2. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-10, Oct, 2000 [dataset]. CNCIPO2000-10, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
	Gallup Canada, Inc.. Gallup Canada Poll # 2000-09, Sep, 2000 [dataset]. CNCIPO2000-09, Version 3. Gallup Canada, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Oct-11-2016.
General Social Surveys	Statistics Canada. General Social Survey, 2012 [Canada]: Cycle 26, Caregiving and Care Receiving (GSS 26). gss-12M0025-E-2012-c-26. http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4502
	Statistics Canada. General Social Survey, 2013 [Canada]: Cycle 27, Social Identity [version 2](GSS 27). gss-89M0032x-E-2013-c27, V2. http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5024
ISSP Research Group Surveys	The data utilized in this publication were documented and made available by the Zentralarchiv Feuer Empirische Sozialforschung, Koeln. The data for the 'ISSP' were collected by independent institutions in each country (see: principal investigators in the study-description-schemes for each participating country). Neither the original collectors nor the Zentralarchiv bear any responsibility for the analyses or interpretations presented here.

	ISSP Research Group (2012): International Social Survey Programme: Family and Changing Gender Roles IV - ISSP 2012. GESIS Data Archive, Cologne. ZA5900 Data file Version 3.0.0 doi:10.4232/1.12339.
Pew Surveys	All Pew surveys retrieved from Pew Research Center for the People & the Press. http://pewresearch.org/ The Center bears no responsibility for the interpretations presented or conclusions reached based on analysis of the data.
	Pew Global Attitudes, 2002. Pew Research Center. http://www.pewglobal.org/2013/05/01/spring-2013-survey-data/
	Pew Global Attitudes, 2003. Pew Research Center. http://www.pewglobal.org/2003/06/03/views-of-a-changing-world-2003/
	Pew Global Attitudes, 2007. Pew Research Center. http://www.pewglobal.org/2007/05/28/spring-2007-survey-data/
	Pew Global Attitudes, 2009. Pew Research Center. http://www.pewglobal.org/2009/07/23/confidence-in-obama-lifts-us-image-around-the-world/
Transatlantic Trends Survey	Wunderlich, Tanja, Astrid Ziebarth, Delancey Gustin, Pierangelo Isernia, Claudia Diehl, and Susan Martin. Transatlantic Trends: Immigration, 2009. ICPSR31801-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-09-01.http://doi.org/10.3886/ICPSR31801.v1
Voice of the People Surveys	Leger Marketing. 2009. Voice of the People, End of Year Survey, November 2008 [Canada]. [public-use microdata file]. Montreal, QC. Leger Marketing [distributor].
	Leger Marketing. 2010. Voice of the People, 1st Edition Survey, April 2009 [Canada]. [public-use microdata file]. Montreal, QC. Leger Marketing [distributor].
	Leger Marketing. 2008. Voice of the People, 1st Edition Survey, November 2007 [Canada]. [public-use microdata file]. Montreal, QC. Leger Marketing [distributor].
World Values Surveys	WORLD VALUES SURVEY Wave 4 1999-2004 OFFICIAL AGGREGATE v.20140429. World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: Asep/JDS, Madrid SPAIN.
	WORLD VALUES SURVEY Wave 5 2005-2008 OFFICIAL AGGREGATE v.20140429. World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: Asep/JDS, Madrid SPAIN.

Appendix A2: Surveys Included in the US Batch

American National Election Study (series)	The American National Election Studies (ANES; www.electionstudies.org). The ANES 2012 Time Series Study [dataset]. Stanford University and the University of Michigan [producers].These materials are based on work supported by the National Science Foundation under grants SES-0937727 and SES-0937715, Stanford University, and the University of Michigan. Any opinions, findings and conclusions or recommendations expressed in these materials are those of the author(s) and do not necessarily reflect the views of the funding organizations.
General Social Survey (series)	Smith, Tom W, Peter Marsden, Michael Hout, and Jibum Kim. *General Social Surveys, 1972-2012* [machine-readable data file] /Principal Investigator, Tom W. Smith; Co-Principal Investigator, Peter V. Marsden; Co-Principal Investigator, Michael Hout; Sponsored by National Science Foundation. --NORC ed.-- Chicago: NORC at the University of Chiago [producer]; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut [distributor], 2015.
National Annenberg Election Survey	National Annenberg Election Survey 2008. The Annenberg Public Policy Center of the University of Pennsylvania. Philadelphia, PA. http://www.annenbergpublicpolicycenter.org/2008-naes-telephone-and-online-data-sets/.
ABC News/Washington Post Polls	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, April 2009. ICPSR27764-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-04-29. http://doi.org/10.3886/ICPSR27764.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, December 2009. ICPSR29045-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-09-17. http://doi.org/10.3886/ICPSR29045.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, February 2009. ICPSR27762-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-04-30. http://doi.org/10.3886/ICPSR27762.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, February 2010. ICPSR30202-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-05-02. http://doi.org/10.3886/ICPSR30202.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, January 2009. ICPSR27761-v1. Ann Arbor, MI: Inter-university

	Consortium for Political and Social Research [distributor], 2010-06-10. http://doi.org/10.3886/ICPSR27761.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, January 2010. ICPSR30201-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-04-13. http://doi.org/10.3886/ICPSR30201.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, July 2009. ICPSR27766-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-10-11. http://doi.org/10.3886/ICPSR27766.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, June 2009. ICPSR27765-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-11-09. http://doi.org/10.3886/ICPSR27765.v2
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, March 2009. ICPSR27763-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-04-29. http://doi.org/10.3886/ICPSR27763.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, November 2009. ICPSR29043-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-10-21. http://doi.org/10.3886/ICPSR29043.v1
	ABC News, and The Washington Post. ABC News/Washington Post Poll, April 2008 [Computer file]. ICPSR24606-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-11-12. doi:10.3886/ICPSR24606.v1
	ABC News, and The Washington Post. ABC News/Washington Post Poll, February 2008 [Computer file]. ICPSR24605-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-09-01. doi:10.3886/ICPSR24605.v1
	ABC News, and The Washington Post. ABC News/Washington Post Poll, January 2008. ICPSR24603-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-10-21. doi:10.3886/ICPSR24603.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, April 2010. ICPSR30204-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-09-21. http://doi.org/10.3886/ICPSR30204.v1
	ABC News, and The Washington Post. ABC News/Washington Post Monthly Poll, June 2010. ICPSR30205-v1. Ann Arbor, MI: Inter-university

