# Sociodemographic representativeness of severe acute respiratory syndrome coronavirus-2 serosurveillance studies with diverse recruitment strategies

Matthew J. Knight

Department of Epidemiology, Biostatistics and Occupational Health

School of Population and Global Health

Faculty of Medicine

McGill University

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Epidemiology

# Table of Contents

# Abstract

**Background.** Serological testing was a key component of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) surveillance. Social distancing interventions, resource limitations, and the need for timely data necessitated serosurveillance studies use creative recruitment strategies, which likely influenced study representativeness. Select demographic subgroups, such as racialized communities, have been underrepresented in previous SARS-CoV-2 serology studies relative to the general population. Characterizing representativeness is crucial to identify gaps in sampling coverage and to evaluate intervention impact on health inequities.

**Objective.** To assess the sociodemographic representativeness of Canadian SARS-CoV-2 serosurveillance research studies with diverse recruitment strategies.

**Methods.** This secondary analysis used demographic data collected between April 2020-November 2023 from SARS-CoV-2 serology studies. The studies included three pre-existing longitudinal cohorts, two convenience samples using blood donations and outpatient laboratory specimens, and one de novo cross-sectional cohort. Specimen counts by age, sex, urbanicity, race/ethnicity, and neighborhood deprivation quintiles were calculated and compared to 2016 Canadian census population counts. For each demographic strata, a representation ratio was derived as the proportion of study specimens divided by the proportion of the population that belonged to the strata. Subgroups were classified as notably underrepresented if greater than 95% of bootstrap replicates produced a representation ratio < 3/4.

**Results.** Racialized minority subgroups were most underrepresented across sex and race/ethnicity strata in pre-existing longitudinal cohorts (representation ratio 0.1-0.3) and blood donors had low representation of females (0.4) but not males (0.9). Rural females were

underrepresented in all probabilistic study designs (representation ratio 0.3-0.7). Representativeness generally improved as participant age increased and material deprivation quintile decreased. 18-26-year-old males were underrepresented across most sex and urbanicity strata.

**Discussion.** This study identified several demographic subgroups are underrepresented across SARS-CoV-2 study designs. Representativeness differed by recruitment strategy for some subgroups, particularly those related to age and rural residence. Ensuring adequate representation is vital to optimize external validity and to enable interventions that address health inequities. These findings highlight the need to further investigate the barriers impeding participation in health research.

**Conclusion.** Nontraditional recruitment strategies were more representative along several sociodemographic strata than traditional probability-based approaches. The influence of study design and sampling strategy on study representativeness is an important consideration for future serosurveillance studies.

# Résumé

**Contexte.** Le dépistage sérologique était un élément clé de la surveillance du coronavirus du syndrome respiratoire aigu sévère 2 (SRAS-CoV-2). La représentativité des études de la sérosurveillance était probablement influencée par les stratégies de recrutement inhabituelles, qui étaient utilisées quand des règles ont nécessité la distanciation physique, quand il y avait des limites de ressources, et quand des données opportunes étaient nécessaires. Certains sous-groupes démographiques, comme les communautés racialisées, étaient sous-représentées dans les études sérologiques du SRAS-CoV-2 par rapport à la population générale. Par conséquent, c'est important de caractériser la représentativité pour identifier les déficits selon la couverture d'échantillonnage et pour évaluer l'impact des interventions sur les inégalités en matière de santé.

**Objectif.** Évaluer la représentativité sociodémographique des études de la sérosurveillance canadienne du SRAS-CoV-2 qui ont utilisé des diverses stratégies de recrutement.

**Méthodes.** Pour cette analyse secondaire, des données démographiques qui étaient recueillis entre avril 2020 et novembre 2023 à partir des études sérologiques du SRAS-CoV-2 étaient utilisées. Les études étaient : trois cohortes longitudinales préexistantes, deux échantillons de convenance et une cohorte transversale de novo. Les chiffres d'échantillons selon l'âge, le sexe, l'urbanité, la race ou l'origine ethnique et les quintiles de défavorisation du quartier étaient calculés et comparés aux chiffres de la population du recensement canadien de 2016. Pour chaque strate démographique, un ratio de représentation était calculé comme la proportion de spécimens d'étude divisée par la proportion de la population qui appartenait aux strates. Les sous-groupes étaient classés comme sous-représentés en manière notamment si plus de 95 pour cent des répétitions de méthode bootstrap produisaient des rapports de représentation < 3/4.

**Résultats**. Dans les cohortes longitudinales préexistantes, les sous-groupes minoritaires racialisés étaient les plus sous-représentés dans les strates de sexe et de race ou de l'origine ethnique (ratio de représentation 0.1-0.3). Aussi, des donneurs de sang avaient moins de représentation féminine (0.4), mais une représentation masculine adéquate (0.9). Des femmes qui vivaient en milieu rural étaient sous-représentées dans toutes les conceptions d'études probabilistes (ratio de représentation 0.3-0.7). La représentativité a généralement amélioré quand l'âge a augmenté et quand le quintile de privation matérielle a diminué. Les hommes qui avaient 18-26 ans étaient sous-représentés dans la plupart des strates de sexe et de l'urbanité dans chaque étude.

**Discussion.** Cette étude a identifié plusieurs sous-groupes démographiques sous-représentés dans les études du SRAS-CoV-2. La représentativité de certaines dimensions sociodémographiques différait selon la stratégie de recrutement. Notamment, les dimensions liées à l'âge des participants et à la résidence rurale. C'est essentiel d'assurer la représentativité pour l'optimisation de la validité externe et pour identifier les interventions qui considèrent des iniquités en santé. Ces résultats soulignent l'importance d'effectuer la recherche qui examinera les obstacles qui empêchent la participation à la recherche en santé.

**Conclusion.** Les stratégies de recrutement non traditionnelles étaient plus représentatives pour plusieurs strates sociodémographiques que les stratégies traditionnelles basées sur les probabilités. C'est important de considérer comment la conception de l'étude et la stratégie d'échantillonnage influencent la représentativité sociodémographique dans les études futures de sérosurveillance.

# Preface

This thesis assesses the sociodemographic representativeness of six Canadian severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) serological surveillance studies relative to the general Canadian population. An overview of SARS-CoV-2 serological surveillance and study representativeness is presented in Chapter 1. A formal literature review of public health disease surveillance, the SARS-CoV-2 pandemic, and the resulting implications for study representativeness is provided in Chapter 2. Chapter 3 details the study methodology. Results are presented as a manuscript in Chapter 4. Chapter 5 provides a detailed discussion of the study findings. A formal summary and conclusion are given in Chapter 6. References are listed in Chapter 7.

# Contribution of Authors

Matthew J. Knight (MJK) performed the literature review, statistical analysis, drafted the manuscript, and wrote all chapters of the thesis. Primary supervisor Dr. W. Alton Russell (WAR) provided guidance during analysis, contributed to interpretation of results, and critically reviewed each chapter of the thesis. MJK and WAR conceived the study with input from co-supervisor Dr. David L. Buckeridge (DLB). WAR and DLB provided access to study datasets and edited the study protocol. DLB reviewed the thesis and provided funding in the form of a stipend.

For the included manuscript, MJK performed data cleaning, statistical analysis, and wrote the first draft of the manuscript. Yuan Yu (YY) provided access to and aided with analysis of the CCAHS-1 and 2016 census data. Jiacheng Chen (JC) assisted with data cleaning of the Alberta Precision Laboratories data. Sheila O'Brien (SB) provided access to the Canadian Blood Services data. Carmen Charlton (CC) provided access to the Alberta Precision Laboratories data. All co-authors critically reviewed the manuscript and contributed to the interpretation of results.

# Acknowledgements

I would like to extend my deepest gratitude to Dr. Alton Russell for his continuous support and guidance throughout this research project. Thank you for welcoming me into your research group, for sharing your expertise, and for being a wonderful mentor. I would also like to extend my most sincere thanks to Dr. David Buckeridge for his insightful guidance and mentorship, and for providing the very generous stipend which allowed me to complete this thesis project. I will always be grateful for the opportunity to learn from you both and can wholeheartedly say I would not be the researcher I am today without your guidance.

I am thankful to the members of the D3Mod lab, the Surveillance lab, and the McGill Clinical and Health Informatics research group for the warm welcome and for fostering such a dynamic research environment. I would especially like to thank Yuan Yu for her support during data analysis and unwavering perseverance with the analysis of the CCAHS-1 dataset. Many thanks to Jiacheng Chen for his guidance during data cleaning of the APL dataset. I would like to extend my gratitude to the data providers and participants of the six datasets, without whom this project would not have been possible.

Finally, I am immensely grateful to my mom, Daniel, Val, and all of my friends and family for their unwavering support. I will forever be thankful for the immense encouragement and guidance you provided through every step of my academic journey. To those who couldn't be here to witness the completion of this project, I dedicate this work to you.

# List of Tables

# List of Figures

# List of Abbreviations

Ab-C: Action to Beat Coronavirus

APL: Alberta Precision Laboratories

CanPath: Canadian Partnership for Tomorrow's Health

CBS: Canadian Blood Services

CCAHS-1: Canadian COVID-19 Antibody and Health Survey 1

CLSA: Canadian Longitudinal Study on Aging

COVID-19: Coronavirus disease 2019

DNA: Deoxyribonucleic acid

RT-PCR: Reverse transcription-polymerase chain reaction

SARS-CoV-2: Severe acute respiratory syndrome coronavirus-2

# Chapter 1: Introduction

## 1.1 Rationale

Public health surveillance can be broadly described as the continuous collection and analysis of health-related data to inform public health action.[1] Surveillance systems have proven vital during outbreaks of pathogens such as human immunodeficiency virus, Ebola virus, and severe acute respiratory syndrome to evaluate community transmission patterns, assess incidence across demographic and geographic strata, and identify emergent infections to enable interventions that mitigate pathogen spread.[2–4] The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic placed enormous pressure on surveillance systems to provide accurate, timely data to guide resource allocation and decision-making. Indeed, government and research institutions leveraged multiple data streams, including case reports, wastewater samples, hospitalizations, and serology specimens to estimate SARS-CoV-2 spread and monitor population immunity.[5–7]

An important characteristic of effective surveillance systems is that they are representative of the target population along demographic, geographic, and other domains related to the research outcome.[8] Representativeness is closely related to the concept of generalizability, which estimates the degree to which study findings may be extrapolated to populations beyond the drawn sample.[9] While generalizability is not necessary to perform valid causal inference,[10] the ability to extrapolate the findings of a representative sample is extremely useful when the entire target population cannot be reasonably measured.[9] Characterizing generalizability is necessary when sampling does not produce a measured population that is representative of the target population.[9]

Serological surveillance, or the testing of blood specimens for antibodies against a given pathogen, is a useful tool to estimate population immunity.[11] Serological surveillance studies can be broadly classified as either convenience samples, probabilistic surveys, or pre-existing longitudinal cohorts. Convenience samples were heavily relied on during the pandemic due to their low cost and ease of data collection.[7] Convenience sampling of study populations that regularly collect serology specimens, such as blood donors or healthcare patients, also permits repeated sample collection over time. Serial cross-sectional designs using these populations were used in multiple countries for SARS-CoV-2 serosurveillance.[12–17] However, the generalizability of convenience samples may be limited if key demographic groups are ineligible for inclusion or poorly represented in the drawn sample.[7] General probabilistic surveys are less prone to recruitment biases due to the use of randomized, weight-based sampling methods.[18] However, low response rates may threaten their generalizability if respondents systematically differ from non-respondents.[19] These designs are also time- and resource-intensive.[20] Sub-studies of pre-existing longitudinal cohorts benefit from use of pre-existing infrastructure and sampling frame,[21] but may also be time-consuming and have limited generalizability due to attrition or if inclusion criteria for the original cohort are restrictive.[22]

Previous SARS-CoV-2 serological surveillance studies have identified gaps in representativeness along several key variable dimensions. Select ethnic communities are underrepresented amongst serology studies conducted in adults[7,23] and children[24] relative to their distribution in the general population. Evaluation of several pre-existing longitudinal cohorts that were used for coronavirus disease 2019 (COVID-19) sub-studies found participants were more educated and had higher income compared to the general population.[25–27] Given some of these underrepresented population subsets (e.g., racialized communities and individuals with low

socioeconomic status) have also been shown to experience greater SARS-CoV-2 burden,[7,28] it is vital to assess differences in representation across recruitment strategies and identify barriers to recruitment. Additionally, understanding differences in study representativeness is crucial to pool data across studies and to inform future interventions that address health inequities. Multiple statistical standardization techniques are routinely used to adjust for imperfect sampling in large population health studies.[29,30] While these techniques can improve sampling imbalances at the analytical stage, they may not be able to adjust for variable levels that are substantially underrepresented in the study sample.[28] Thus, characterizing the effect of recruitment strategy on representativeness is useful to inform the design of future serosurveillance studies.

## 1.2 Objective

The sociodemographic representativeness of SARS-CoV-2 serosurveillance studies conducted in Canada remains unclear. The primary objective of this study is to assess the sociodemographic representativeness of six SARS-CoV-2 serosurveillance studies with various recruitment strategies to the Canadian population. Representativeness will be assessed by age, sex, urbanicity, race/ethnicity, and neighborhood levels of deprivation.

# Chapter 2: Literature Review

## 2.1 Public health surveillance

### 2.1.1 Design and application

Public health surveillance is integral to preserving the health and well-being of the general population. By monitoring and analyzing health-related data, public health surveillance systems can produce valuable information to help mitigate disease outbreaks, identify high-risk groups, and inform program design at a low cost.[31–33] Although a primary aim of public health surveillance is to monitor disease in a population, surveillance data can also be used to assess the effectiveness and equity of interventions and identify areas for improvement.[32] A key component of public health surveillance is providing actionable data to inform policymaking and improve public health.[8,32]

The structure and function of a public health surveillance system are primarily determined by the outcome of interest.[8,34] For example, surveillance systems that monitor local health outcomes, such as outbreaks of foodborne illness, may only require collection and reporting of basic demographic and laboratory data to local health authorities. Other public health surveillance systems may require municipal, state, and national-level data linkage of multiple data streams to effectively assess disease status in the target population.[8,32] The nature and severity of disease often dictate whether public health officers directly engage with a medical network (active surveillance) or rely upon medical practitioners, public health laboratories, and other health operators to report incident cases (passive surveillance).[11] Surveillance systems frequently use a combination of data streams, such as environmental monitoring, clinical medical records, administrative databases, and other disease reporting systems to monitor health-related outcomes.[34–36] Recently, digital surveillance systems that

monitor social media, web search data, and other virtual data streams have been deployed and appear promising for public health surveillance.[37,38]

The design of efficient public health surveillance systems requires careful consideration of data standardization, quality, timeliness, and reporting infrastructure.[8] Standardization of case definitions and measurement protocols is essential to accurately estimate disease prevalence and permit comparison across reporting sites.[28,35] Efficient data sharing and reporting infrastructures are necessary for timely knowledge mobilization and to prevent workflow redundancies resulting from a lack of communication between surveillance streams.[39] Rapid reporting systems are key to identifying pathogen outbreaks and coordinating resources for containment.[31] Data privacy, technological barriers, and economic resources should also be considered when designing surveillance systems.[8]

Public health surveillance systems have proven vital to the mitigation of disease within Canada and abroad. The World Health Organization Influenza Surveillance Network has helped mitigate and eliminate several influenza A pandemics through effective collaboration, pooling of resources, and rapid notification of emerging variants.[31] One foodborne pathogen surveillance system operating between 1994-2009 in the United States provided an estimated $507 million in savings by reducing medical and workforce losses incurred due to foodborne illness.[40] Within Canada, public health surveillance networks have provided estimates of chronic disease incidence,[35] assessed the burden of opioid- and stimulant-induced harms,[41] and identified risk factors for severe poisoning and injuries in youth and adults.[42]

### 2.1.2 Representativeness of surveillance systems

An important characteristic of effective surveillance systems is that they are representative of the target population. A representative surveillance system accurately models

the relevant demographic, geographic, and clinical attributes of the target population in the population under surveillance.[8] Some assessments of representativeness in the literature are vague and lack specification of the target population, underlying assumptions, and the degree to which a study population is representative.[9,43] Prior studies have argued representativeness should not be prioritized during the design stage since representative sampling does not enhance internal validity or causal inference.[10] However, representative sampling is necessary when attempting to describe the health status of a population within a specified time interval for which risk factors and health outcomes may be inequitably distributed across subgroups.[43] Use of a representative study population is also required when applying statistical inference to a target population that cannot be reasonably measured.[9,44]

There are multiple dimensions along which a study may be representative of a target population. A study may be representative if the distribution of sample characteristics matches the target population. This representative subset may lead a study to be representative by interpretation of effect, whereby one may assume the effect measure interpretation to be similar between the study and target populations.[43] In some cases, a representative subset is not required to ensure estimates are representative in interpretation if statistical adjustment methods can be applied to correct for imbalanced sampling.[30] Representativeness may also be characterized as the similarity in effect estimate when evaluated in the study and hypothetical target population. For a study to be representative in estimate, the distribution of covariates must be effectively modeled in the study population and relevant confounding variables accounted for during analysis.[43] The former can be achieved by the use of a probabilistic sampling strategy, while the latter requires rigorous study design and background knowledge.