	Consortium for Political and Social Research [distributor], 2011-10-12. http://doi.org/10.3886/ICPSR30205.v1
CBS News Polls (includes CBS/New York Times, CBS/60 Minutes/Vanity Fair)	CBS News. CBS News Monthly Poll #1, January 2010. ICPSR31562-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-07-08. http://doi.org/10.3886/ICPSR31562.v1
	CBS News. CBS News Monthly Poll, August 2009. ICPSR27803-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-12-06. http://doi.org/10.3886/ICPSR27803.v1
	CBS News. CBS News Monthly Poll, June 2010. ICPSR31575-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-09-07. http://doi.org/10.3886/ICPSR31575.v1
	CBS News. CBS News Monthly Poll, March 2010. ICPSR31566-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-08-15. http://doi.org/10.3886/ICPSR31566.v1
	CBS News, and The New York Times. CBS News/New York Times Monthly Poll #2, May 2010. ICPSR31573-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-08-11. http://doi.org/10.3886/ICPSR31573.v1
	CBS News, and The New York Times. CBS News/New York Times Monthly Poll, July 2009. ICPSR27802-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-04-12. http://doi.org/10.3886/ICPSR27802.v1
	CBS News, and Vanity Fair. CBS News/Vanity Fair Monthly Poll, May 2010. ICPSR31572-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-08-11. http://doi.org/10.3886/ICPSR31572.v1
Pew Surveys	All Pew surveys retrieved from Pew Research Center for the People & the Press. http://pewresearch.org/ The Center bears no responsibility for the interpretations presented or conclusions reached based on analysis of the data.
	America's Place in the World 2009, 2013
	Biennial Media Consumption 2008, 2010, 2012
	Civic Engagement Aug 2012
	Economic Survey Feb 2008
	Gender and Generations Nov 2012
	General Public Science Survey April 2009
	Generations Sept 2011
	Generational Change and Veteran GP Sept 2011
	Global Attitudes: March 2008, March 2011
	Higher Education/Housing Survey 2011
	Media Consumption 2010
	Middle Class Update July 2012
	Millennial Survey Jan 2010

	Omnibus Surveys: June 2009, Aug 2009, Nov 2009, Jan 2010, Feb Week 2 2013, March Week 2 & Week 4 2013, June Week 2 2013
	Political and Economic Surveys: Feb 2008, July 2008
	Political and Future Survey April 2010
	Political Communications Jan 2012
	Political Polarization Survey 2014
	Political Typology Survey Group A and Group B 2011
	Post-Debate Survey Sept 2008
	Post-Election Survey Nov 2010
	Political Surveys: January, February, March, April, May, Mid-Oct, Late Oct, December 2008
	Political Surveys: January, February, March, June, July, December 2009
	Political Surveys: January, February, March, June, October, December 2010
	Political Surveys: January, February, April, May, June, July, August, December 2011
	Political Surveys: January, February, March, September, Early Oct, Late Oct, December 2012
	Political Surveys: January, February, March, June, July, September, October, December 2013
	Political Surveys: January, February, April, July, August 2014
	Racial Attitudes in America II Oct 2009
	Religion and Politics Survey Nov 2011
	Religion and Public Life Surveys: July 2008, Aug 2009, July 2010, June 2012
	Social Trends: Aging 2009
	Social Trends: Changing American Family 2010
	Social Trends: Work/Generation Gap 2009
	Trust in Government Survey March 2010
	Values Survey: March 2009, April 2012
	Voter Attitude Surveys: June 2008, June 2012
	Youth and Economy 2011

Appendix B: Survey Level Variables Calculated for each Geographical Area Included in the Study

Table B.1

Survey Level Variables Calculated Using Ca2001 Batch to Estimate the Adult Jewish Population in Toronto

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
12440402	6426	0.031	0.0002	0.002	0.027	0.036	1.00	7.66	5307
12450103	156	0.013	0.0065	0.009	0.003	0.051	1.94	70.28	294
12451003	151	0.019	0.0062	0.011	0.006	0.057	1.81	57.34	280
12454100	151	0.019	0.0064	0.011	0.006	0.059	1.83	57.33	277
12454101	164	0.037	0.0059	0.015	0.017	0.080	1.80	40.23	288
12454198	152	0.059	0.0062	0.019	0.031	0.110	1.45	32.34	221
12454199	149	0.062	0.0066	0.020	0.032	0.116	1.88	32.57	271
12454200	156	0.006	0.0064	0.006	0.001	0.044	1.83	99.70	277
12454201	154	0.044	0.0061	0.016	0.021	0.090	1.74	37.01	271
12454298	153	0.059	0.0062	0.019	0.031	0.109	1.46	32.35	224
12454299	152	0.032	0.0063	0.014	0.013	0.075	1.75	44.58	275
12454300	157	0.027	0.0087	0.015	0.009	0.080	1.73	57.18	191
12454301	154	0.019	0.0063	0.011	0.006	0.058	1.77	57.42	272
12454398	155	0.032	0.0062	0.014	0.013	0.075	1.45	44.01	225
12454399	147	0.012	0.0060	0.009	0.003	0.047	1.67	70.37	278
12454400	148	0.019	0.0063	0.011	0.006	0.058	1.80	57.30	274
12454401	149	0.066	0.0063	0.020	0.036	0.119	1.79	30.82	263
12454498	161	0.050	0.0059	0.017	0.025	0.096	1.40	34.48	227
12454499	178	0.008	0.0040	0.006	0.002	0.031	1.14	70.61	277
12470300	244	0.025	0.0040	0.010	0.011	0.054	1.77	40.42	416
124910504	393	0.000					0.00		460
124911000	289	0.032	0.0038	0.011	0.016	0.062	1.48	34.26	369
124930100	114	0.004	0.0043	0.004	0.001	0.030	0.53	100.21	120
124930200	116	0.010	0.0052	0.007	0.002	0.040	0.56	71.36	104
124930300	114	0.019	0.0075	0.012	0.006	0.064	0.91	62.71	117
124930400	119	0.019	0.0093	0.013	0.005	0.072	1.18	70.71	125
124930499	126	0.009	0.0093	0.009	0.001	0.064	1.23	99.60	126
124930500	108	0.037	0.0096	0.019	0.013	0.098	0.99	51.08	97
124930599	118	0.015	0.0092	0.012	0.003	0.068	1.18	77.88	121
124930600	126	0.028	0.0066	0.014	0.011	0.072	0.87	48.45	126
124930700	107	0.006	0.0056	0.006	0.001	0.039	0.62	100.10	109
124930799	116	0.045	0.0094	0.020	0.018	0.107	1.17	45.78	115
124930800	111	0.016	0.0070	0.011	0.004	0.057	0.79	66.37	110
124930899	112	0.015	0.0049	0.008	0.005	0.045	0.58	58.07	115
124930900	108	0.026	0.0087	0.015	0.008	0.079	1.09	57.31	118
124931000	115	0.023	0.0074	0.013	0.007	0.068	0.88	57.26	112
124931099	108	0.035	0.0114	0.020	0.011	0.104	1.31	56.68	110
124931100	114	0.036	0.0122	0.021	0.011	0.109	1.48	58.42	112
124931199	116	0.018	0.0090	0.013	0.004	0.070	1.12	70.41	120
124931200	117	0.044	0.0100	0.021	0.017	0.110	1.25	47.57	116
124950702	70	0.038	0.0123	0.022	0.012	0.112	0.90	56.88	69
124970197	243	0.038	0.0054	0.014	0.018	0.078	1.99	37.66	362
124990300	96	0.000					0.00		139

Table B.2

Survey Level Variables Calculated Using Ca2001 Batch to Estimate the Adult Jewish Population in Montreal