### 2.1.3 Methods to assess representativeness

A variety of methods have been utilized to assess study population representativeness. The distribution of measured characteristics in the study population is often compared to a large administrative population dataset that reasonably estimates the true variable distribution in the target population, such as census data.[7,16,26,45] Formal hypothesis testing may be performed to identify statistically significant differences in the variable distributions between the study and target populations.[46] In some cases, study estimates standardized for age, sex, marital status, and other sociodemographic indicators may be compared to the target population to better reflect any discrepancies in representativeness following statistical adjustment.[45] While these are relatively simple approaches, these comparison measures can be a useful heuristic to assess if study results are likely to generalize to the target population and to assess the influence of recruitment strategy on representativeness. Indeed, standardized study estimates were compared to a large population registry to evaluate the effect of recruitment on the representativeness of the LifeLines cohort in the Netherlands,[45] while the representativeness of racialized students in California medical school populations was assessed by dividing the proportion of racialized students by the proportion of racialized individuals in the general population.[47]

Other quantitative, index-based measures to assess representativeness have been developed in the clinical research sector. Rather than evaluate the representativeness of a drawn sample, these methods assess the representation of one[48] or several[49] characteristics amongst the inclusion criteria of a sample of clinical studies. Thus, while these methods may provide useful insights into specific variable dimensions that are underrepresented across studies compared to the general population, they do not assess the representativeness of any given study population.[48,49]

## 2.2 Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic

### 2.2.1 Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) epidemiology

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a single-stranded RNA virus first identified in December 2019 in Wuhan, China. A strong body of evidence suggests SARS-CoV-2 originated as a product of zoonotic transmission between humans and bats.[50] Transmission primarily occurs via the transfer of infected liquid particles that penetrate the mucosal membranes of the eyes, nose, and mouth.[51–53] Cellular infection occurs when a spike protein located on the viral exterior engages with extracellular angiotensin-converting enzyme 2 receptors, facilitating membrane fusion and cellular penetration, followed by viral replication and replicon release.[54]

The clinical manifestation of SARS-CoV-2 infection is typically mild. Infected individuals often present with fever or respiratory symptoms, but may also exhibit symptoms such as sore throat, cough, and fatigue.[55] Severe cases of coronavirus disease 2019 (COVID-19) may require mechanical ventilation support and hospitalization.[56] Individuals who are older, immunocompromised, unvaccinated, or who have other comorbidities are at higher risk of developing severe COVID-19.[57,58] Patients may develop 'Long COVID' where symptoms persist months after initial infection; one study evaluating 273,618 confirmed COVID-19 cases found 37% of individuals experienced 1 or more symptoms 3-6 months after infection.[59]

The prevalence and burden of SARS-CoV-2 varied substantially by geography throughout the pandemic. Early estimates from serological studies (April 14 2020 – May 11 2020) collected during the pre-vaccine era (December 2019-November 2020) suggested population-weighted SARS-CoV-2 prevalence was greater in Wuhan, China (6.92%)[60] compared to other countries (Spain [4.6%],[61] United States [1.01%]).[62] Relative to later pandemic periods,

estimates of SARS-CoV-2 prevalence from studies conducted between April 2020 – November 2020 remained low overall but were highly variable across countries (0.3% [Canada],[63] 0.73% [India],[64] 6.0% [England],[65] and 7.9% [Switzerland]).[66] Regional-level estimates were often higher in large metropolitan areas compared to surrounding rural regions, likely due to greater population density.[67] A nationwide study in Spain identified up to five-fold differences in prevalence (1.2%-14.4%) between regions.[61] This may be attributable to differences in policies regarding school closures, face masks, and social distancing, but also political factors and individual adherence to public health recommendations.[67–69]

Despite the emergence of viral variants, SARS-CoV-2 prevalence remained around 5% or less in Canada during the vaccine era (December 2020 – November 2021).[7] Prevalence of natural infection rose sharply with the emergence of the Omicron variant (December 2021 – March 2023) in Canada due to spike protein mutations that provided enhanced transmissibility and ability to escape immune surveillance.[14,70,71] As of January 7, 2024, the World Health Organization estimated the burden of SARS-CoV-2 has surpassed 774,000,000 cases and 7,000,000 deaths worldwide, although this is likely a conservative estimate.[72]

The burden of SARS-CoV-2 is disproportionately distributed across the general population. Racialized groups have a greater odds of hospitalization with COVID-19[73] and experience higher SARS-CoV-2 prevalence in Canada,[7,15,25] the United States,[74] and abroad[28,65] compared to white individuals. Studies conducted in Germany[75] and Canada[76] estimate the odds of infection to be 1.87 and 1.33 times higher in individuals with low education levels compared to highly educated individuals, respectively. A higher likelihood of infection has also been associated with other proxy measures of low socioeconomic status, such as high neighborhood material deprivation.[7,14] Risk of hospitalization and in-hospital mortality due to COVID-19 is greater for older

individuals, which may be driven by poorer immune function.[77,78] While the risk of severe COVID-19 outcomes is substantially lower for younger age groups, younger populations often experience greater SARS-CoV-2 prevalence compared to older adults.[7,15] Along with other factors, social interaction tendencies and attitudes towards public health restrictions likely contribute to the distribution of SARS-CoV-2 in both demographic groups.[7] Quantifying the distribution of COVID-19 among demographic groups is essential to coordinate resource allocation and reduce inequities.

Non-pharmaceutical interventions played a critical role in mitigating the spread of SARS-CoV-2 during the pandemic. A prospective cohort study of 198,077 individuals in the United States found the risk of predicted COVID-19 was 62% lower for individuals who constantly used face masks and 31% lower for communities with the highest grade of social distancing.[79] Other countries, such as Sweden, adopted very limited non-pharmaceutical interventions with the goal of permitting community infection to build a natural 'herd immunity' in the population.[80] This strategy was largely unsuccessful due to the rapid mutation of the virus and low COVID-19 prevalence in many countries[61,63–66], including Sweden, and resulted in elongated spikes in infection with greater mortality in the Swedish population compared to other Nordic countries.[80] When the strain of SARS-CoV-2 used for vaccine development matches the strain circulating through the general population, vaccination against SARS-CoV-2 has been shown to be up to 96% effective in preventing COVID-19 and is highly effective in preventing COVID-19-related hospitalizations among high-risk groups, such as older adults.[81,82] However, due to antibody waning over time, booster doses are necessary to prevent a resurgence in COVID-19 cases and hospitalizations.[83] Thus, both pharmaceutical and non-pharmaceutical interventions are key to containing the spread of SARS-CoV-2.

2.2.2 Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) surveillance

Many infectious disease surveillance systems operating prior to the onset of the COVID-19 pandemic were ill-equipped to effectively monitor SARS-CoV-2. Clinical case-based reporting systems, used early in the pandemic, grossly underestimated COVID-19 prevalence because the majority of SARS-CoV-2 cases are either asymptomatic or unreported to the medical system.[33,84] Case-based estimates may also be biased by individuals who frequently seek out medical care and by limited testing capacity.[85] In comparison, wastewater surveillance methods that test community sewage samples for SARS-CoV-2 are unbiased by medical-seeking behaviors and may provide more accurate prevalence estimates than case-based methods since they capture both symptomatic and asymptomatic cases.[86] Because SARS-CoV-2 can be detected in wastewater specimens before individuals become symptomatic and report to medical authorities, wastewater monitoring can provide early warning of SARS-CoV-2 community transmission.[87]A study conducted in the Netherlands identified a linear correlation between wastewater levels and COVID-19 prevalence, suggesting wastewater monitoring could provide enhanced warning of spikes in COVID-19 case counts.[88] However, this correlation has been shown to fluctuate across time and geography.[89] Additional limitations to wastewater surveillance include the inability to identify individual characteristics associated with SARS-CoV-2 infection and technical difficulties in deriving prevalence estimates from wastewater data.[6]

Throughout the SARS-CoV-2 pandemic, governments and other public health organizations utilized several types of biological assays to inform SARS-CoV-2 surveillance outcomes.[90] Among diagnostic tests used to identify active SARS-CoV-2 infection, reverse transcription-polymerase chain reaction (RT-PCR) is considered the gold-standard due to its high

sensitivity and specificity. Briefly, RT-PCR converts viral ribonucleic acid to deoxyribonucleic acid (DNA) and repeatedly replicates the DNA in order to detect SARS-CoV-2 within a pre-determined number of test cycles.[90] However, because RT-PCR can identify SARS-CoV-2 in individuals with low viral loads, a positive test does not guarantee an individual is infectious.[91] Additionally, RT-PCR testing is costly, requires 1-2 days to obtain results, and may produce false negatives due to test contamination or erroneous specimen handling and storage.[90] A common alternative to RT-PCR is the use of antigen-based diagnostic tests which recognize SARS-CoV-2 surface proteins. While these tests provide several benefits, including low cost, near-immediate results, and the ability to perform self-testing, they offer reduced test sensitivity compared to RT-PCR.[92–94] Although not formally a diagnostic test, antibody testing of serology specimens also proved effective for SARS-CoV-2 surveillance activities to assess the burden of prior SARS-CoV-2 infection and estimating antibody dynamics.[95] Regardless of the chosen analytical test, lack of standardization in specimen processing, equipment calibration, and test positivity thresholds limits the comparison of results across studies.[96,97]

Serological surveillance was a major component of many countries' SARS-CoV-2 surveillance strategies. In locations where approved vaccines only target the spike protein of the SARS-CoV-2 virus, such as Canada, prior natural infection can be inferred by testing specimens for antibodies against the nucleocapsid protein of the virus, while antibodies against the nucleocapsid, spike, or receptor-binding domain proteins are suggestive of either natural infection or vaccination.[98] Reporting nucleocapsid and spike or receptor-binding assay results in conjunction can be used to infer vaccination levels in the population.[7] Importantly, unlike wastewater surveillance, serosurveillance offers the ability to identify associations between individual characteristics and SARS-CoV-2 test positivity.[33] Indeed, serosurveillance has been

successfully used to identify subgroup differences in SARS-CoV-2 infection,[14,25,99,100] assess vaccine uptake and antibody waning over time,[7,101] and estimate infection prevalence throughout the pandemic.[63] However, because antibody generation generally occurs between 10-15 days after symptom onset, serology testing is a poor indicator of current infection rates.[102]

## 2.3 Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) serosurveillance

### 2.3.1 Study design

Many SARS-CoV-2 serosurveillance studies relied upon convenience sampling to recruit participants or obtain blood specimens for testing.[13,14,16,17,103] Convenience samples using leftover, or residual, blood specimens from blood donors or healthcare patient populations are advantageous for serosurveillance due to low costs, pre-established infrastructure, and ability to repeat sample collection over time.[7,104] Within Canada, residual blood convenience samples have provided key insights into temporal patterns in SARS-CoV-2 infection,[13–15,70] antibody waning,[7] and subgroup differences associated with SARS-CoV-2 test positivity.[13–15] Convenience sampling of serology specimens originally collected for alternative purposes may be necessary in contexts where probabilistic designs are anticipated to fail; a study of 533 individuals receiving routine medical care in Haiti elected to avoid probabilistic sampling designs due to concerns that strong negative attitudes towards COVID-19 would severely reduce participation rates.[105] However, retrospective convenience samples relying on existing specimens often cannot collect additional data pertaining to potentially confounding variables, such as ethnicity, education status, and income level, beyond what is collected by the data provider.[21,106,107]

Probabilistic sampling strategies provide a robust framework for deriving valid statistical inference from serosurveillance studies. The use of random, non-zero weights to select participants from a sampling frame may provide improved coverage and representativeness of

the target population compared to convenience sampling.[108] Survey quality metrics such as sampling errors, non-response rates, and characteristics associated with non-respondents may also be calculated.[19] Probabilistic serosurveys in Spain and Canada have accounted for population density, COVID-19 prevalence, and low-population regions in their sampling strategies while achieving near-national level coverage with adequate precision.[109,110]

Sub-studies of pre-existing longitudinal cohorts also present several advantages for SARS-CoV-2 serosurveillance. Use of pre-existing sampling frames and study infrastructure may reduce recruitment periods[21] and time to publication. Repeating specimen collection at multiple time periods allows for effective modeling of immunity since recruitment biases likely remain relatively constant over time, although attrition may introduce additional biases.[8,111] In comparison to convenience samples, closed serosurveillance cohorts are conducive to longitudinal analysis while providing the additional benefit of working within a probabilistic framework.[112] Longitudinal cohorts likely collect more participant data than residual blood convenience studies and also provide the ability to perform further data collection.[107,113] The use of home-based serology tests, often necessitated by social distancing restrictions, may result in reduced test performance and limited comparability to studies that employed venous specimen collection techniques.[114] However, home-based collection methods may be more logistically feasible and provide enhanced population coverage during pandemic periods compared to designs requiring participants to travel for venous specimen collection.[115] Home-based serology tests are especially useful for serosurveillance studies involving children, as parents are more willing to consent to needle prick collection methods than venous collection methods.[116]

2.3.2 Representativeness to the general population

Several elements of serosurveillance study design may influence study representativeness. Statistical inference from convenience samples is susceptible to selection and non-response biases if the characteristics of the sample are not reflective of the true variable distribution in the target population.[106,117] While weighting techniques are available to adjust for imbalances in representativeness, they cannot adjust for unmeasured covariates and may be ineffective if the calibration dataset contains participation biases.[18] Lack of a sampling frame prevents assessment of metrics to quantify representativeness, including response rates, sampling errors, or characteristics of non-respondents.[118] Low response rates, common to many probabilistic serosurveys,[99,109,119] may also impede generalizability if respondents systematically differ from non-respondents along variable dimensions related to the outcome of interest.[19] The degree to which serosurveillance study design influences representativeness to the general population remains unclear.