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
12440402	3162	0.024	0.0003	0.003	0.019	0.029	1.14	11.10	3843
12450103	164	0.007	0.0070	0.007	0.001	0.048	1.21	99.63	171
12451003	163	0.006	0.0064	0.006	0.001	0.044	1.13	99.70	170
12454100	155	0.021	0.0068	0.012	0.007	0.062	1.51	57.14	214
12454101	159	0.020	0.0067	0.012	0.007	0.061	1.53	57.36	223
12454198	174	0.006	0.0057	0.006	0.001	0.040	1.03	99.74	179
12454199	171	0.039	0.0056	0.015	0.018	0.081	1.44	38.19	248
12454200	146	0.000					0.00		200
12454201	171	0.005	0.0053	0.005	0.001	0.037	1.29	99.79	239
12454298	175	0.000					0.00		189
12454299	150	0.059	0.0072	0.021	0.030	0.116	1.79	34.83	239
12454300	146	0.006	0.0064	0.006	0.001	0.044	1.10	99.74	170
12454301	164	0.021	0.0068	0.012	0.007	0.062	1.61	57.09	232
12454398	155	0.032	0.0062	0.014	0.013	0.075	1.20	44.01	186
12454399	175	0.032	0.0064	0.014	0.013	0.075	1.65	44.85	254
12454400	163	0.011	0.0053	0.008	0.003	0.042	1.31	70.41	239
12454401	160	0.019	0.0065	0.011	0.006	0.059	1.52	57.98	225
12454498	164	0.000					0.00		184
12454499	166	0.004	0.0038	0.004	0.001	0.027	0.92	99.99	233
12470300	302	0.020	0.0047	0.010	0.008	0.051	1.85	48.38	374
124910504	328	0.003	0.0016	0.002	0.001	0.013	0.58	70.74	342
124911000	431	0.006	0.0033	0.005	0.002	0.026	1.08	71.85	313
124930100	107	0.017	0.0088	0.012	0.004	0.068	0.86	72.12	94
124930200	105	0.016	0.0084	0.012	0.004	0.064	0.82	72.05	94
124930300	107	0.009	0.0093	0.009	0.001	0.064	0.98	99.70	102
124930400	103	0.031	0.0083	0.016	0.011	0.084	0.84	51.91	99
124930499	105	0.014	0.0050	0.008	0.005	0.045	0.49	58.77	94
124930500	107	0.011	0.0112	0.011	0.002	0.076	1.21	99.50	104
124930599	101	0.010	0.0103	0.010	0.001	0.070	1.04	99.63	96
124930600	104	0.011	0.0105	0.011	0.001	0.072	1.15	99.57	107
124930700	97	0.013	0.0129	0.013	0.002	0.087	1.23	99.39	93
124930799	101	0.010	0.0104	0.010	0.001	0.071	1.12	99.64	103
124930800	100	0.000					0.00		89
124930899	104	0.010	0.0104	0.010	0.001	0.071	0.99	99.64	94
124930900	99	0.030	0.0085	0.016	0.010	0.083	0.82	53.78	90
124931000	113	0.004	0.0043	0.004	0.001	0.030	0.46	100.15	104
124931099	106	0.000					0.00		99
124931100	96	0.042	0.0168	0.026	0.012	0.138	1.68	63.65	92
124931199	106	0.020	0.0097	0.014	0.005	0.076	1.00	70.12	99
124931200	100	0.000					0.00		92
124950702	46	0.000					0.00		59
124970197	350	0.029	0.0037	0.010	0.015	0.058	1.23	35.50	329
124990300	136	0.000					0.00		137

Table B.3

Survey Level Variables Calculated Using Ca2001 Batch to Estimate the Adult Jewish Population in Vancouver

Survey Id	Sample Size	Proportion Jewish Adult Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
12440402	2681	0.004	0.0004	0.001	0.002	0.007	0.90	32.31	2193
12450103	88	0.012	0.0122	0.012	0.002	0.082	1.24	99.36	100
12451003	94	0.031	0.0104	0.018	0.010	0.094	1.17	57.62	105
12454100	88	0.021	0.0105	0.015	0.005	0.082	1.06	70.09	97
12454101	93	0.042	0.0105	0.021	0.016	0.109	1.10	49.72	99
12454198	89	0.011	0.0111	0.011	0.002	0.076	0.95	99.47	84
12454199	96	0.022	0.0107	0.015	0.005	0.083	1.16	70.06	107
12454200	95	0.000					0.00		105
12454201	96	0.022	0.0107	0.015	0.005	0.083	1.17	70.01	106
12454298	83	0.024	0.0118	0.017	0.006	0.091	0.95	69.87	79
12454299	91	0.008	0.0076	0.008	0.001	0.052	0.75	99.86	101
12454300	100	0.000					0.00		106
12454301	95	0.000					0.00		105
12454398	82	0.037	0.0118	0.021	0.012	0.107	1.04	56.69	86
12454399	92	0.015	0.0076	0.011	0.004	0.059	0.80	72.26	104
12454400	106	0.021	0.0102	0.015	0.005	0.079	1.16	70.18	110
12454401	89	0.009	0.0089	0.009	0.001	0.061	0.88	99.72	97
12454498	93	0.022	0.0105	0.015	0.005	0.082	0.93	69.97	88
12454499	95	0.012	0.0117	0.012	0.002	0.079	1.40	99.52	114
12470300	140	0.023	0.0064	0.012	0.008	0.064	1.20	52.31	178
124910504	180	0.009	0.0048	0.006	0.002	0.036	0.80	74.36	160
124911000	117	0.009	0.0090	0.009	0.001	0.062	1.35	99.62	143
124930100	56	0.000					0.00		53
124930200	48	0.000					0.00		41
124930300	55	0.000					0.00		53
124930400	49	0.012	0.0124	0.012	0.002	0.084	0.66	100.09	53
124930499	55	0.040	0.0191	0.028	0.010	0.146	1.09	69.50	53
124930500	56	0.009	0.0094	0.009	0.001	0.064	0.54	100.18	56
124930599	56	0.019	0.0184	0.019	0.003	0.122	1.04	99.33	53
124930600	57	0.000					0.00		52
124930700	52	0.000					0.00		49
124930799	57	0.010	0.0095	0.010	0.001	0.065	0.59	100.19	59
124930800	56	0.036	0.0174	0.025	0.009	0.133	1.08	69.76	59
124930899	47	0.022	0.0213	0.021	0.003	0.139	1.00	99.22	45
124930900	43	0.000					0.00		39
124931000	58	0.000					0.00		53
124931099	52	0.000					0.00		43
124931100	59	0.000					0.00		58
124931199	52	0.000					0.00		50
124931200	53	0.011	0.0110	0.011	0.002	0.075	0.56	100.01	49
124950702	26	0.000					0.00		26
124970197	213	0.012	0.0035	0.007	0.004	0.035	0.73	52.80	212
124990300	59	0.000					0.00		84

Table B.4

Survey Level Variables Calculated Using Ca2001 Batch to Estimate the Adult Jewish Population in Quebec