The characteristics of study populations available for serosurveillance may also influence study representativeness. The generalizability of blood donor studies may be biased by self-selection to donate, eligibility criteria, and donors' tendency to be in better health than the general population.[7,120] A similar 'healthy participant' effect has been observed in respondents to prospective epidemiologic cohort studies[121] and general population research.[122] Study populations recruited from healthcare settings may be in worse health than the general population or have greater tendencies to seek out medical care.[13] Generalizability may be limited when using pre-existing longitudinal cohorts or if inclusion criteria for the original cohort are restrictive.[22] Characterizing the extent to which study population selection and attrition influences generalizability is key to reducing the influence of recruitment biases in future studies.

Several statistical techniques are available to improve study representativeness. Calibration weighting, commonly known as raking, iteratively weights sample data to reflect the observed distribution of key covariates in the general population.[30,123] Multilevel regression and post-stratification uses regression modeling to predict the outcome of interest for population strata, after which strata-specific predictions are weighted according to the distribution of variables in the general population.[29] While these methods are preferable over crude estimation, they may fail to provide valid estimates of population subgroups that are either included with low frequency such that estimates are unstable or are not included in the study sample.[30] Given statistical modeling can reduce, but not eliminate, biases associated with imbalanced sampling, assessing the influence of recruitment strategy on study representativeness is necessary to improve the design of future serosurveillance studies.

# Chapter 3: Methods

## 3.1 Study overview

This study assessed the sociodemographic representativeness of six Canadian serosurveillance studies as compared to the 2016 Canadian census. The studies included three pre-existing longitudinal cohorts, two convenience samples, and one de novo probabilistic survey conducted between April 2020 – November 2023.[13,14,25–27] Following the framework presented by Rudolph et al., we classified a study as representative if the sociodemographic composition of the study population was aligned to the age- and region-matched 2016 Canadian census.[43] While this may produce an interpretation of effect that is representative of the target population, this definition does not assume the magnitude of effect to be representative within an estimated margin of error and makes no assumption of the underlying sampling mechanism. All studies utilized antibody-based assays to assess the presence of SARS-CoV-2 antibodies in participant serology specimens. This study received ethical approval from the McGill University Faculty of Medicine and Health Sciences Review Board.

## 3.2 Data sources

### 3.2.1 Convenience samples

Blood donor data were provided by the Canadian Blood Services (CBS) coronavirus disease 2019 (COVID-19) Seroprevalence study. CBS collects blood donations from fixed and mobile sites in all Canadian provinces except Québec. To donate blood, donors must be at least 17 years old, have a blood hemoglobin level of 125 g/l (women) or 130 g/l (men), be afebrile, and not be COVID-19 positive or been in contact with a positive case during the two weeks prior to donation.[7,124] A small specimen is routinely retained from each donation for additional

screening for infectious pathogens, known as retention samples, which this study used to conduct a serial cross-sectional study of specimens collected between May 2020 – November 2023.[14,104] Retention samples from all weeks of the month were available for testing during most of the study period, although the study was restricted to specimens collected in the second half of the month during August 2020 – December 2020 and February 2021 – November 2021. A combination of convenience sampling (May 2020 – July 2020, December 2021), random sampling (January 2021 – May 2021, January 2022 – November 2023), and stratified random sampling by age and region (July 2021 – November 2021) was used to select specimens for testing.[7,14,104]

Outpatient data was provided by Alberta Precision Laboratories (APL). Venous residual blood specimens from individuals receiving routine outpatient laboratory testing were collected monthly from provincial laboratories in major metropolitan cities (Calgary, Edmonton) and laboratories in surrounding regions (Lethbridge, Medicine Hat, Grand Prairie, Red Deer) between April 2020 – October 2022.[12,13] Specimens from metropolitan regions were tested one day per month, while specimens from surrounding regions were collected sequentially over three to seven days per month to increase rural representation. All residents of Alberta who sought medical care and provided a venous blood specimens for laboratory testing were eligible for the study.[12,13]

### 3.2.2 Pre-existing longitudinal cohorts

The Action to Beat Coronavirus (Ab-C) study is a longitudinal open cohort recruited from the Angus Reid Forum. The Angus Reid Forum is a marketing panel conceived using a two-stage stratified design that sampled participants from 300 regional units across Canada, stratified by age, gender, and education status, to produce a nationally representative sample.

Census metropolitan area, age, sex, and education status were used as sampling strata to invite Angus Reid Forum members to the Ab-C study.[25] Members who were age 18 or older, spoke French or English, and were willing to provide serology specimens via dried blood spot were eligible for inclusion. Participants received compensation in the form of points from the Angus Reid Forum, which can be used to purchase rewards such as pre-paid credit cards. Individuals aged 60 and older were oversampled to provide sufficient power for stratified analyses. Specimens and survey responses were collected at four time points between May 2020 – April 2022.[25]

Data for the Canadian Partnership for Tomorrow's Health (CanPath) COVID-19 Antibody study was collected from six regional cohorts in Alberta, British Columbia, Manitoba, Ontario, Québec, and the Atlantic provinces (New Brunswick, Newfoundland and Labrador, Nova Scotia, Prince Edward Island).[125–129] Inclusion criteria varied for each original cohort. Recruitment strategies also differed by cohort but included stratified sampling from administrative healthcare datasets,[129] random digit dialing by health region combined with further household sampling,[128] and convenience designs relying on advertising, word of mouth, and community events to recruit participants.[125–127] No additional eligibility criteria were specified for the COVID-19 Antibody sub-study. The sampling frame for the COVID-19 Antibody sub-study was tailored to include individuals with a higher risk of COVID-19, including those residing in long-term care facilities or regions with low socioeconomic status.[130] Survey responses and dried blood spot specimens were collected between February 2021 – November 2021.

Data for the Canadian Longitudinal Study on Aging (CLSA) COVID-19 Antibody study were collected from CLSA participants residing in the 10 Canadian provinces between October

2020 – August 2021.[131] The original CLSA cohort recruited participants aged 45-85 between 2010-2015 into either a tracking or comprehensive sub-cohort and will continue follow-up every three years until 2033. Participants were sampled from provincial health databases, the Canadian Community Health Survey – Healthy Aging, and using household random digit dialing.[26,132] Census data were used to oversample areas with low education and socioeconomic status since these populations are typically poorly represented in population health studies.[26] Individuals who were cognitively and physically independent and who spoke French or English were eligible for inclusion in the original cohort, while individuals who were institutionalized, member of the Canadian armed forces, or who resided in Indigenous reserves, the Canadian territories, or select rural regions were ineligible.[133] No additional eligibility criteria were specified for the COVID-19 Antibody sub-study. A random sample stratified by age or age and province was used to recruit participants from the comprehensive and tracking sub-cohorts, respectively.[131] Participants provided survey responses and a sample for antibody testing. Participants either went to one of eleven collection sites to provide a venous specimen or person or used a kit to provide a dried blood spot if pandemic restrictions prevented travel to a collection site.[131]

### 3.2.3 De novo probabilistic cohort

The Canadian COVID-19 Antibody and Health Survey 1 (CCAHS-1) is a prospectively sampled probabilistic survey. Questionnaire data and dried blood spots were collected between November 2020 – April 2021 from participants aged 1 or older residing in the 10 Canadian provinces and capital cities of the three Canadian territories (Iqaluit, Yellowknife, and Whitehorse). Participants who were institutionalized, member of the Canadian armed forces, resided in Indigenous communities, or resided outside the territorial capitals were ineligible for inclusion.[134] The sampling frame for participants aged 25 and older was constructed using a

comprehensive list of household addresses (Dwelling Universe File), while data from the 2016 Canadian census, Canadian child benefit program, and Canadian Revenue Agency were used to construct the sampling frame for participants aged 1-24. Specimen collection was dispersed across 30 geographic strata to reflect the distribution of COVID-19 in the general population. The sample size of regional strata with large population counts or higher COVID-19 prevalence was increased to produce adequately precise estimates but was adjusted to ensure a sufficient quantity of specimens were collected from sparsely populated regions.[134] Stratified random sampling was used to select participants directly from the sampling frame (ages 1-24) and randomly select one participant from a selected household (ages 25 and older). Sample data were weighted to adjust for probability of selection, survey non-response, and dried blood spot non-response. Sample data were further adjusted using a supplementary administrative dataset to ensure the regional distribution of age and sex were representative of the general population.[134]

## 3.3 Data processing

I used unique identifiers provided by each data provider to calculate participant and specimen counts. Participant identifiers were derived if they were not provided but excluded specimens that could not be mapped to a participant (n = 324 [APL]). I excluded participants who did not meet the inclusion criteria of their respective study or who were missing age, province or territory of residence, and serology test result data. Given only one study collected specimens in the Canadian territories (CCAHS-1),[134] I restricted the primary analysis to the 10 Canadian provinces and assessed territorial specimen representativeness in a separate analysis. I did not calculate representativeness of the 0-17 age group in the CBS study because the minimum donor age was 17.

The sociodemographic composition of the general population was estimated using the long-form 2016 Canadian census.[135] The long-form census is a mandatory survey distributed to 25% of the Canadian population that collects data on a variety of topics including participant health, ethnicity, employment, and housing. Data are weighted to account for survey design, coverage, non-response, and sampling errors to ensure representativeness to the general population.[135] Weighted 2016 Canadian census counts were rounded to the nearest multiple of 0 or 5 and weighted CCAHS-1 specimen counts were rounded to base 2000 in accordance with Statistics Canada's data privacy regulations.

I extracted participant age, sex, postal code, date of sample collection, and self-reported race/ethnicity from each dataset and the 2016 census. I calculated participant age as either the age at specimen collection (Ab-C, APL, CBS) or questionnaire completion (CanPath, CLSA, CCAHS-1, 2016 census) and categorized age as 0-17 years, 18-26 years, 27-36 years, 37-46 years, 47-56 years, or 57 and older. For Ab-C specimens collected between December 2020 – April 2021 and July 2021 – September 2021, I used the 2019 baseline age since the age at collection was not measured. I categorized sex as male or female and removed alternative responses (n = 138 [Ab-C]) for all analyses involving participant sex. I used the first three characters of the postal code to classify participant's residence as urban or rural. A mapping table provided by CBS was used to assign full postal codes where available (CBS, APL, CCAHS-1, CLSA, 2016 census) to a quintile of the Pampalon material and social deprivation indices.[136,137] Material deprivation is a composite measure of employment, education, and income that estimates an individual's access to materials or services that provide a good quality of life. Social deprivation is a composite measure of family structure, marital status, and other social structures that estimates the strength of an individual's social network.[136] I calculated date of specimen

collection as either the precise date a specimen was provided by the participant (CBS, APL), the date of questionnaire completion (CanPath, CLSA, CCAHS-1) or the date the specimen was received by the data provider (Ab-C).

To account for the differential measurement of race between study datasets, I dichotomized self-reported race/ethnicity groups into 'white' or 'racialized minority' according to the population group variable from the 2016 Canadian census (Tables S1-S2).[138] In the primary analysis, I classified individuals who self-identified as both white and a racialized minority, along with Indigenous-identifying individuals, as racialized minorities. I excluded individuals who self-identified as Indigenous from the 2016 census dataset when assessing representativeness for studies where data on Indigenous identity was unavailable (CLSA, CanPath). For studies that performed multiple rounds of specimen collection (Ab-C, CBS, CLSA), I imputed missing variables when available for another collection. I assessed representativeness using a complete cases approach for each set of demographic strata (e.g., records missing race/ethnicity observations were excluded when stratified by age, sex, and race/ethnicity, but not when stratified by age, sex, and urban or rural residence).[139]

## 3.4 Statistical analysis

I summarized the variable distributions of each study dataset (Table S3). To assess the representativeness of each study relative to the general population along one or more sociodemographic dimensions, I calculated a representation ratio as the proportion of study specimens belonging to sociodemographic strata divided by the proportion of weighted 2016 census counts belonging to the strata. Because the CCAHS-1 study was designed to be representative only after statistical calibration with a supplemental administrative dataset, I used weighted CCAHS-1 counts in the numerator of the representation ratio. Unweighted counts were

used to assess representativeness for all other datasets. I restricted census counts by age and region to match studies' composition (Table 1).

I used bootstrapping to identify notably underrepresented sociodemographic subgroups. Bootstrapping is a resampling technique that performs sampling with replacement to generate N-simulated samples. This provides a distribution of the test statistic and permits the estimation of standard errors and uncertainty intervals.[140] I produced 5000 bootstrap resamples to generate a distribution of representation ratios for sociodemographic strata. I classified a subgroup as notably underrepresented if greater than 95% of bootstrap resamples produced representation ratios less than 3/4.

As mentioned in Chapter 2.3.2, multiple statistical adjustment methods can be applied to correct imbalanced sampling if an adequate number of samples are collected for underrepresented strata.[29,123] For each dataset, I assessed the feasibility of statistical adjustment by imposing four increasing levels of stratification and calculating the number of cells with counts greater than 25. Stratifying variables included age group, sex, urban or rural residence, self-identified race/ethnicity, and sample collection date categorized into 2-month intervals.

To assess the influence of the race/ethnicity classification on study representativeness, I conducted a sensitivity analysis that classified individuals who self-identified as both white and a racialized minority as white. I also performed a second sensitivity analysis which included Indigenous-identifying individuals in the census dataset when calculating representativeness for studies that did not provide data on Indigenous identity (CLSA, CanPath). All analyses were completed using R version 4.3.1.[141]

# Chapter 4: Results

## 4.1 Preface

Serological surveillance studies were vital during the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic to estimate coronavirus disease 2019 (COVID-19) prevalence and monitor population immunity. Resource limitations, public health interventions, and need for timely data required studies to use diverse recruitment strategies to obtain serology specimens, which likely influenced their representativeness to the general Canadian population. While various methods to assess sample representativeness have been proposed in the literature, few have directly compared the sociodemographic representativeness of studies with diverse recruitment strategies. This study adds to the current literature by assessing the sociodemographic representativeness of six serosurveillance studies with diverse recruitment strategies to the general Canadian population. The present manuscript has been submitted to the American Journal of Public Health and submitted as a conference abstract to the Association for the Advancement of Blood and Biotherapies.

4.2 Manuscript

# SOCIODEMOGRAPHIC CHARACTERISTICS OF COVID-19 SEROSURVEILLANCE STUDIES WITH DIVERSE RECRUITMENT STRATEGIES

Matthew J. Knight[1], Yuan Yu[1], Jiacheng Chen[1], Sheila F. O'Brien[2,3], David L. Buckeridge[1,4], Carmen Charlton[5,6], W. Alton Russell[1,4]

[1]School of Population and Global Health, McGill University, Montreal, Canada

[2]Canadian Blood Services, Ottawa, Canada

[3]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada

[4]COVID-19 Immunity Task Force, Montreal, Canada

[5]Canadian Blood Services, Edmonton, Canada

[6]Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, Canada

**Corresponding author:** Matthew J. Knight matthew.knight@mail.mcgill.ca, +1-613-854-9571

Suite 1200, 2001 McGill College Avenue, Montreal, QC, Canada H3A 1G1

# Abstract

**Objective.** To assess the demographic representativeness of six SARS-CoV-2 serological surveillance research studies in Canada.