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
12450101	409	0.011	0.0036	0.006	0.003	0.033	1.52	58.19	410
12450103	420	0.003	0.0030	0.003	0.000	0.021	1.21	99.85	402
12451003	411	0.003	0.0028	0.003	0.000	0.020	1.13	99.87	385
12454100	413	0.011	0.0036	0.006	0.004	0.034	1.51	57.38	404
12454101	406	0.011	0.0037	0.006	0.004	0.034	1.53	57.58	405
12454198	428	0.002	0.0023	0.002	0.000	0.016	1.03	99.92	438
12454199	433	0.023	0.0031	0.008	0.011	0.046	1.38	36.72	448
12454200	401	0.000					0.00		389
12454201	428	0.003	0.0030	0.003	0.000	0.021	1.28	99.86	429
12454298	442	0.000					0.00		452
12454299	414	0.032	0.0040	0.011	0.016	0.064	1.78	35.27	440
12454300	386	0.002	0.0024	0.002	0.000	0.017	1.10	99.92	448
12454301	413	0.014	0.0035	0.007	0.005	0.036	1.50	50.92	417
12454398	417	0.015	0.0026	0.006	0.007	0.033	1.14	41.26	442
12454399	440	0.021	0.0033	0.008	0.010	0.046	1.53	39.28	458
12454400	422	0.006	0.0029	0.004	0.001	0.023	1.30	70.51	436
12454401	409	0.011	0.0036	0.006	0.003	0.033	1.52	58.19	410
12454498	426	0.000					0.00		442
12454499	420	0.004	0.0019	0.003	0.001	0.015	0.84	71.14	417
12470300	703	0.008	0.0020	0.004	0.003	0.022	1.85	48.72	895
124910504	943	0.006	0.0009	0.002	0.003	0.013	0.90	37.41	960
124911000	1130	0.010	0.0013	0.004	0.005	0.020	1.09	35.11	825
124930100	225	0.011	0.0038	0.006	0.003	0.034	0.79	59.14	203
124930200	230	0.007	0.0035	0.005	0.002	0.027	0.82	72.24	227
124930300	233	0.004	0.0043	0.004	0.001	0.030	0.98	99.88	223
124930400	215	0.015	0.0040	0.008	0.005	0.041	0.83	52.07	206
124930499	232	0.006	0.0021	0.004	0.002	0.019	0.49	58.74	221
124930500	230	0.005	0.0052	0.005	0.001	0.036	1.21	99.80	228
124930599	228	0.005	0.0046	0.005	0.001	0.032	1.04	99.86	215
124930600	228	0.005	0.0049	0.005	0.001	0.034	1.15	99.83	228
124930700	218	0.006	0.0055	0.006	0.001	0.039	1.23	99.78	218
124930799	222	0.005	0.0049	0.005	0.001	0.034	1.12	99.84	220
124930800	220	0.000					0.00		210
124930899	222	0.007	0.0039	0.005	0.002	0.030	0.85	72.69	214
124930900	223	0.013	0.0038	0.007	0.004	0.037	0.81	54.07	205
124931000	238	0.002	0.0020	0.002	0.000	0.014	0.46	100.11	229
124931099	220	0.000					0.00		214
124931100	216	0.019	0.0079	0.012	0.005	0.066	1.69	64.61	203
124931199	230	0.009	0.0044	0.006	0.002	0.035	1.00	70.50	222
124931200	220	0.000					0.00		218
124950503	112	0.027	0.0177	0.022	0.005	0.124	1.97	81.39	113
124950702	111	0.000					0.00		110
124970197	898	0.012	0.0014	0.004	0.006	0.024	1.21	33.78	855
124990300	440	0.000					0.00		416

Table B.5

Survey Level Variables Calculated Using Ca2001 Batch to Estimate the Adult Jewish Population in Ontario

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
12450101	484	0.031	0.0026	0.009	0.017	0.054	1.71	28.89	634.52
12450103	491	0.006	0.0029	0.004	0.001	0.023	1.94	70.50	661.01
12451003	485	0.008	0.0027	0.005	0.003	0.025	1.80	57.57	640.09
12454100	484	0.012	0.0024	0.005	0.005	0.028	1.63	45.49	650.60
12454101	507	0.017	0.0025	0.007	0.008	0.036	1.72	38.14	671.63
12454198	449	0.025	0.0021	0.007	0.014	0.045	1.36	28.92	623.30
12454199	460	0.032	0.0027	0.009	0.018	0.056	1.78	29.11	642.43
12454200	502	0.006	0.0022	0.004	0.002	0.020	1.48	58.73	658.95
12454201	499	0.018	0.0026	0.007	0.009	0.038	1.74	37.43	655.29
12454298	456	0.031	0.0023	0.008	0.018	0.053	1.50	27.35	632.79
12454299	474	0.022	0.0022	0.007	0.012	0.041	1.44	31.48	646.72
12454300	552	0.009	0.0024	0.005	0.003	0.025	1.63	50.62	662.10
12454301	501	0.011	0.0024	0.005	0.005	0.028	1.59	45.65	661.06
12454398	466	0.020	0.0023	0.007	0.010	0.038	1.46	33.88	628.90
12454399	471	0.012	0.0021	0.005	0.005	0.026	1.34	42.38	650.64
12454400	541	0.012	0.0022	0.005	0.005	0.028	1.51	42.45	667.40
12454401	484	0.031	0.0026	0.009	0.017	0.054	1.71	28.89	634.52
12454498	450	0.031	0.0024	0.009	0.018	0.054	1.59	27.92	638.71
12454499	509	0.006	0.0016	0.003	0.002	0.017	1.12	50.09	660.53
12470300	754	0.012	0.0016	0.004	0.006	0.024	2.12	37.43	1230.80
124910504	1143	0.013	0.0009	0.003	0.008	0.022	1.28	26.34	1340.46
124911000	856	0.017	0.0012	0.004	0.010	0.029	1.39	26.06	1131.58
124930100	308	0.002	0.0016	0.002	0.000	0.011	0.52	100.11	326.00
124930200	308	0.004	0.0019	0.003	0.001	0.015	0.55	71.34	288.26
124930300	312	0.007	0.0028	0.004	0.002	0.024	0.91	62.89	318.61
124930400	316	0.016	0.0044	0.008	0.006	0.045	1.42	51.50	321.11
124930499	327	0.007	0.0036	0.005	0.002	0.028	1.17	70.58	314.53
124930500	296	0.017	0.0031	0.007	0.007	0.039	0.89	42.78	279.28
124930599	318	0.009	0.0035	0.006	0.003	0.030	1.12	61.65	308.38
124930600	312	0.011	0.0026	0.005	0.004	0.029	0.86	48.68	320.08
124930700	293	0.002	0.0021	0.002	0.000	0.015	0.62	100.06	290.60
124930799	313	0.017	0.0036	0.008	0.007	0.041	1.17	46.38	310.59
124930800	302	0.009	0.0027	0.005	0.003	0.026	0.82	55.38	301.73
124930899	307	0.007	0.0018	0.004	0.003	0.019	0.56	50.29	308.78
124930900	296	0.010	0.0034	0.006	0.003	0.032	1.08	57.60	299.20
124931000	306	0.017	0.0029	0.007	0.007	0.037	0.92	41.77	303.54
124931099	300	0.016	0.0040	0.008	0.006	0.042	1.27	49.72	308.06
124931100	310	0.017	0.0044	0.008	0.006	0.045	1.39	51.31	303.52
124931199	316	0.007	0.0035	0.005	0.002	0.028	1.11	70.70	310.27
124931200	316	0.017	0.0039	0.008	0.006	0.042	1.25	48.24	309.02
124950503	168	0.000					0.00		178.81
124950702	161	0.022	0.0054	0.011	0.008	0.057	0.88	49.63	158.52
124970197	833	0.017	0.0016	0.005	0.009	0.030	2.03	31.35	1247.27
124990300	412	0.008	0.0029	0.005	0.003	0.026	1.82	58.55	620.86

Table B.6

Survey Level Variables Calculated Using Ca2001 Batch to Estimate the Adult Jewish Population in in Manitoba

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
12450101	110	0.006	0.0064	0.006	0.001	0.044	0.41	99.90	63
12450103	112	0.009	0.0087	0.009	0.001	0.060	0.54	99.61	62
12451003	115	0.006	0.0056	0.006	0.001	0.039	0.35	99.95	59
12454100	111	0.035	0.0093	0.018	0.012	0.093	0.62	52.05	63
12454101	99	0.000					0.00		61
12454198	101	0.030	0.0097	0.017	0.010	0.089	0.68	56.88	68
12454199	121	0.017	0.0089	0.012	0.004	0.068	0.62	72.40	69
12454200	104	0.000					0.00		59
12454201	108	0.020	0.0098	0.014	0.005	0.076	0.65	70.27	65
12454298	115	0.026	0.0084	0.015	0.008	0.078	0.60	57.00	69
12454299	105	0.071	0.0122	0.030	0.031	0.156	0.84	41.46	65
12454300	117	0.029	0.0113	0.018	0.008	0.095	0.77	62.52	66
12454301	115	0.000					0.00		67
12454398	111	0.027	0.0087	0.015	0.009	0.080	0.61	56.98	69
12454399	105	0.075	0.0139	0.032	0.031	0.168	0.93	43.22	63
12454400	104	0.000					0.00		63
12454401	110	0.006	0.0064	0.006	0.001	0.044	0.41	99.90	63
12454498	110	0.046	0.0087	0.020	0.019	0.105	0.61	43.69	67
12454499	115	0.012	0.0121	0.012	0.002	0.082	0.81	99.64	64
12470300	324	0.017	0.0043	0.009	0.006	0.045	0.54	50.48	119
124910504	196	0.000					0.00		148
124911000	104	0.005	0.0051	0.005	0.001	0.036	0.65	100.12	121
124930100	36	0.000					0.00		37
124930200	33	0.000					0.00		36
124930300	35	0.000					0.00		36
124930400	34	0.000					0.00		35
124930499	32	0.000					0.00		33
124930500	27	0.064	0.0350	0.047	0.014	0.244	1.02	73.94	27
124930599	31	0.000					0.00		32
124930600	33	0.000					0.00		31
124930700	34	0.000					0.00		33
124930799	33	0.000					0.00		30
124930800	34	0.000					0.00		33
124930899	34	0.017	0.0172	0.017	0.002	0.114	0.61	99.96	34
124930900	33	0.021	0.0204	0.020	0.003	0.134	0.57	99.69	27
124931000	30	0.000					0.00		29
124931099	37	0.000					0.00		33
124931100	35	0.000					0.00		32
124931199	32	0.000					0.00		30
124931200	32	0.000					0.00		29
124950503	16	0.000					0.00		17
124950702	13	0.000					0.00		13
124970197	175	0.019	0.0051	0.010	0.007	0.051	0.66	52.36	128
124990300	66	0.022	0.0216	0.022	0.003	0.141	1.81	98.65	81

Table B.7

Survey Level Variables Calculated Using Ca2001 Batch to Estimate the Adult Jewish Population in in British Columbia

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
12450101	198	0.004	0.0039	0.004	0.001	0.027	0.87	99.90	223
12450103	198	0.011	0.0055	0.008	0.003	0.044	1.27	70.32	226
12451003	201	0.023	0.0046	0.010	0.009	0.054	1.07	45.04	221
12454100	196	0.010	0.0048	0.007	0.002	0.038	1.06	70.47	215
12454101	211	0.023	0.0046	0.010	0.009	0.055	1.06	44.96	221
12454198	196	0.010	0.0049	0.007	0.002	0.038	1.00	70.49	203
12454199	205	0.010	0.0050	0.007	0.002	0.039	1.16	70.49	233
12454200	205	0.005	0.0054	0.005	0.001	0.038	1.26	99.73	227
12454201	207	0.010	0.0050	0.007	0.003	0.040	1.17	70.44	228
12454298	192	0.020	0.0048	0.010	0.007	0.051	0.98	49.59	200
12454299	202	0.012	0.0067	0.009	0.003	0.050	1.51	76.12	228
12454300	224	0.000					0.00		231
12454301	210	0.000					0.00		232
12454398	190	0.020	0.0049	0.010	0.008	0.052	1.03	49.57	205
12454399	200	0.007	0.0035	0.005	0.002	0.027	0.79	72.41	230
12454400	226	0.010	0.0048	0.007	0.002	0.038	1.16	70.62	236
12454401	198	0.004	0.0039	0.004	0.001	0.027	0.87	99.90	223
12454498	202	0.009	0.0045	0.006	0.002	0.036	0.94	70.44	207
12454499	208	0.006	0.0059	0.006	0.001	0.041	1.39	99.72	227
12470300	341	0.011	0.0025	0.005	0.004	0.028	1.12	47.22	429
124910504	538	0.005	0.0018	0.003	0.002	0.016	0.86	59.89	461
124911000	329	0.005	0.0027	0.004	0.001	0.021	1.13	74.42	396
124930100	116	0.000					0.00		111
124930200	103	0.000					0.00		92
124930300	112	0.000					0.00		105
124930400	108	0.006	0.0058	0.006	0.001	0.040	0.65	100.04	113
124930499	114	0.020	0.0098	0.014	0.005	0.077	1.09	70.08	104
124930500	118	0.014	0.0079	0.011	0.003	0.060	0.89	74.09	109
124930599	117	0.009	0.0091	0.009	0.001	0.062	1.04	99.67	109
124930600	114	0.000					0.00		103
124930700	112	0.000					0.00		108
124930799	115	0.005	0.0053	0.005	0.001	0.037	0.59	100.06	107
124930800	111	0.019	0.0092	0.013	0.005	0.072	1.07	70.20	113
124930899	102	0.010	0.0100	0.010	0.001	0.068	0.99	99.63	97
124930900	100	0.013	0.0124	0.013	0.002	0.084	1.22	99.38	94
124931000	115	0.008	0.0084	0.008	0.001	0.058	0.92	99.73	106
124931099	108	0.000					0.00		96
124931100	119	0.000					0.00		115
124931199	107	0.000					0.00		101
124931200	107	0.005	0.0055	0.005	0.001	0.038	0.56	100.07	99
124950503	58	0.000					0.00		60
124950702	58	0.000					0.00		58
124970197	413	0.008	0.0017	0.004	0.003	0.020	0.69	47.09	402
124990300	160	0.000					0.00		222

Table B.8

Survey Level Variables Calculated Using Ca2011 Batch to Estimate the Adult Jewish Population in Montreal