**Methods.** We analyzed three pre-existing longitudinal cohorts, two convenience samples using residual blood, and one de novo probabilistic survey conducted between April 2020 – November 2023. We calculated study specimen counts by age, sex, urbanicity, race/ethnicity, and neighborhood deprivation quintiles. For each demographic strata, we derived a representation ratio by dividing the proportion of study specimens by the proportion of population in the strata.

**Results**. The six studies included 1,321,675 specimens. When stratifying by age group and sex, 65% of racialized minority subgroups were moderately underrepresented (representation ratio < 0.75). Representation was generally higher for older Canadians, urban neighborhoods, and neighborhoods with low material deprivation. Rural representation was highest in a study that used outpatient laboratory blood specimens. Racialized minority representation was highest in a de novo probabilistic survey cohort.

**Conclusions.** While no study had adequate representation of all subgroups, less traditional recruitment strategies excelled in some dimensions of representativeness. Understanding demographic representativeness and barriers to recruitment are important considerations when designing population health surveillance studies.

# Introduction

In April 2020, the largest serological surveillance program in Canada's history was established to monitor population immunity to severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), informing COVID-19 epidemiology and antibody dynamics. Between April 2020 and February 2021, many studies began testing blood specimens for SARS-CoV-2 antibodies.[1–6] Challenged by social distancing measures to curb the coronavirus disease 2019 (COVID-19) pandemic, studies used diverse strategies to recruit participants or obtain residual blood samples for testing. Recruitment directly impacts the extent to which study participants represent the general population.[7] Comparing the demographic representativeness of concurrent SARS-CoV-2 serosurveillance studies can inform surveillance strategies for diverse pathogens in Canada and abroad.

SARS-CoV-2 surveillance studies can be broadly categorized as convenience samples, de novo probabilistic surveys, or pre-existing longitudinal cohorts. In many countries, convenience samples of residual blood specimens were used due to low operational costs and capability to repeat sample collection over time.[1,2,8,9] These study designs may suffer from selection bias if demographic subgroups are poorly represented or ineligible to be sampled.[10] General probabilistic serosurveys can mitigate recruitment biases by using stratified, weight-based approaches to recruitment. De novo designs allow tailoring recruitment to study objectives. Probabilistic serosurveys were successfully deployed to monitor SARS-CoV-2 in multiple countries.[4,5,11-13] However, these designs are time-intensive, require many resources, and have important limitations, such as low response rates.[6,12,14] Use of pre-existing longitudinal cohorts may be more efficient due to established study infrastructure and sampling frame, but

generalizability may be limited by attrition or if inclusion criteria for the original cohort are restrictive.

In this study, we assessed the demographic representativeness of six SARS-CoV-2 serosurveillance studies by comparing the composition of each study population to the 2016 Canadian census according to age, sex, race/ethnicity, urbanicity, and neighborhood measures of socioeconomic deprivation.

# Methods

## Data

We assessed representativeness by analyzing demographic data from six Canadian study populations (Table 1). Here, we define a study to be representative if the sociodemographic composition of the study population matches that of the census-based target population; we make no assumptions of the underlying sampling mechanism or whether inferences from the study population are representative. This similarity suggests the interpretation of an effect measure may be generalizable to the target population, but does not assume the quantitative effect estimate, within a given uncertainty interval, will be identical between the study and target populations.[15] Studies' recruitment strategies included one fully de novo cross-sectional probabilistic sample (the Canadian COVID-19 Antibody and Health Survey 1 [CCAHS-1]), one open longitudinal cohort recruited from a marketing research panel (Action to Beat Coronavirus [Ab-C]), two pre-existing closed longitudinal cohorts (the Canadian Longitudinal Study on Aging COVID-19 Antibody Study [CLSA], the Canadian Partnership for Tomorrow's Health COVID-19 Antibody Study [CanPath]), and two serial cross-sectional convenience samples that

used residual blood from blood donations (Canadian Blood Services [CBS]) and from specimens collected for outpatient laboratory testing (Alberta Precision Laboratories [APL]). The included studies tested specimens collected from April 2020 to November 2023 with sample sizes ranging from 11,050 (CCAHS-1) to 1,039,298 (CBS). Inclusion criteria and enrollment procedures for each study have been described previously.[1,2,6,13,16,17]

From each dataset, we extracted participants' age, sex, postal code, date of specimen collection, and self-reported race/ethnicity. The first three digits of postal code were used to classify participants' residence as urban or rural, and the full postal code was used to assign participants' neighborhood to a quintile of the Pampalon material and social deprivation indices. Material deprivation is a composite measure of education, employment, and income reflecting access to essential material resources. Social deprivation is a composite measure of people living alone, single-parent families, and people who are either separated, divorced, and/or widowed, reflecting the fragility of social networks. Both index measures are derived from the 2016 Canadian census.[18] Date of specimen collection was provided as the actual collection date (CBS, APL), date of questionnaire completion (CanPath, CCAHS-1, CLSA) or date of specimen receipt (Ab-C).

Race/ethnicity information was unavailable for the APL study. Deprivation indices were available for the CBS, APL, CCAHS-1, and CLSA studies. Specimen counts for the CCAHS-1 study were rounded to base 2000 in accordance with data usage guidelines. Age was calculated as the age at specimen collection (Ab-C, APL, CBS) or questionnaire completion (CanPath, CLSA, CCAHS-1) and categorized as 0-17 years, 18-26 years, 27-36 years, 37-46 years, 47-56

years, or 57 years and older. For Ab-C specimens collected between December 2020 – April 2021 and July 2021 – September 2021, we used the 2019 baseline age since the age at current collection could not be calculated. We categorized sex as male or female and excluded participants who provided alternative responses (n = 138 [Ab-C]) from all analyses involving participant sex. Because race/ethnicity data collection varied between studies and differed from census categorization, we re-classified participants as 'white' and 'racialized minority' and did not analyze specific racialized minority groups (Tables S1-S2). For studies allowing multiple encounters with participants, we imputed missing variables when available for another encounter (CBS, Ab-C, CLSA). We classified participants who identified as both white and a racialized minority as a racialized minority, and we considered Indigenous identities as a racialized minority but conducted sensitivity analyses with different classifications. Because only CCAHS-1 collected specimens from the capital cities of the Canadian territories, we restricted our primary analysis to specimens collected from Canadian provinces but assessed territorial representativeness in a separate analysis. We excluded participants who did not meet the inclusion criteria of their respective study and who were missing age, province/territory of residence, or serology test result data. For the CBS study, we did not assess the representativeness of the 0–17-year-old age group because there were no donors younger than 17. We calculated specimen counts using complete cases within each set of demographic strata (e.g., participants missing race/ethnicity were excluded when stratifying by age, sex, and race/ethnicity but not when stratifying by age, sex, and urbanicity).

## Representation ratio analysis

To assess the representativeness of subgroups defined by one or more sociodemographic variable, we derived a representation ratio by dividing the proportion of specimens in a sociodemographic subgroup by the proportion of general population in the subgroup using weighted 2016 Canadian census counts.[19] For each study, census counts were restricted by age and province/territory to match studies' composition (Table 1) and rounded to the nearest multiple of zero or five. The representation ratio numerator used unweighted counts for all studies except CCAHS-1, which was designed to be representative only after statistical calibration with a supplemental administrative dataset.[6] Census counts by race/ethnicity were derived using the population group variable.[19] We did not classify Indigenous-identifying individuals as racialized minorities for the CLSA and CanPath studies since data on Indigenous status was not available. We performed bootstrapping (n = 5000) to identify notably underrepresented demographic subgroups and considered a subgroup to be notably underrepresented if representation ratios were below 3/4 in over 95% of bootstrap resamples.

## Sample count by strata analysis

In some cases, statistical adjustment or subsampling may allow derivation of representative population statistics from large but unbalanced study populations if there are sufficient samples from less represented strata.[20] To inform whether this would be feasible in our study populations, we assessed the number of strata with counts greater than 25 when grouped by age, sex, urbanicity, race/ethnicity, and date of specimen collection binned into two-month intervals. All analyses were conducted in R version 4.3.1.[21] Analytical code will be available in a public

repository upon publication. This study was approved by the McGill Faculty of Medicine and Health Sciences Institutional Review Board.

# Results

## Study population

During data pre-processing, we excluded 3,718 observations for CBS (0.4%), 3,871 for APL (1.8%), 2,052 for Ab-C (7.6%), 2,024 for CLSA (10.5%), and 4,258 for CanPath (16.4%) due to missing data or failure to meet study inclusion criteria. We analyzed the remaining 1,035,580 (CBS), 210,905 (APL), 25,110 (Ab-C), 21,720 (CanPath), 17,310 (CLSA), and 11,050 (CCAHS-1) observations. For the Ab-C, CLSA, and CanPath studies, the minimum age of participants included in our analysis (Table 1) was older than the minimum age specified in their inclusion criteria.[13,16,17] Across studies, the largest number of observations were in the 57 and older age group (34.4% [CCAHS-1] – 91.4% [CLSA]) (Table S3). Observations for the 18-26-year-old age group were generally low (0.0% [CanPath] – 5.6% [APL]) with the exception of the CBS (10.6%) and CCAHS-1 (11.8%) studies. Among studies for which neighborhood deprivation was available, specimen counts across social deprivation quintile were balanced, but only 8.2% (CBS), 8.4% (APL), 9.6% (CLSA), and 13.1% (CCAHS-1) of specimens were provided from the most materially deprived quintile of neighborhoods. Across studies, most observations were for participants who self-identified as white (78.2% [Ab-C] – 94.7% [CLSA]) and female (52.3% [CLSA] – 65.6% [CanPath]), except that 58.2% of CBS observations were from males. Rural specimens accounted for 8.6% (CanPath) – 17.6% (CCAHS-1) of all specimens across studies. Convenience samples collected substantially more specimens for each demographic strata

compared to other recruitment strategies, but specimen counts per strata were greater in the CBS study compared to APL (Figures S1-S4).

## Representation ratio analysis

Studies generally had sufficient representation across sexes (representation ratio 0.7-1.3) and, when available, by social deprivation (Figure 1, Figure S5). Racialized minority subgroups were underrepresented (representation ratio < 1) in multiple age and sex strata in all studies (Figure 2). Racialized minority representation, while still low, was often better in older age groups (Ab-C, CanPath, and CLSA). In contrast, racialized minority representation was better for younger age groups among women for CBS. While APL was sufficiently representative (representation ratio 0.8-1.3) across quintiles of material deprivation and rural regions, CBS observations skewed towards less materially deprived neighborhoods and urban regions, although rural representation for CBS was better than the three longitudinal cohort studies. Urban regions produced larger representation ratios by age and sex strata than rural regions in all studies (Figures 1-2). 18-26-year-old males were underrepresented across most sex and urbanicity strata in all studies for which they were eligible to be sampled.

Among 18–46-year-olds, specimens collected from the Ab-C open cohort produced greater representation ratios across sex and urbanicity strata compared to CanPath (Figure 2). Representation ratios of 18-46-year-old rural residents were generally larger across age and sex strata in the CCAHS-1 study than several studies with probabilistic recruitment strategies (Ab-C, CanPath). Racialized minorities aged 47 years and older were sufficiently represented (representation ratio 0.8-1.3) in the Ab-C open cohort but were underrepresented in the CanPath

and CLSA closed cohorts, except for males aged 57 and older in the CanPath study. Of the two convenience samples, the CBS study was more representative of participants aged 18-46 across sex and urbanicity strata, whereas the APL study was more representative of individuals aged 47 and older. In CCAHS-1, the only study that sampled in the three Canadian territories, 0-17-year-olds were underrepresented across sexes in territorial specimens. (Figure S6). A sensitivity analysis reclassifying mixed race/ethnicity study participants as white had little impact on findings, except all racialized minority subgroups in the Ab-C open cohort that were generally well represented (Figure 2) became underrepresented because racialized minority counts decreased by 55% (Figures S7-S8). A sensitivity analysis in which the denominator for CLSA and CanPath representation ratios included Indigenous-identifying individuals as racialized minorities had little impact on findings (Figure S9).

### Sample count analysis

The convenience samples with large overall sample size produced substantially more cells with counts greater than 25 across 4 levels of stratification compared to all other study designs in the primary analysis (Table 2) and in sensitivity analysis (Table S4). Among studies with probabilistic recruitment strategies, pre-existing closed cohorts (CLSA, CanPath) produced a greater proportion of cells with counts greater than 25 than other probabilistic recruitment strategies (Ab-C, CCAHS-1) for all strata.

## Discussion

In this study, we observed considerable variability in the sociodemographic representativeness of six serosurveillance studies with diverse recruitment strategies. No study was adequately

representative of all sociodemographic subgroups. When planning future surveillance studies, the strengths and weaknesses of each recruitment strategy must be considered in light of the research question and pathogen of interest.

Probabilistically sampled surveys have traditionally been considered the 'gold standard' for obtaining representative samples.[7] Use of administrative datasets to construct sampling frames for large population survey studies often provides superior population coverage compared to non-probability samples that rely on participant self-selection. The underlying statistical framework also permits estimation of sampling errors and characteristics associated with non-response.[22] While resource constraints may limit the ability of probabilistic designs to perform repeated specimen collection, non-probability sampling from sources with a continuous stream of residual blood specimens, such as blood donors, may be conducive to longitudinal trend modelling which can also incorporate complex geographic structures.[23] The generalizability of probabilistic designs may also be limited if differences between respondents and non-respondents are non-random.[7] Bias may be introduced via a 'healthy volunteer' effect whereby cohort participants are healthier than the general population,[24] similar to the 'healthy donor' bias documented in blood donor research cohorts.[25] Where available, response rates of the included studies were fairly low (23% [CCAHS-1],[6] 25% [Ab-C];[13] these response rates exclude individuals who completed a questionnaire but did not provide a blood sample). This suggests non-response bias could partially explain some of the observed differences in representativeness between study designs. The above response rates are consistent with other probabilistic serosurveys,[12] although response rates as high as 69% have been reported.[11]

Many large-scale SARS-CoV-2 serosurveillance studies in the United States and other countries relied on blood donor and healthcare patient study populations for serology specimens.[22,26] Blood donors have traditionally been dismissed as a population for public health surveillance, while the potential for expanded screening of residual outpatient laboratory samples remains unclear. Our findings suggest they are representative along multiple sociodemographic dimensions relative to the general population. Future studies should evaluate the potential of linking surveillance cohorts to administrative datasets to improve characterization of representativeness and derivation of statistical weights for adjustment. Gaps in demographic representation may be overcome by using multipronged surveillance approaches that synthesize data from multiple sources.[27] Yet differences in choice of assay, use of venous blood draws or dried blood samples, and the format or availability of variables can curtail the ability to synthesize data across studies.[28,29]

Racialized minorities were underrepresented (representation ratio < 1) across all studies included in our analysis. Language barriers and skepticism of research or medical institutions may contribute to poor representation of these populations.[30,31] While use of stratified random sampling or sampling weights may improve sample representativeness to the target population, they fail to address the underlying individual and societal factors governing participation in health research. Direct engagement and collaboration with community members throughout the research cycle may help mitigate these recruitment barriers by facilitating trust, reducing misinformation, and ensuring study materials are accessible.[31,32] Racialized minorities may be better represented in healthcare cohorts like APL,[26] though a lack of race-based data in Canadian administrative healthcare datasets may make this difficult to measure.[33] Notably, representation

of racialized minorities improved as age increased in most studies requiring participant opt-in (Ab-C, CLSA, CanPath), but young minorities exhibited better representativeness compared to older subgroups in the CBS study. Lack of a standardized definition of participant race/ethnicity impeded comparison across studies and prevented assessment of representativeness by specific minority group.