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
1245411	167	0.023	0.0059	0.012	0.009	0.061	1.00	50.08	158
12454306	189	0.009	0.0085	0.008	0.001	0.058	1.53	99.46	172
12454406	190	0.011	0.0054	0.008	0.003	0.042	1.01	71.32	178
124810112	120	0.013	0.0070	0.010	0.003	0.054	0.86	71.93	123
124810114	147	0.034	0.0066	0.015	0.014	0.079	1.00	44.03	147
124910511	173	0.039	0.0083	0.018	0.016	0.094	1.17	46.02	135
124911008	301	0.005	0.0029	0.004	0.001	0.022	0.92	74.45	312
124950407	93	0.000					0.00		93
124950509	73	0.022	0.0211	0.021	0.003	0.138	1.43	98.67	66
124990206	185	0.027	0.0069	0.014	0.010	0.071	1.43	50.88	199

Table B.9

Survey Level Variables Calculated Using Ca2011 Batch to Estimate the Adult Jewish Population in in Toronto

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
1245411	182	0.026	0.0052	0.012	0.011	0.061	1.23	44.72	220
12454306	231	0.017	0.0042	0.008	0.006	0.044	1.27	50.27	286
12454406	237	0.047	0.0043	0.014	0.026	0.084	1.39	30.16	295
124810112	185	0.014	0.0048	0.008	0.004	0.044	0.90	58.76	185
124810114	198	0.060	0.0058	0.019	0.032	0.109	1.22	31.12	200
124910511	132	0.074	0.0100	0.027	0.035	0.148	1.79	36.72	165
124911008	315	0.008	0.0023	0.004	0.003	0.023	0.88	52.84	363
124950407	143	0.021	0.0069	0.012	0.007	0.063	1.00	57.16	143
124950509	107	0.059	0.0126	0.027	0.023	0.141	1.48	46.28	110
124990206	156	0.006	0.0064	0.006	0.001	0.045	1.72	99.73	261

Table B.10

Survey Level Variables Calculated Using Ca2011 Batch to Estimate the Adult Jewish Population in Vancouver

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
1245411	45	0.000					0.00		65
12454306	92	0.019	0.0093	0.013	0.005	0.073	1.11	70.19	112
12454406	92	0.000					0.00		104
124810112	65	0.018	0.0178	0.018	0.003	0.118	1.18	99.05	66
124810114	86	0.000					0.00		86
124910511	175	0.000					0.00		192
124911008	150	0.034	0.0093	0.018	0.012	0.092	1.19	52.25	120
124950407	44	0.000					0.00		44
124950509	38	0.024	0.0234	0.024	0.003	0.152	0.74	99.03	31
124990206	89	0.000					0.00		103

Table B.11

Survey Level Variables Calculated Using Ca2011 Batch to Estimate the Adult Jewish Population in Quebec

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
1245411	340	0.014	0.0029	0.007	0.006	0.035	0.96	45.13	306
12454306	445	0.004	0.0036	0.004	0.001	0.025	1.53	99.79	410
12454406	439	0.005	0.0024	0.003	0.001	0.019	1.01	71.52	409
124530312	3493	0.009	0.0004	0.002	0.006	0.014	1.91	20.36	4941
124530313	4582	0.007	0.0003	0.001	0.005	0.010	1.59	20.49	5437
124710909	307	0.007	0.0024	0.004	0.002	0.022	0.52	58.24	207
124810112	315	0.005	0.0028	0.004	0.001	0.022	0.86	72.14	310
124810114	318	0.019	0.0030	0.008	0.008	0.041	0.99	40.52	320
124910511	944	0.007	0.0015	0.003	0.003	0.017	1.18	46.95	770
124911008	781	0.007	0.0011	0.003	0.003	0.015	0.89	39.93	782
124921112	88	0.000					0.00		145
124950407	204	0.000					0.00		204
124950509	161	0.008	0.0080	0.008	0.001	0.055	1.44	99.66	178
124980409	403	0.000					0.00		310
124980707	225	0.000					0.00		196
124981108	379	0.021	0.0057	0.011	0.008	0.058	1.70	51.51	277
124990206	466	0.011	0.0029	0.006	0.004	0.030	1.43	51.24	474

Table B.12

Survey Level Variables Calculated Using Ca2011 Batch to Estimate the Adult Jewish Population in Ontario

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
1245411	407	0.011	0.0023	0.005	0.005	0.028	1.22	45.01	496
12454306	535	0.009	0.0018	0.004	0.004	0.021	1.25	45.04	665
12454406	553	0.022	0.0019	0.006	0.012	0.039	1.40	29.20	693
124530312	6081	0.015	0.0002	0.002	0.012	0.019	1.46	10.82	8015
124530313	5968	0.017	0.0002	0.002	0.013	0.021	1.98	11.58	8774
124710909	312	0.008	0.0026	0.004	0.002	0.024	0.84	58.94	303
124810112	450	0.015	0.0045	0.008	0.005	0.043	2.14	54.53	478
124810114	491	0.027	0.0023	0.008	0.015	0.047	1.17	29.20	508
124910511	934	0.014	0.0015	0.005	0.008	0.027	1.74	31.92	1164
124911008	905	0.013	0.0013	0.004	0.007	0.024	1.44	31.45	1060
124921112	249	0.006	0.0055	0.006	0.001	0.038	1.55	99.72	281
124950407	332	0.009	0.0030	0.005	0.003	0.028	1.00	57.51	332
124950509	260	0.027	0.0059	0.013	0.010	0.066	1.49	47.28	244
124980409	549	0.038	0.0041	0.012	0.020	0.072	2.40	32.75	526
124980707	334	0.037	0.0028	0.010	0.022	0.064	0.95	27.39	316
124981108	421	0.019	0.0022	0.007	0.010	0.038	1.11	33.79	465
124990206	449	0.005	0.0023	0.003	0.001	0.019	1.68	70.61	712

Table B.13

Survey Level Variables Calculated Using Ca2011 Batch to Estimate the Adult Jewish Population in Manitoba

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
1245411	84	0.010	0.0096	0.007	0.001	0.065	0.49	99.84	49
12454306	112	0.000		0.004			0.00		62
12454406	111	0.019	0.0100	0.003	0.005	0.076	0.67	72.34	63
124530312	1146	0.006	0.0013	0.002	0.003	0.015	0.92	45.04	717
124530313	1551	0.009	0.0014	0.001	0.004	0.019	1.08	39.29	783
124710909	26	0.000		0.004			0.00		31
124810112	47	0.047	0.0227	0.004	0.012	0.171	0.97	69.43	41
124810114	61	0.013	0.0129	0.008	0.002	0.087	0.65	99.72	50
124910511	134	0.024	0.0128	0.003	0.006	0.096	1.39	73.35	106
124911008	130	0.004	0.0040	0.003	0.001	0.028	0.43	100.05	103
124921112	65	0.018	0.0179		0.003	0.118	0.76	99.25	42
124950407	29	0.000					0.00		29
124950509	23	0.000		0.008			0.00		27
124980409	69	0.000					0.00		49
124980707	16	0.000					0.00		25
124981108	42	0.000		0.011			0.00		41
124990206	54	0.017	0.0172	0.006	0.002	0.114	0.91	99.35	51