Several other dimensions of representativeness varied across studies. The Ab-C open cohort was substantially more representative of participants aged 18-46 years old across sex and urbanicity strata compared to the CanPath longitudinal closed cohort (Figure 2). Given both the CLSA and CanPath studies began recruitment of participants aged 45-85 in 2010 and 35-74 in 2009, respectively, the age distribution of both COVID-19 sub-studies unsurprisingly skewed towards older adults.[16,17] Between convenience samples, individuals who resided in highly materially deprived areas were notably underrepresented when using blood donations (representation ratio 0.4–0.6), but not when using outpatient labs (representation ratio 0.9-1.2). Donor eligibility criteria, along with the 'healthy donor effect' or other unmeasured socioeconomic factors, may homogenize the demographic composition of the sampled donor pool.[25] Rural regions had consistently worse representation compared to their urban counterparts in all studies, which may be related to urban-centric study recruitment patterns or willingness to travel for specimen collection (Figures 1-2).

The study had several limitations. First, our analysis only considered representativeness by age, sex, race/ethnicity, urbanicity, and neighborhood deprivation. Many other sociodemographic dimensions are important considerations regarding representativeness in serosurveillance studies,

particularly those related to health and disability. Indeed, we hypothesize a 'healthy participant' sampling bias may have led to underrepresentation of individuals with poor health and/or disability in all study populations except outpatient labs.[24,25] Prior analyses of the pre-existing longitudinal cohorts included in our study have also indicated participants are more educated and/or have higher income than the general population,[13,16,17] as are blood donors in the United States.[34] Second, the measurement of race/ethnicity was inconsistent between studies. Race/ethnicity options for CBS blood donors include four mutually exclusive categories, while the Ab-C, CCAHS-1, CanPath, CLSA, and census datasets permitted selection of multiple racial/ethnic identities. This necessitated dichotomizing the race/ethnicity variable as white or racialized minority and may have biased the CBS representation estimate if individuals who identified as mixed race/ethnicity selected their race/ethnicity as white during donation. Additionally, due to missing Indigenous identity data, we modified our representation assessment for the CLSA and CanPath studies by omitting Indigenous-identifying individuals from the census dataset. Our sensitivity analysis suggests this did not substantially impact our findings (Figure S9). Third, our criteria for defining notably underrepresented subgroups and focus on strata with fewer than 25 samples were largely arbitrary. We plotted the representation ratio bootstrap distributions to assess the influence of our artificial threshold for underrepresentation (Figures S10-S16). Fourth, we did not analyze factors shaping the sociodemographic composition of each study, including intentional oversampling. For example, the CCAHS-1 study used a complex stratified random sampling strategy that oversampled geographic regions with greater COVID-19 prevalence and less populated regions to improve estimate precision. Less populated areas of Canada often have fewer racialized minorities, which likely contributes to lower representation ratios.[6] Understanding the causes and consequences of

each study's sociodemographic composition requires more detailed analysis than is presented here. Finally, our study is not a comprehensive assessment of all SARS-CoV-2 serology studies conducted in Canada. Demographic groups excluded here have been evaluated elsewhere.[35]

## Public health implications

Understanding variability in demographic representation between study designs is an important consideration when planning serosurveillance studies, which increasingly leverage pre-existing samples or study cohorts. We found that underrepresentation of racialized minorities and younger age groups was common and not restricted to convenience samples, which had better representation for some sociodemographic strata. Identifying coverage barriers is vital to support adequate representation and detection of disease trends within demographic subgroups. We also observed differences in the measurement of participant race/ethnicity between studies. This highlights the need to adopt a standardized approach to the measurement of self-identified race/ethnicity.

# Declarations

**Funding:** This work was supported by funding from the Government of Canada through the COVID-19 Immunity Task Force and from Canadian Blood Services (CBS). Ce projet a été soutenu par un financement du Gouvernement du Canada, par le biais du Secrétariat du groupe de travail sur l'immunité COVID-19 et par le Société Canadienne du Sang.

**Acknowledgements:** This research was made possible using data collected by Canadian Blood Services' SARS-CoV-2 Seroprevalence Study; the Canadian Partnership for Tomorrow's Health (CanPath – formerly CPTP) COVID-19 Antibody study and its regional cohorts the BC Generations Project, Alberta's Tomorrow Project, the Ontario Health Study, CARTaGENE, Manitoba Tomorrow Project and the Atlantic Partnership for Tomorrow's Health; the Action to Beat Coronavirus in Canada (Ab-C) study; the Alberta Precision Laboratories study; the Canadian COVID-19 Antibody and Health Survey (CCAHS) conducted by Statistics Canada, and the Canadian Longitudinal Study on Aging (CLSA). We thank the participants, staff, and researchers of all six studies. We also thank Scott McLeish for the insightful discussions which greatly improved the quality of the manuscript. Funding for the CLSA is provided by the Government of Canada through the Canadian Institutes of Health Research (CIHR) under grant reference: LSA 94473 and the Canada Foundation for Innovation, as well as the provinces, Newfoundland, Nova Scotia, Quebec, Ontario, Manitoba, Alberta, and British Columbia. Funding for support of the CLSA COVID-19 questionnaire-based study is provided by the Juravinski Research Institute, Faculty of Health Sciences, McMaster University, the Provost Fund from McMaster University, the McMaster Institute for Research on Aging, the Public Health Agency of Canada/CIHR grant reference CMO 174125 and the government of Nova Scotia. This research has been conducted using the CLSA Baseline Comprehensive Dataset v7.0,

Baseline Tracking Dataset v4.0, Follow-up 1 Comprehensive Dataset v4.0, Follow-up 1 Tracking Dataset v3.0, Follow-up 2 Comprehensive Dataset v1.0, Follow-up 2 Tracking Dataset v1.0, COVID-19 Questionnaire Study Dataset v1.1, and COVID-19 Seroprevalence Study Dataset v1.0 under Application Number 2209005. The CLSA is led by Drs. Parminder Raina, Christina Wolfson, and Susan Kirkland. We appreciate the funding that made each study possible, including the support each study received from the Government of Canada through the COVID-19 Immunity Task Force. The opinions expressed in this manuscript are the author's own and do not reflect the views of the data providers and/or their institutions.

**Conflicts:** The authors have no conflicts to declare.

**Ethics/Consent:** All studies analyzed were approved by a Research Ethics Board or Institutional Review Board of a Canadian institution as reported previously. This secondary analysis of six studies was approved by the McGill University Research Ethics Board (study number 22-03-077).

**Data and materials:** The authors are not authorized to share individual-level data from any study. Processes are available for researchers to request access to datasets for studies that have undergone institutional ethical approval. Data from Canadian Blood Services and Alberta Precision Laboratories may be made available upon request, subject to internal review, privacy legislation, data sharing agreements, and research ethics approval. The CCAHS-1 study by Statistics Canada can be analyzed for approved projects at Research Data Centres located across Canada (https://www.statcan.gc.ca/en/microdata/data-centres/access). Data are available from the Canadian Longitudinal Study on Aging (www.clsa-elcv.ca) for researchers who meet the criteria for access to de-identified CLSA data. Access to the Ab-C study data can be requested

through the COVID-19 Immunity Task Force Databank (https://portal.citf.mcgill.ca/). Access to the CanPath data can be requested through the CanPath data portal (https://portal.canpath.ca/).

**Code availability:** Analytical code will be available in a public repository upon publication.

**Authors' contributions:** WAR and MJK designed the study with input from DLB, SFO, and CC. MJK, YY, and JC contributed to data analysis. MJK and WAR drafted the initial manuscript. All authors revised the manuscript and approved the final version for publication.

# References


1. O'Brien SF, Caffrey N, Yi QL, et al. Cross-Canada Variability in Blood Donor SARS-CoV-2 Seroprevalence by Social Determinants of Health. *Microbiology Spectrum*. 2023;11(1):e03356-22. doi:10.1128/spectrum.03356-22

2. Charlton CL, Nguyen LT, Bailey A, et al. Pre-Vaccine Positivity of SARS-CoV-2 Antibodies in Alberta, Canada during the First Two Waves of the COVID-19 Pandemic. *Microbiology Spectrum*. 2021;9(1):e00291-21. doi:10.1128/Spectrum.00291-21

3. Jha, Prabhat. Action to beat Coronavirus Study: Dataset contributed to the CITF Databank. Version DRAFT. Borealis. doi:https://doi.org/doi:10.5683/SP3/LA2IKO

4. Canadian Longitudinal Study on Aging. Data Support Document SARS-CoV-2 Antibodies. April 2024.

5. Canadian Partnership for Tomorrow's Health. COVID-19 Initiatives. CanPath - Canadian Partnership for Tomorrow's Health. https://canpath.ca/covid-19-initiatives/. Accessed April 26, 2024.

6. Statistics Canada. Canadian COVID-19 Antibody and Health Survey (CCAHS). October 2020. https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1287991. Accessed May 20, 2024.

7. Cornesse C, Bosnjak M. Is there an association between survey characteristics and representativeness? A meta-analysis. *Survey Research Methods*. 2018;12(1):1-13. doi:10.18148/srm/2018.v12i1.7205

8. Abu-Raddad LJ, Chemaitelly H, Ayoub HH, et al. Characterizing the Qatar advanced-phase SARS-CoV-2 epidemic. *Sci Rep*. 2021;11(1):6233. doi:10.1038/s41598-021-85428-7

9. Assefa N, Regassa LD, Teklemariam Z, et al. Seroprevalence of anti-SARS-CoV-2 antibodies in women attending antenatal care in eastern Ethiopia: A facility-based surveillance. *BMJ Open*. 2021;11(11):e055834. doi:10.1136/bmjopen-2021-055834

10. Clapham H, Hay J, Routledge I, et al. Seroepidemiologic Study Designs for Determining SARS-COV-2 Transmission and Immunity. *Emerg Infect Dis*. 2020;26(9):1978-1986. doi:10.3201/eid2609.201840

11. Pérez-Gómez B, Pastor-Barriuso R, Fernández-de-Larrea N, et al. SARS-CoV-2 Infection During the First and Second Pandemic Waves in Spain: The ENE–COVID Study. *Am J Public Health*. 2023;113(5):533-544. doi:10.2105/AJPH.2023.307233

12. Ward H, Atchison C, Whitaker M, et al. Design and Implementation of a National Program to Monitor the Prevalence of SARS-CoV-2 IgG Antibodies in England Using Self-Testing: The REACT-2 Study. *Am J Public Health*. 2023;113(11):1201-1209. doi:10.2105/AJPH.2023.307381

13. Tang X, Sharma A, Pasic M, et al. Assessment of SARS-CoV-2 Seropositivity During the First and Second Viral Waves in 2020 and 2021 Among Canadian Adults. *JAMA Network Open*. 2022;5(2):e2146798. doi:10.1001/jamanetworkopen.2021.46798

14. Patel EU, Bloch EM, Tobian AAR. Severe Acute Respiratory Syndrome Coronavirus 2 Serosurveillance in Blood Donor Populations. *The Journal of Infectious Diseases*. 2022;225(1):1-4. doi:10.1093/infdis/jiab517

15. Rudolph JE, Zhong Y, Duggal P, Mehta SH, Lau B. Defining representativeness of study samples in medical and population health research. *BMJ Med*. 2023;2(1):e000399. doi:10.1136/bmjmed-2022-000399

16. Raina P, Wolfson C, Kirkland S, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *International Journal of Epidemiology*. 2019;48(6):1752-1753j. doi:10.1093/ije/dyz173

17. Dummer TJB, Awadalla P, Boileau C, et al. The Canadian Partnership for Tomorrow Project: A pan-Canadian platform for research on chronic disease prevention. *CMAJ*. 2018;190(23):E710-E717. doi:10.1503/cmaj.170292

18. Pampalon R, Hamel D, Gamache P, Philibert MD, Raymond G, Simpson A. An area-based material and social deprivation index for public health in Québec and Canada. *Can J Public Health*. 2012;103(8 Suppl 2):S17-22. doi:10.1007/BF03403824

19. Statistics Canada. Census Profile, 2016 Census. February 2017. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E. Accessed April 26, 2024.

20. Downes M, Carlin JB. Multilevel regression and poststratification as a modeling approach for estimating population quantities in large population health studies: A simulation study. *Biometrical Journal*. 2020;62(2):479-491. doi:10.1002/bimj.201900023

21. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2023. https://www.R-project.org/.

22. Lindan CP, Desai M, Boothroyd D, et al. Design of a population-based longitudinal cohort study of SARS-CoV-2 incidence and prevalence among adults in the San Francisco Bay Area. *Annals of Epidemiology*. 2022;67:81-100. doi:10.1016/j.annepidem.2021.11.001

23. Yu Y, Knight MJ, Gibson D, O'Brien SF, Buckeridge DL, Russell WA. Temporal trends in disparities in COVID-19 seropositivity among Canadian blood donors. *International Journal of Epidemiology*. 2024;53(3):dyae078. doi:10.1093/ije/dyae078

24. Froom P, Melamed S, Kristal-Boneh E, Benbassat J, Ribak J. Healthy Volunteer Effect in Industrial Workers. *Journal of Clinical Epidemiology*. 1999;52(8):731-735. doi:10.1016/S0895-4356(99)00070-0

25. Atsma F, Veldhuizen I, Verbeek A, Kort W de, Vegt F de. Healthy donor effect: Its magnitude in health research among blood donors. *Transfusion*. 2011;51(8):1820-1828. doi:10.1111/j.1537-2995.2010.03055.x

26. Anand S, Montez-Rath M, Han J, et al. Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: A cross-sectional study. *The Lancet*. 2020;396(10259):1335-1344. doi:10.1016/S0140-6736(20)32009-2

27. Tancredi S, Chiolero A. Designing Surveillance at a Population Level. *Am J Public Health*. 2023;113(11):1143-1145. doi:10.2105/AJPH.2023.307425

28. Mulchandani R, Brown B, Brooks T, et al. Use of dried blood spot samples for SARS-CoV-2 antibody detection using the Roche Elecsys ® high throughput immunoassay. *Journal of Clinical Virology*. 2021;136:104739. doi:10.1016/j.jcv.2021.104739

29. Patel EU, Bloch EM, Clarke W, et al. Comparative Performance of Five Commercially Available Serologic Assays To Detect Antibodies to SARS-CoV-2 and Identify Individuals with High Neutralizing Titers. *Journal of Clinical Microbiology*. 2021;59(2):10.1128/jcm.02257-20. doi:10.1128/jcm.02257-20

30. Etti M, Fofie H, Razai M, Crawshaw AF, Hargreaves S, Goldsmith LP. Ethnic minority and migrant underrepresentation in Covid-19 research: Causes and solutions. *eClinicalMedicine*. 2021;36. doi:10.1016/j.eclinm.2021.100903

31. Bonevski B, Randell M, Paul C, et al. Reaching the hard-to-reach: A systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Medical Research Methodology*. 2014;14(1):42. doi:10.1186/1471-2288-14-42

32. Ekezie W, Czyznikowska BM, Rohit S, et al. The views of ethnic minority and vulnerable communities towards participation in COVID-19 vaccine trials. *Journal of Public Health*. 2021;43(2):e258-e260. doi:10.1093/pubmed/fdaa196

33. Khan MM, Kobayashi K, Vang ZM, Lee SM. Are visible minorities "invisible" in Canadian health data and research? A scoping review. *International Journal of Migration, Health and Social Care*. 2017;13(1):126-143. doi:10.1108/IJMHSC-10-2015-0036

34. Patel EU, Bloch EM, Grabowski MK, et al. Sociodemographic and behavioral characteristics associated with blood donation in the United States: A population-based study. *Transfusion*. 2019;59(9):2899-2907. doi:10.1111/trf.15415

35. Atkinson A, Albert A, McClymont E, et al. Canadian SARS-CoV-2 serological survey using antenatal serum samples: A retrospective seroprevalence study. *Canadian Medical Association Open Access Journal*. 2023;11(2):E305-E313. doi:10.9778/cmajo.20220045

# Tables

Table 1: Summary of Canadian serological study designs included in the study. Serological specimens were collected from various regions across Canada between April 2020 - November 2023.

| Study | Design | Age | Region | Specimen type | Study time and size | De novo recruitment |
|---|---|---|---|---|---|---|
| Action to Beat Coronavirus (Ab-C) | Pre-existing longitudinal open research cohort | ≥ 18 | AB, BC, MB, NB, NL, NS, ON, PE, QC, SK, YT[a] | Dried blood spot | 25,110 specimens from 10,621 participants<br>May 2020 - April 2022 | No |
| Alberta Precision Laboratories (APL) | Serial cross-sectional convenience sample | ≥ 0 | AB | Heparinized plasma<br>Plasma<br>Serum | 210,905 specimens from 187,887 participants<br>April 2020 - October 2022 | No |
| Canadian Blood Services (CBS) | Serial cross-sectional random sample | ≥ 18 | AB, BC, MB, NB, NL, NS, ON, PE, SK | Serum | 1,035,580 specimens from 446,187 participants<br>May 2020 - November 2023 | No |
| Canadian Covid-19 Antibody and Health Survey 1 (CCAHS-1) | Prospective cross-sectional cohort with direct (ages 1-24 ) or multi-stage (ages ≥ 25) sampling | ≥ 1 | AB, BC, MB, NB, NL, NT, NS, NU, ON, PE, QC, SK, YT | Dried blood spot | 11,050 specimens from 11,050 participants<br>November 2020 - April 2021 | Yes |
| Canadian Longitudinal Study on Aging (CLSA)[b] | Pre-existing longitudinal closed research cohort | ≥ 51 | AB, BC, MB, NB, NL, NS, ON, PE, QC, SK | Dried blood spot<br>Plasma | 17,310 specimens from 17,310 participants<br>October 2020 - August 2021 | No |
| Canadian Partnership for Tomorrow's Health (CanPath)[c] | Pre-existing longitudinal closed research cohort | ≥ 25 | AB, BC, MB, NB, NL, NS, ON, PE, QC | Dried blood spot | 21,720 specimens from 21,717 participants<br>February 2021 - November 2021 | No |

*Note*. AB = Alberta; BC = British Columbia; MB = Manitoba; NB = New Brunswick; NL = Newfound and Labrador; NT = Northwest Territories; NS = Nova Scotia; NU = Nunavut; ON = Ontario; PE = Prince Edward Island; QC = Quebec; SK = Saskatchewan; YT = Yukon.

[a]Four specimens were collected from Yukon territory and were excluded from all analyses.

[b]Composed of comprehensive sub-cohort and tracking sub-cohort that recruited participants from seven and 10 provinces, respectively.

[c]Composed of 6 distinct regional cohorts.

Table 2: Percentage of study demographic subgroups with greater than 25 collected specimens after stratification. Date of sample collection was binned into 2-month intervals. Subgroups with counts above threshold value might produce more stable estimates when statistically adjusted. Study specimens were collected from various regions across Canada between April 2020 - November 2023. All specimen counts were unweighted.

| Study (specimen count) | Months sampled | Demographic subgroups | | | |
|---|---|---|---|---|---|
| | | Age, Sex, Province, Month | Age, Sex, Province, Urban, Month | Age, Sex, Province, Race/Ethnicity, Month | Age, Sex, Province, Race/Ethnicity, Urban, Month |
| CBS blood donor (1,035,580) | 41 | 92% | 74% | 70% | 52% |
| APL outpatient laboratory (210,905) | 27 | 94% | 84% | NA | NA |
| Ab-C open cohort (25,110) | 18 | 31% | 20% | 20% | 12% |
| CanPath closed cohort (21,720) | 10 | 40% | 31% | 32% | 26% |
| CLSA closed cohort (17,310) | 11 | 50% | 36% | 34% | 27% |
| CCAHS-1 closed cohort (11,050) | 6 | 33% | 20% | 21% | 13% |

# Figures

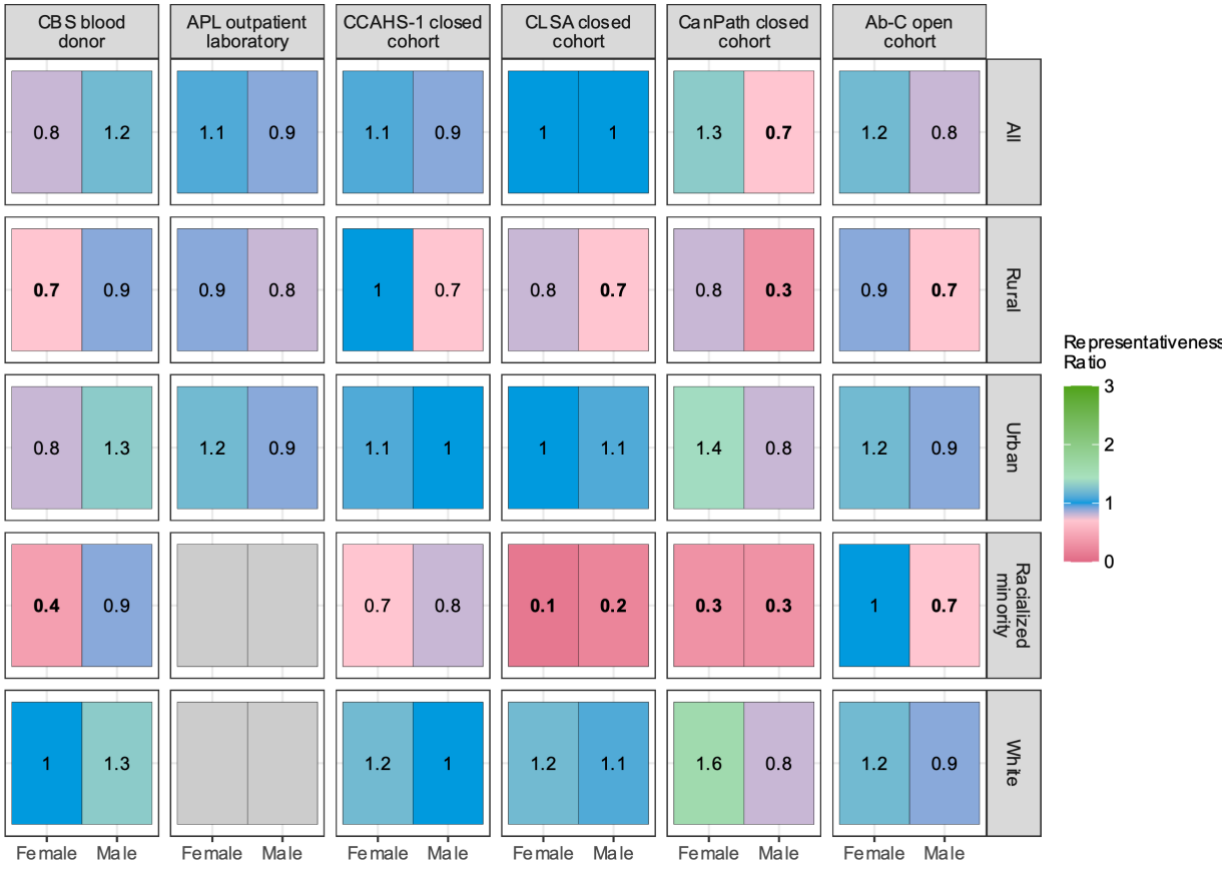


Figure 1: Demographic representativeness of Canadian serological studies compared to the general population by sex, urbanicity, and racial/ethnic identity. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Study specimens were collected from various regions across Canada between April 2020 - November 2023. Bolded representation ratios indicate greater than 95% of subgroup bootstrap replicates produced representation ratios below 0.75. Bootstrapping was not performed for studies with weighted representation ratios (CCAHS-1).

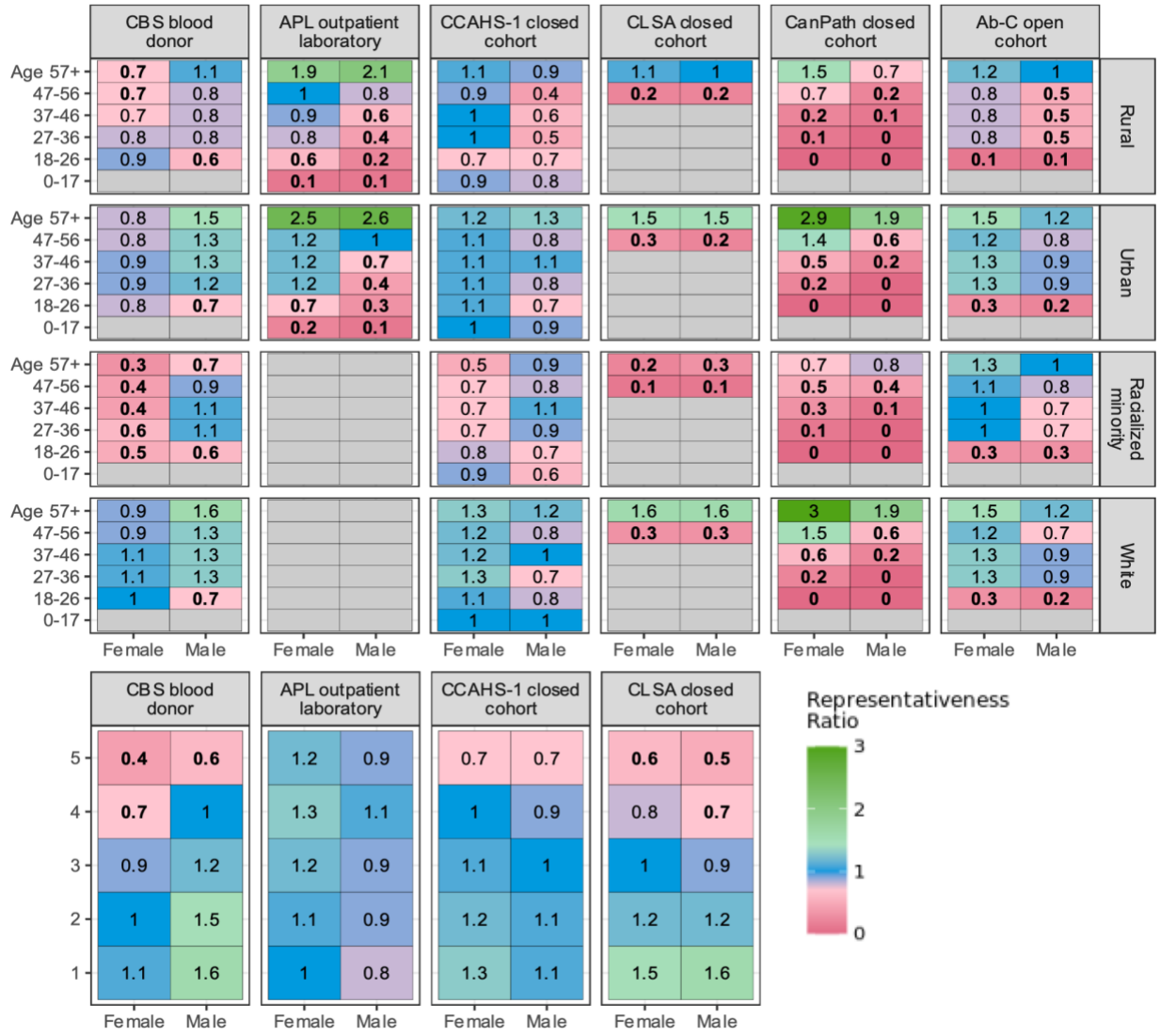


Figure 2: Sociodemographic representativeness of Canadian serological studies compared to the general population by age group, sex, urbanicity, racial/ethnic identity, and material deprivation quintile. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Material deprivation scores were not available for Ab-C and CanPath studies. Study specimens were collected from various regions across Canada between April 2020 - November 2023. Bolded representation ratios indicate greater than 95% of subgroup bootstrap replicates produced representation ratios below 0.75. Bootstrapping was not performed for studies with weighted counts (CCAHS-1).

# Supplemental Materials for 'Sociodemographic characteristics of COVID-19 serosurveillance studies with diverse recruitment strategies'

Matthew J. Knight[1], Yuan Yu[1], Jiacheng Chen[1], Sheila O'Brien[2,3], David L. Buckeridge[1,4], Carmen Charlton[5,6], W. Alton Russell[1,4]

[1]School of Population and Global Health, McGill University, Montreal, Canada

[2]Canadian Blood Services, Ottawa, Canada

[3]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada

[4]COVID-19 Immunity Task Force, Montreal, Canada

[5]Canadian Blood Services, Edmonton, Canada

[6]Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, Canada

# Supplemental tables

Table S1: Classification table of self-reported racial/ethnic identities. Participants who identified as non-white in race/ethnicity, or who identified as Indigenous, were classified as a racialized minority. Ambiguous or missing responses were removed for all analyses which included race/ethnicity as a stratum.

| | | **Response classification** | | |
|---|---|---|---|---|
| **Study** | **Race/ethnicity definition** | **White** | **Racialized minority** | **Removed** |
| CBS | Ethnicity | White | Aboriginal, Asian, Other | Missing |
| CCAHS-1 | Are you X? | White | South Asian, Chinese, Black, Filipino, Arab, Latin American, Southeast Asian, West Asian, Korean, Japanese, Other | NA |
| Ab-C | Ethnicity | English / Irish / Scottish, French, Other European (e.g. German, Russian, Italian, Norwegian, etc.) | Indigenous / First Nations / Inuit / Metis, Filipino, South Asian (Indian, Pakistani, Sri Lankan, etc.), Other Asian (Vietnamese, Korean, Japanese, etc.), Oceania (e.g., Australian, New Zealander, Fijian, etc.), Other (Please specify), Caribbean (e.g. Jamaica, Cuba, Trinidad), Central or South American, African (e.g. South African, Ethiopian, Nigerian, etc.), Middle Eastern / Central Asian (e.g. Lebanese, Iranian, Turkish, Syrian, etc.), Chinese | Rather Not Say, NA |
| CanPath | Race/ethnicity | White | Arab, Black, Chinese, Filipino, Japanese, Korean, Latin American / Hispanic, South Asian, Southeast Asian, West Asian, Other, Other (Specify) | Rather not say |

| | | | | |
|---|---|---|---|---|
| CLSA | Ethnicity | White | South Asian, Chinese, Black, Filipino, Latin American, Arab, Southeast Asian, West Asian, Korean, Japanese, Other (Please specify) | Rather Not Say, Refused, NA |

*Note*. NA = Not applicable.