Table B.14

Survey Level Variables Calculated Using Ca2011 Batch to Estimate the Adult Jewish Population in British Columbia

Survey Id	Sample Size	Proportion Adult Jewish Population	Var	SE	CIL	CIH	DEFF	CV	Weighted Sample Size
1245411	134	0.000					0.00		166
12454306	206	0.009	0.0043	0.006	0.002	0.034	1.10	70.48	245
12454406	210	0.004	0.0045	0.004	0.001	0.031	1.12	99.83	239
124530312	2620	0.005	0.0005	0.002	0.003	0.010	1.32	29.94	2749
124530313	2746	0.008	0.0011	0.003	0.004	0.017	3.39	36.48	3071
124710909	65	0.013	0.0127	0.013	0.002	0.086	1.45	99.62	108
124810112	161	0.007	0.0072	0.007	0.001	0.050	1.18	99.65	165
124810114	170	0.000					0.00		179
124910511	358	0.000					0.00		402
124911008	437	0.011	0.0032	0.006	0.004	0.032	1.19	52.95	360
124921112	88	0.000					0.00		130
124950407	113	0.000					0.00		113
124950509	75	0.011	0.0113	0.011	0.002	0.077	0.74	99.76	64
124980409	174	0.019	0.0041	0.009	0.008	0.047	0.86	46.12	192
124980707	89	0.011	0.0110	0.011	0.002	0.075	1.33	99.58	116
124981108	152	0.000					0.00		147
124990206	206	0.010	0.0103	0.010	0.001	0.071	2.54	99.35	240

Appendix C: Information Concerning Software and Syntaxes

The purpose of this appendix is to provide information related to the software and the syntaxes used in this study.

Meta-Analysis of Complex Surveys (MACS)

I used STATA statistical software to generate MACS cross-survey estimates. To this end, I wrote a series of STATA do-files. The first set of do-files consisted of STATA survey commands to generate the weighted estimates of the proportion of the adult Jewish population in each geographical area included in the study for every survey in the US and Canadian batches. Along with the estimates of the proportions of the adult Jewish population, I also calculated the variances, design effects, standard errors, and coefficient of variations associated with each estimate. At the end of this appendix there is an example of the STATA commands used to generate these variables for each survey. Appendix B includes the survey statistical variables calculated for each Canadian geographical area included in the study. The information generated by this syntax was also used for the PDCS estimate calculations.

To verify this STATA syntax, I replicated the calculation of the estimates of the proportions of the adult Jewish population in the chosen geographical areas for one of the surveys included in Ca2001 Batch with the SPSS statistical package. Given that the SPSS basic module is not appropriate for the analysis of standard errors of complex surveys, I only verified the point estimates.

The second set of do-files generated the survey-level specific weights for every geographical area in each survey using the estimated variances and design effects generated by the first set of do-files according to the specifications of the MCAS approach.

The third set generated the proportions of the adult Jewish population in each geographical area (for each batch), along with associated variances of this population (using survey estimates and survey-level specific weights). To verify the estimates generated with this set of do-files, I replicated the process (for a specific geographical location) in Excel. Table C1 presents the calculations made in Excel to estimate the Ca2001 proportion of the adult Jewish population in Toronto.

Table C1

2001 MACS Estimates of the Proportions of the Adult Jewish Population in Toronto - Excel calculations

Survid	N	Prop (jw=1)	Var	CIL	CIH	DEFF	CV	Sum Weight	SE	w = (1/var)/deff	prop * w
12440402	6426	0.0312	0.0002	0.027	0.036	1.0	8	5307	0.002	5464.29	170.28
12450103	156	0.0131	0.0065	0.003	0.051	1.9	70	294	0.009	79.60	1.04
12451003	151	0.0187	0.0062	0.006	0.057	1.8	57	280	0.011	89.97	1.68
12454100	151	0.0194	0.0064	0.006	0.059	1.8	57	277	0.011	85.67	1.66
12454101	164	0.0368	0.0059	0.017	0.080	1.8	40	288	0.015	93.50	3.44
12454198	152	0.0592	0.0062	0.031	0.110	1.4	32	221	0.019	111.48	6.60
12454199	149	0.0621	0.0066	0.032	0.116	1.9	33	271	0.020	80.70	5.01
12454200	156	0.0065	0.0064	0.001	0.044	1.8	100	277	0.006	84.95	0.55
12454201	154	0.0445	0.0061	0.021	0.090	1.7	37	271	0.016	94.15	4.19
12454298	153	0.0588	0.0062	0.031	0.109	1.5	32	224	0.019	111.32	6.55
12454299	152	0.0316	0.0063	0.013	0.075	1.7	45	275	0.014	90.96	2.88
12454300	157	0.0266	0.0087	0.009	0.080	1.7	57	191	0.015	66.25	1.76
12454301	154	0.0192	0.0063	0.006	0.058	1.8	57	272	0.011	88.90	1.71
12454398	155	0.0323	0.0062	0.013	0.075	1.4	44	225	0.014	110.61	3.57
12454399	147	0.0121	0.0060	0.003	0.047	1.7	70	278	0.009	100.06	1.21
12454400	148	0.0193	0.0063	0.006	0.058	1.8	57	274	0.011	87.60	1.69
12454401	149	0.0660	0.0063	0.036	0.119	1.8	31	263	0.020	89.18	5.89
12454498	161	0.0497	0.0059	0.025	0.096	1.4	34	227	0.017	120.91	6.01
12454499	178	0.0079	0.0040	0.002	0.031	1.1	71	277	0.006	221.07	1.75
12470300	244	0.0245	0.0040	0.011	0.054	1.8	40	416	0.010	141.12	3.46
124910504	393	0.0043	0.0043	0.001	0.030	0.5	100	120	0.004	440.16	1.89
124911000	289	0.0320	0.0038	0.016	0.062	1.5	34	369	0.011	179.16	5.74
124930100	114	0.0043	0.0043	0.001	0.030	0.5	100	120	0.004	440.16	1.89
124930200	116	0.0101	0.0052	0.002	0.040	0.6	71	104	0.007	346.81	3.52
124930300	114	0.0190	0.0075	0.006	0.064	0.9	63	117	0.012	146.43	2.79
124930400	119	0.0186	0.0093	0.005	0.072	1.2	71	125	0.013	91.38	1.70