Table S2: Classification table of self-reported racial/ethnic identities provided by free text response in CLSA and CanPath studies. Participants who identified as non-white in race/ethnicity, or who identified as Indigenous, were classified as a racialized minority. Ambiguous or missing responses were removed for all analyses which included race/ethnicity as a stratum.

| **Response classification** | |
|---|---|
| **White** | **Racialized minority** |
| Responses which included "White" or "Caucasian". | Responses which included "Caribbean", "Indian", "Indo-Caribbean", "Asian", "African", "South American", "Iraqi", "Ugandan", "Trinidadian", "Chinese", "Japanese", "Armenian", "Black", "Guyanese", "Latin-American", "Khoisan", "Hispanic", "Persian", "Assyrian", "Latino", "Afro-Latino American", "Berbere", "Amazigh", "Maghrebine", "Moroccan Jewish", "Arab", "Afrikaans", "Jamaican", "Indo-European", "Colombian", "Middle East", "Cherokee", "Zimbawayan", "Indonesian", "Punjabi", "Cambodian", "Maurician", "Taiwanese", "Afghan", "Tunisian", "Mexican", "Metis", "Aboriginal", or "Indigenous". |

Table S3: Summary of demographic characteristics by research study. Scores of 1 and 5 indicate the lowest and highest quantiles of deprivation, respectively. Raw CCAHS-1 counts were rounded to base 2000 according to data usage guidelines.

| | CBS blood donor | APL outpatient laboratory | CCAHS-1 closed cohort | CLSA closed cohort | CanPath closed cohort | Ab-C open cohort |
|---|---|---|---|---|---|---|
| n | 1035580 | 210905 | 11050 | 17310 | 21720 | 25110 |
| **Age Group (%)** | | | | | | |
| 0-17 | 0 (0.0) | 7111 (3.4) | 1750 (15.8) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| 18-26 | 109542 (10.6) | 11763 (5.6) | 1300 (11.8) | 0 (0.0) | 3 (0.0) | 991 (3.9) |
| 27-36 | 183368 (17.7) | 26354 (12.5) | 1200 (10.9) | 0 (0.0) | 355 (1.6) | 4244 (16.9) |
| 37-46 | 176478 (17.0) | 27646 (13.1) | 1450 (13.1) | 0 (0.0) | 1226 (5.6) | 4219 (16.8) |
| 47-56 | 192898 (18.6) | 30443 (14.4) | 1550(14.0) | 1485 (8.6) | 3687 (17.0) | 4396 (17.5) |
| 57+ | 373294 (36.0) | 107588 (51.0) | 3800 (34.4) | 15825 (91.4) | 16449 (75.7) | 11260 (44.8) |
| **Material deprivation quintile (%)** | | | | | | |
| 1 | 267483 (25.8) | 54152 (25.7) | 2550 (23.1) | 5345 (30.9) | 0 (0.0) | 0 (0.0) |
| 2 | 225349 (21.8) | 38727 (18.4) | 2250(20.4) | 3884 (22.4) | 0 (0.0) | 0 (0.0) |
| 3 | 188746 (18.2) | 32684 (15.5) | 1950 (17.6) | 3132 (18.1) | 0 (0.0) | 0 (0.0) |
| 4 | 145317 (14.0) | 26976 (12.8) | 1850 (16.7) | 2404 (13.9) | 0 (0.0) | 0 (0.0) |
| 5 | 85043 (8.2) | 17743 (8.4) | 1450 (13.1) | 1659 (9.6) | 0 (0.0) | 0 (0.0) |
| Missing | 123642 (11.9) | 40623 (19.3) | 950 (8.6) | 886 (5.1) | 21720 (100.0) | 25110 (100.0) |
| **Social deprivation quintile (%)** | | | | | | |
| 1 | 197381 (19.1) | 37530 (17.8) | 1600 (14.5) | 2822 (16.3) | 0 (0.0) | 0 (0.0) |
| 2 | 194957 (18.8) | 28600 (13.6) | 2100 (19.0) | 3432 (19.8) | 0 (0.0) | 0 (0.0) |
| 3 | 182655 (17.6) | 35489 (16.8) | 2250 (20.4) | 3479 (20.1) | 0 (0.0) | 0 (0.0) |
| 4 | 167547 (16.2) | 33994 (16.1) | 2200 (19.9) | 3426 (19.8) | 0 (0.0) | 0 (0.0) |
| 5 | 169398 (16.4) | 34669 (16.4) | 1950 (17.6) | 3265 (18.9) | 0 (0.0) | 0 (0.0) |
| Missing | 123642 (11.9) | 40623 (19.3) | 950 (8.6) | 886 (5.1) | 21720 (100.0) | 25110 (100.0) |
| **Race/ethnicity (%)** | | | | | | |
| Racialized minority | 189304 (18.3) | 0 (0.0) | 1400 (12.7) | 461 (2.7) | 1509 (6.9) | 5296 (21.1) |
| White | 846276 (81.7) | 0 (0.0) | 9200 (83.3) | 16389 (94.7) | 19829 (91.3) | 19633 (78.2) |
| Missing | 0 (0.0) | 210905 (100.0) | 450 (4.1) | 460 (2.7) | 382 (1.8) | 181 (0.7) |
| **Sex (%)** | | | | | | |
| Female | 433366 (41.8) | 119362 (56.6) | 6150 (55.7) | 9045 (52.3) | 14258 (65.6) | 14815 (59.3) |
| Male | 602214 (58.2) | 91530 (43.4) | 4900 (44.3) | 8265 (47.7) | 7462 (34.4) | 10157 (40.7) |
| Missing | 0 (0.0) | 13 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| **Urbanicity (%)** | | | | | | |
| Rural | 133393 (12.9) | 26045 (12.3) | 1950 (17.6) | 2429 (14.0) | 1870 (8.6) | 3156 (12.6) |
| Urban | 902180 (87.1) | 184859 (87.7) | 9050 (81.9) | 14881 (86.0) | 19850 (91.4) | 21923 (87.4) |
| Missing | 7 (0.0) | 1 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |

Table S4: Sensitivity analysis on the percentage of study demographic subgroups with greater than 25 collected specimens after stratification. Study participants identifying as mixed race/ethnicity (white and racialized minority) classified as white. Date of sample collection was binned into 2-month intervals. Subgroups with counts above threshold value might produce more stable estimates when statistically adjusted. All specimen counts were unweighted.

| | | Demographic subgroups | | | |
|---|---|---|---|---|---|
| Study (specimen count) | Months sampled | Age, Sex, Province, Month | Age, Sex, Province, Urban, Month | Age, Sex, Province, Race/Ethnicity, Month | Age, Sex, Province, Race/Ethnicity, Urban, Month |
| CBS blood donor (1,035,580) | 41 | 92% | 74% | 70% | 52% |
| APL outpatient laboratory (210,905) | 27 | 94% | 84% | NA | NA |
| Ab-C open cohort (25,110) | 18 | 31% | 20% | 21% | 14% |
| CanPath closed cohort (21,720) | 10 | 40% | 31% | 33% | 27% |
| CLSA closed cohort (17,310) | 11 | 50% | 36% | 34% | 27% |
| CCAHS-1 closed cohort (11,050) | 6 | 33% | 20% | 21% | 13% |

# Supplemental figures

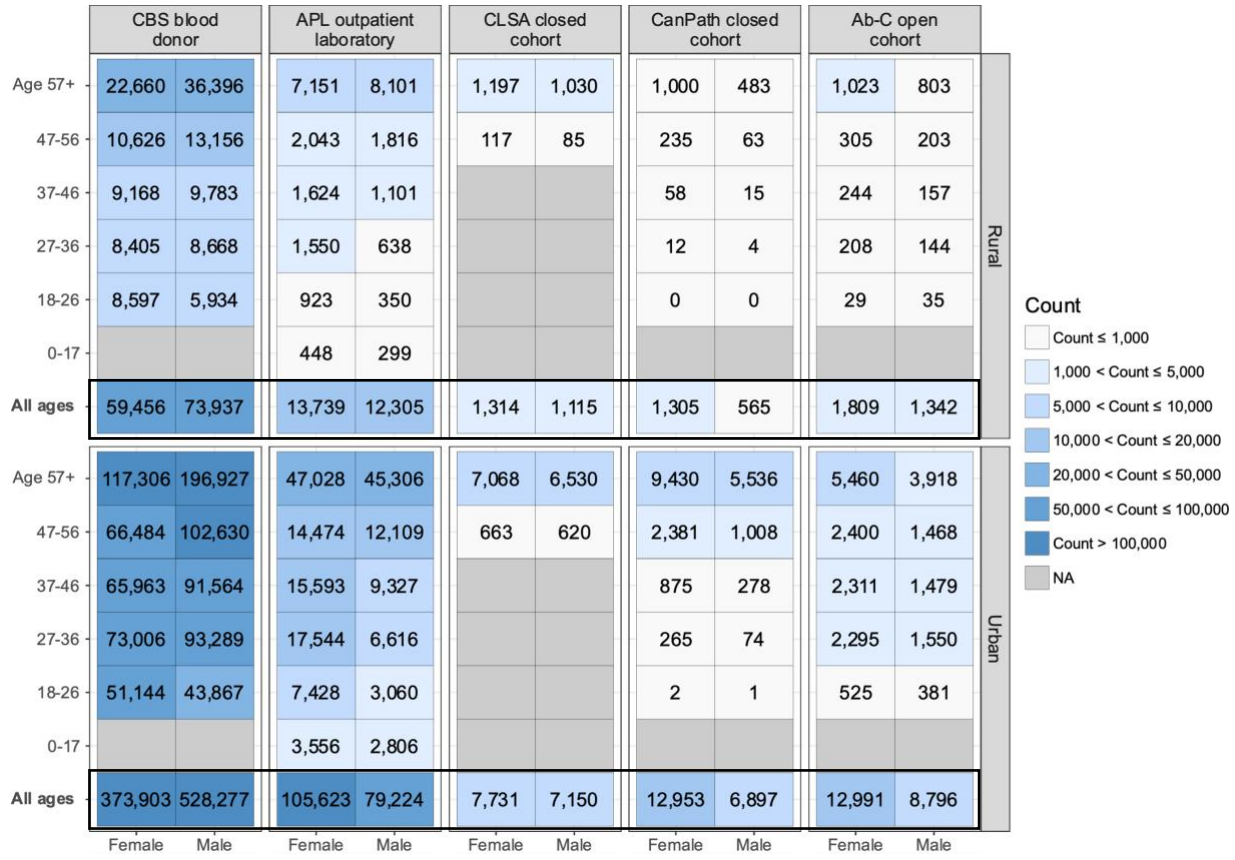


Figure S1: Demographic composition of Canadian serological studies by participant age, sex, and urbanicity. Counts were calculated as the number of serological specimens contributed by each study subgroup. CBS tested 3410 donations from 17-year-old donors, but they were excluded from our analysis. CCAHS-1 counts were not included due to privacy regulations.

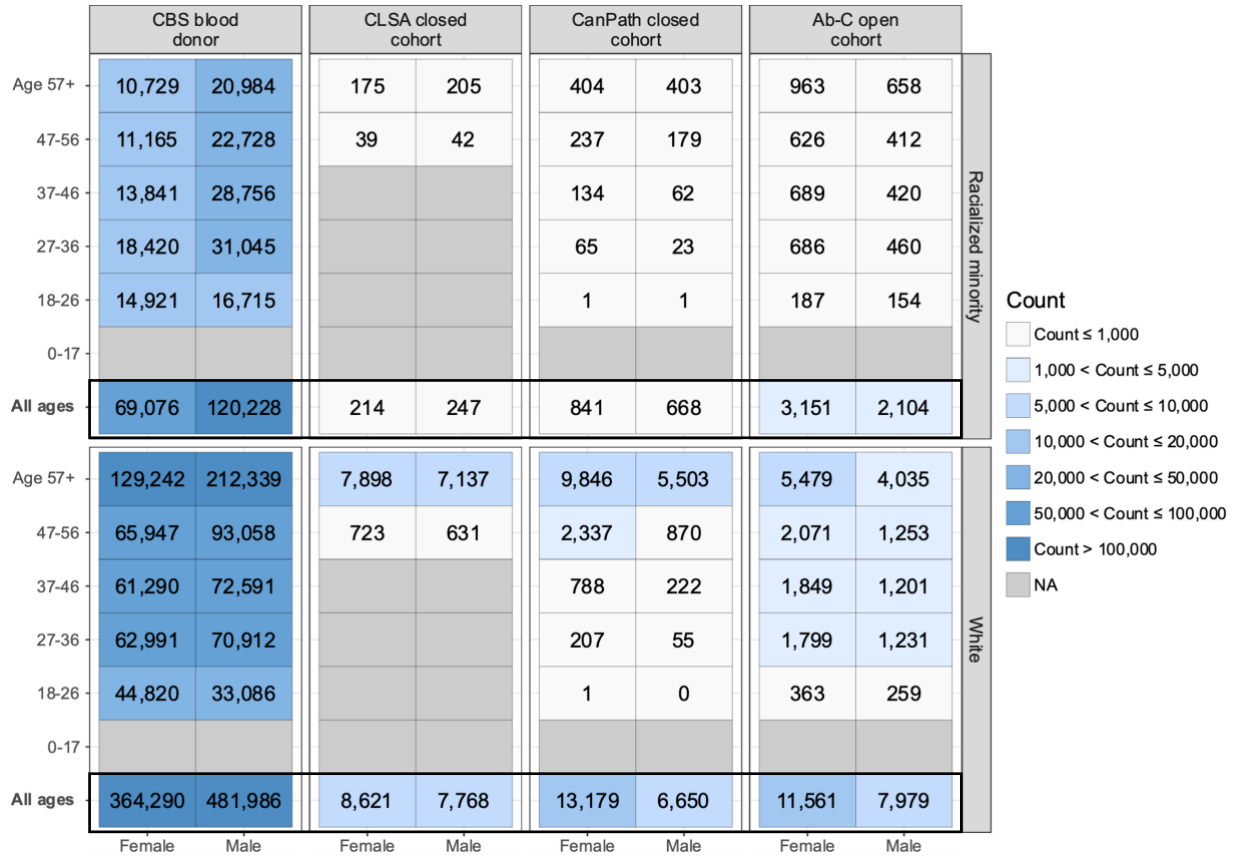


Figure S2: Demographic composition of Canadian serological studies by participant age, sex, and self-identified race/ethnicity. Counts were calculated as the number of serological specimens contributed by each study subgroup. CBS tested 3410 donations from 17-year-old donors, but they were excluded from our analysis. CCAHS-1 counts were not included due to privacy regulations.