124930499	126	0.0093	0.0093	0.001	0.064	1.2	100	126	0.009	88.00	0.82
124930500	108	0.0369	0.0096	0.013	0.098	1.0	51	97	0.019	104.48	3.86
124930599	118	0.0152	0.0092	0.003	0.068	1.2	78	121	0.012	91.69	1.40
124930600	126	0.0282	0.0066	0.011	0.072	0.9	48	126	0.014	173.84	4.90
124930700	107	0.0056	0.0056	0.001	0.039	0.6	100	109	0.006	290.61	1.61
124930799	116	0.0448	0.0094	0.018	0.107	1.2	46	115	0.020	90.77	4.06
124930800	111	0.0159	0.0070	0.004	0.057	0.8	66	110	0.011	179.89	2.86
124930899	112	0.0146	0.0049	0.005	0.045	0.6	58	115	0.008	351.69	5.12
124930900	108	0.0264	0.0087	0.008	0.079	1.1	57	118	0.015	106.02	2.80
124931000	115	0.0227	0.0074	0.007	0.068	0.9	57	112	0.013	152.88	3.47
124931099	108	0.0354	0.0114	0.011	0.104	1.3	57	110	0.020	66.87	2.37
124931100	114	0.0359	0.0122	0.011	0.109	1.5	58	112	0.021	55.26	1.98
124931199	116	0.0181	0.0090	0.004	0.070	1.1	70	120	0.013	100.14	1.81
124931200	117	0.0443	0.0100	0.017	0.110	1.3	48	116	0.021	79.68	3.53
124950702	70	0.0380	0.0123	0.012	0.112	0.9	57	69	0.022	90.80	3.45
124970197	243	0.0379	0.0054	0.018	0.078	2.0	38	362	0.014	93.35	3.54
124990300	96	0.0043	0.0043	0.001	0.030	0.5	100	120	0.004	440.16	1.89

sum(w)	sum(prop*w)	estimate=(sum(prop*w) /sum(w))
11612.54	297.93	0.026

Pooled Design-Based Cross-Survey (PDCS)

I used STATA statistical software to generate PDCS cross-survey estimates. Similar to MACS, I wrote three sets of do-files to calculate the different variables needed to generate the PDCS cross-survey estimates of the proportion of the adult Jewish population.

The first set of do-files calculated coefficient k for every geographical area and for every survey included in each batch. As I described in Chapters 2 and 3, coefficient k is a function of the sample sizes and of the coefficients of variation of the different surveys that are used to generate an estimate for a specific geographical area in a batch. I used Excel to replicate the calculation of k coefficients for the Toronto metropolitan area for batch Ca2011.

The second set of do-files calculated a new weight for every record included in all the surveys in the three batches. The new weight was calculated as a function of coefficient k (of the survey and geographical area) and the original weight of the record.

The last set generated the estimates of the proportions of the adult Jewish population in each geographical area for every batch using the new weights calculated in the previous step. Survey STATA commands were then used to obtain the point estimates and Survey STATA jackknife commands to generate variances. An example of a jackknife STATA command may be found at the end of this appendix.

Bayesian Multilevel Regression with Post-Stratification

To generate the BMRP cross-survey estimates of the proportions of the adult Jewish population, I used R and STAN languages. R programs were used to set up the data, run STAN programs, and post-stratify the Bayesian results so as to generate the cross-survey estimates. Stan is an open-source C++ program that performs Bayesian inference. Gellman, Lee, and Guo (2015) explain the use of this program as follows:, “to use Stan, a user writes a Stan program that directly computes the log-posterior density. This code is then compiled and run along with data. The result is a set of posterior simulations of the parameters in the model (or a point estimate, if Stan is set to optimize)” (Gellman, Lee, and Guo, 2015, p. 1).

As described in the methodological section, BMRP estimates are generated in two steps. In the first step, the Bayesian multilevel logistic regression model is built for each scenario. At the end of this appendix there is an example of a section of a STAN program. A series of programs written on R were used to carry out the second step, that is, post-stratification. These R programs processed the outputs from the STAN analyses along with census counts to generate cross-survey estimates.

Both the STAN and the R post-stratification programs are similar to the programs used for the past few years by SSRI researchers to estimate the proportion of the U.S. Jewish population.

STATA and Stan Syntax

STATA Commands to obtain survey level estimates – example.

Do file - Estimates by Survey

#Declare survey design for dataset
svyset _n [pweight=*weight*]

Proportion of Jewish population for survey *surv* metropolitan area for the Canadian metropolitan area o*met*
svy, subpop(if survid==*surv* & metro==*met*):prop jw

Obtaining stored results

# subpopulation observations
e(N_sub)

# Estimate of subpopulation size
e(N_subpop)

#Design effects
estat effects

#Coefficient of variation
estat cv

Jackknife STATA commands –example.

#Declare survey design for dataset
svyset _n [pweight=*weight*], vce(jackknife) singleunit(missing)

```
#Estimates using jackknife
svy jackknife, subpop(if metro==met) : proportion jw
```

Stan syntax - Canadian Metro areas using Ca2001 data example.

```
# Stan program - Canadian Metro areas using Ca2001 data
data {
  int<lower=0> N;
  int<lower=0> n_surv;
  int<lower=0,upper=1> curreljw[N];
  int<lower=0,upper=1> age1[N];
…
  int<lower=0,upper=1> metc3[N];
  int<lower=0, upper=n_surv> survs[N];
}

parameters {
    real   b_cons;
    real   b_age1;
…
    real   b_e5xm3;
  vector[n_surv] b_surv;
  real<lower=0,upper=2.5> sigma_surv;
}

model {
  vector[N] p;
  b_cons  ~  normal(-4.82,100);
  b_age1  ~  normal(-1.34,100);
  b_e5xm3  ~  normal(-0.23,100);
…
  b_surv ~ normal(0, sigma_surv);
  sigma_surv ~ cauchy(0, 2.5);

 for (i in 1:N)
     p[i] = fmax(0, fmin(1, inv_logit(b_cons +
     b_age1*age1[i] + b_age2*age2[i] + b_age3*age3[i] + b_age4*age4[i] +
     b_edu1*edu1[i] + b_edu2*edu2[i] + b_edu4*edu4[i] + b_edu5*edu5[i] +
     b_metc1*metc1[i] + b_metc2*metc2[i] + b_metc3*metc3[i] +
     b_a1xe1*age1[i]*edu1[i] + b_a2xe1*age2[i]*edu1[i] +
     b_a3xe1*age3[i]*edu1[i] + b_a4xe1*age4[i]*edu1[i] +
     b_a1xe2*age1[i]*edu2[i] + b_a2xe2*age2[i]*edu2[i] +
     b_a3xe2*age3[i]*edu2[i] + b_a4xe2*age4[i]*edu2[i] +
     b_a1xe4*age1[i]*edu4[i] + b_a2xe4*age2[i]*edu4[i] +
     b_a3xe4*age3[i]*edu4[i] + b_a4xe4*age4[i]*edu4[i] +
     b_a1xe5*age1[i]*edu5[i] + b_a2xe5*age2[i]*edu5[i] +
     b_a3xe5*age3[i]*edu5[i] + b_a4xe5*age4[i]*edu5[i] +
     b_e1xm1*edu1[i]*metc1[i] +  b_e1xm2*edu1[i]*metc2[i] +
     b_e1xm3*edu1[i]*metc3[i] +
     b_e2xm1*edu2[i]*metc1[i] +  b_e2xm2*edu2[i]*metc2[i] +
     b_e2xm3*edu2[i]*metc3[i] +
     b_e4xm1*edu4[i]*metc1[i] +  b_e4xm2*edu4[i]*metc2[i] +
     b_e4xm3*edu4[i]*metc3[i] +
```

```
    b_e5xm1*edu5[i]*metc1[i] +  b_e5xm2*edu5[i]*metc2[i] +
    b_e5xm3*edu5[i]*metc3[i] +
    b_surv[survs[i]])));
  curreljw ~ bernoulli(p);
}
```