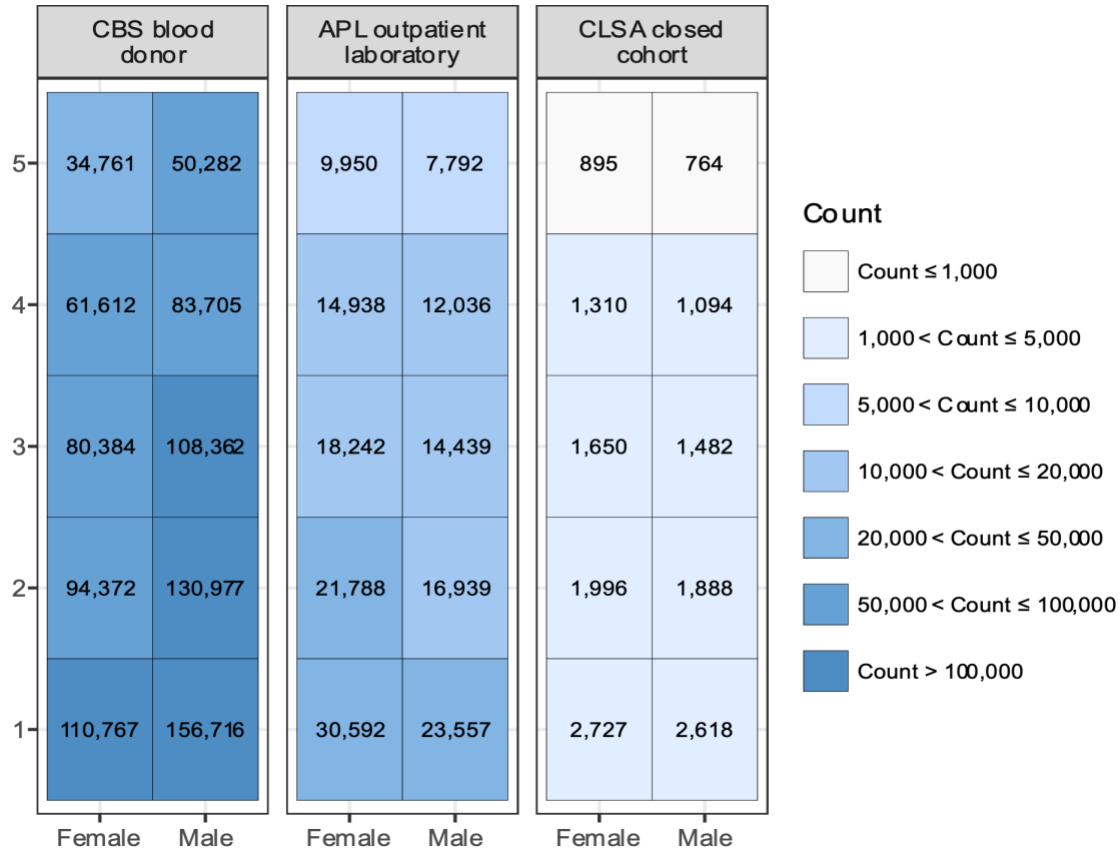


Figure S3: Demographic composition of Canadian serological studies by participant sex and material deprivation quintile score. Counts were calculated as the number of serological specimens contributed by each study subgroup. Scores of 1 and 5 indicate the lowest and highest quintiles of deprivation, respectively. CCAHS-1 counts were not included due to privacy regulations.

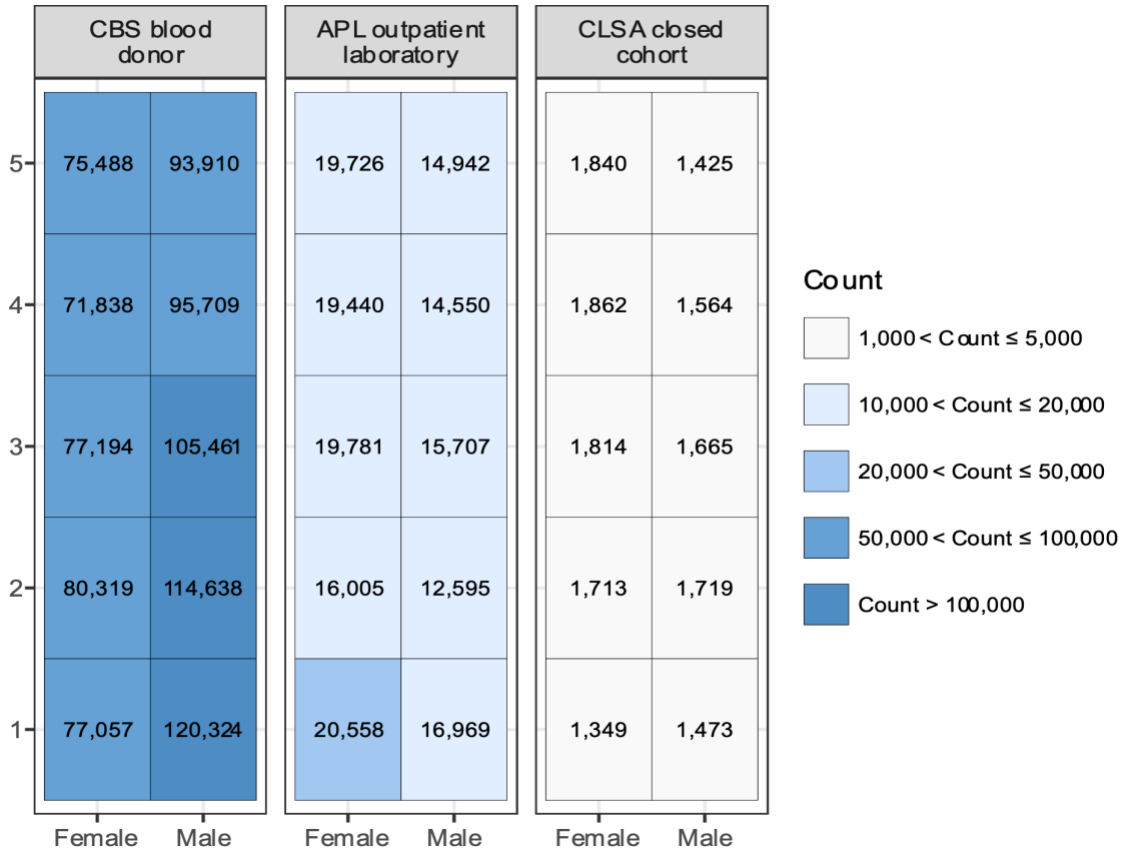


Figure S4: Demographic composition of Canadian serological studies by participant sex and social deprivation quintile score. Counts were calculated as the number of serological specimens contributed by each study subgroup. Scores of 1 and 5 indicate the lowest and highest quintiles of deprivation, respectively. CCAHS-1 counts were not included due to privacy regulations.

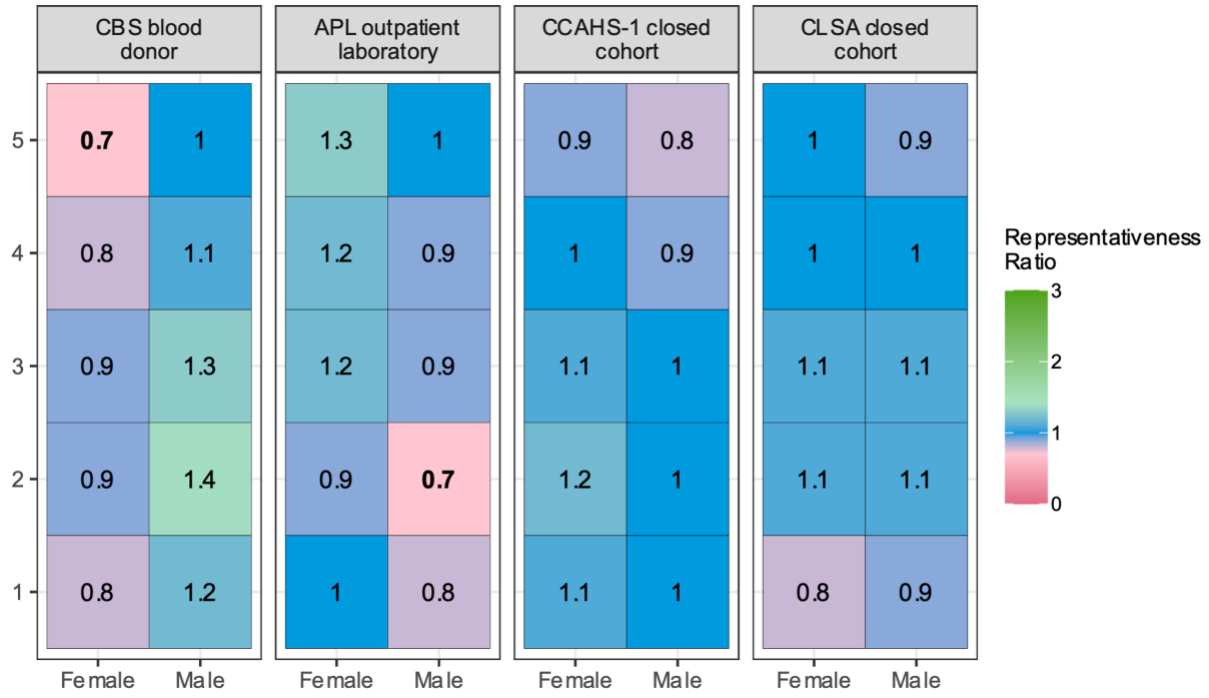


Figure S5: Sociodemographic representativeness of Canadian serological studies compared to the general population by sex and social deprivation quintile. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Social deprivation scores were not available for Ab-C and CanPath studies. Bolded representation ratios indicate greater than 95% of subgroup bootstrap replicates produced representation ratios below 0.75. Bootstrapping was not performed for studies with weighted counts (CCAHS-1).

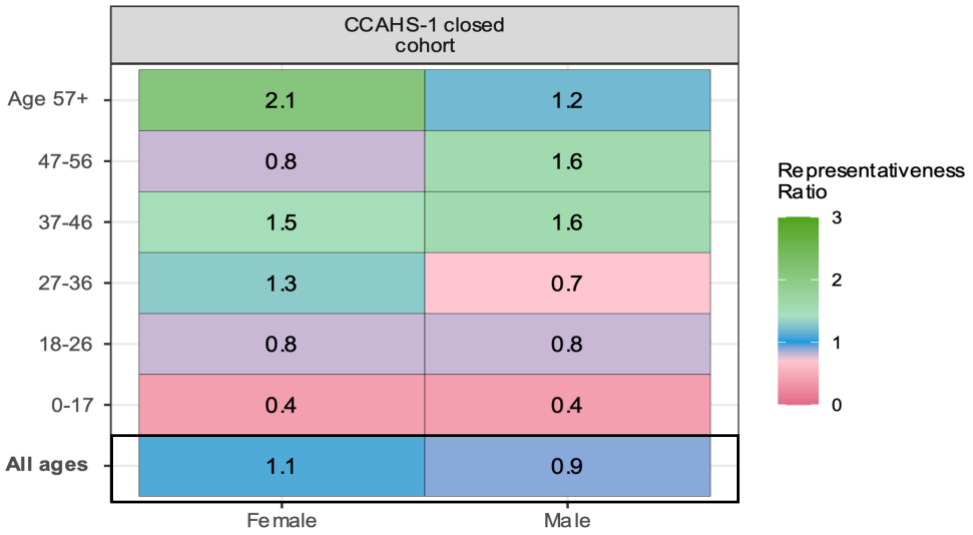


Figure S6: Demographic representativeness of CCAHS-1 serological study specimens collected from the capital cities of the Canadian territories compared to the general population by age group and sex. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Bootstrapping was not performed because representation ratios were calculated using weighted study counts. The population distribution of the weighted CCAHS-1 data was assumed to reflect the total territorial population distribution for this analysis.

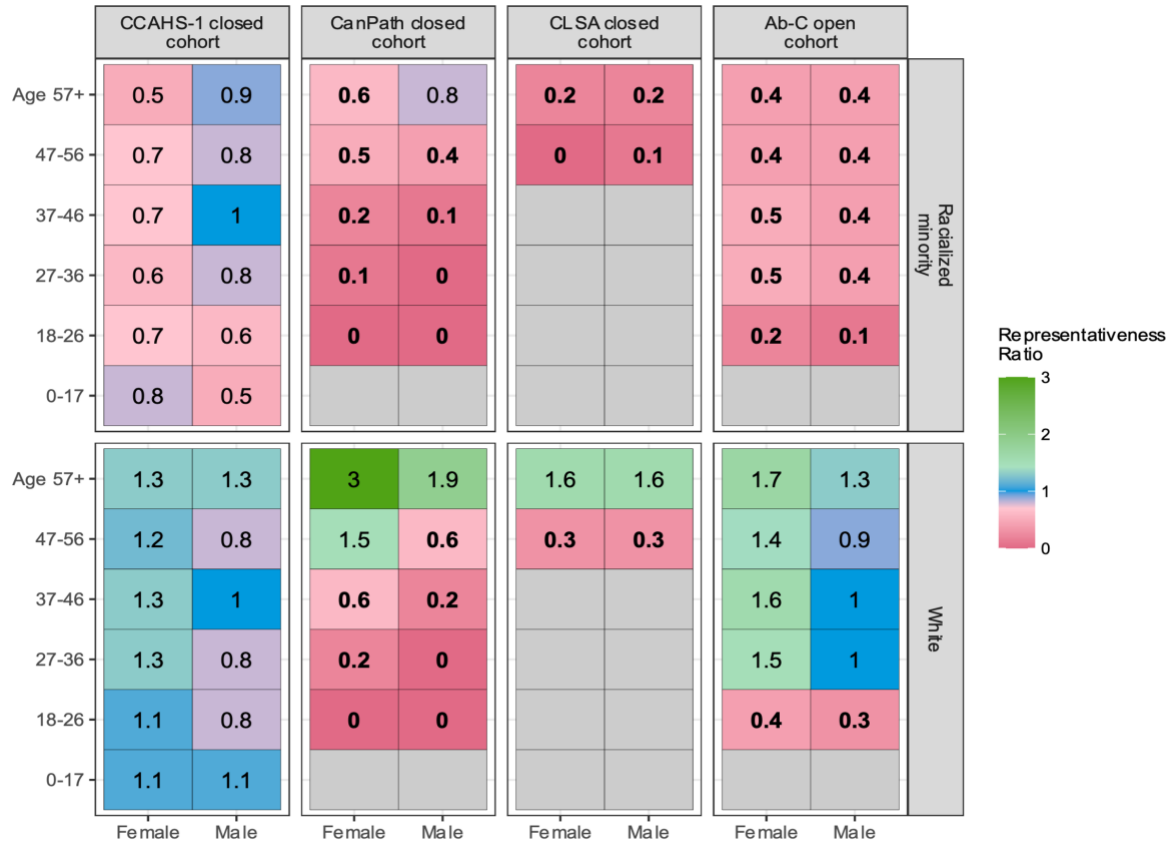


Figure S7: Sensitivity analysis on the demographic representativeness of Canadian serological studies compared to the general population where study participants identifying as mixed race/ethnicity were classified as white. Representativeness was assessed by age group, sex, and racial/ethnic identity. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Bolded representation ratios indicate greater than 95% of subgroup bootstrap replicates produced representation ratios below 0.75. Bootstrapping was not performed for studies with weighted counts (CCAHS-1).

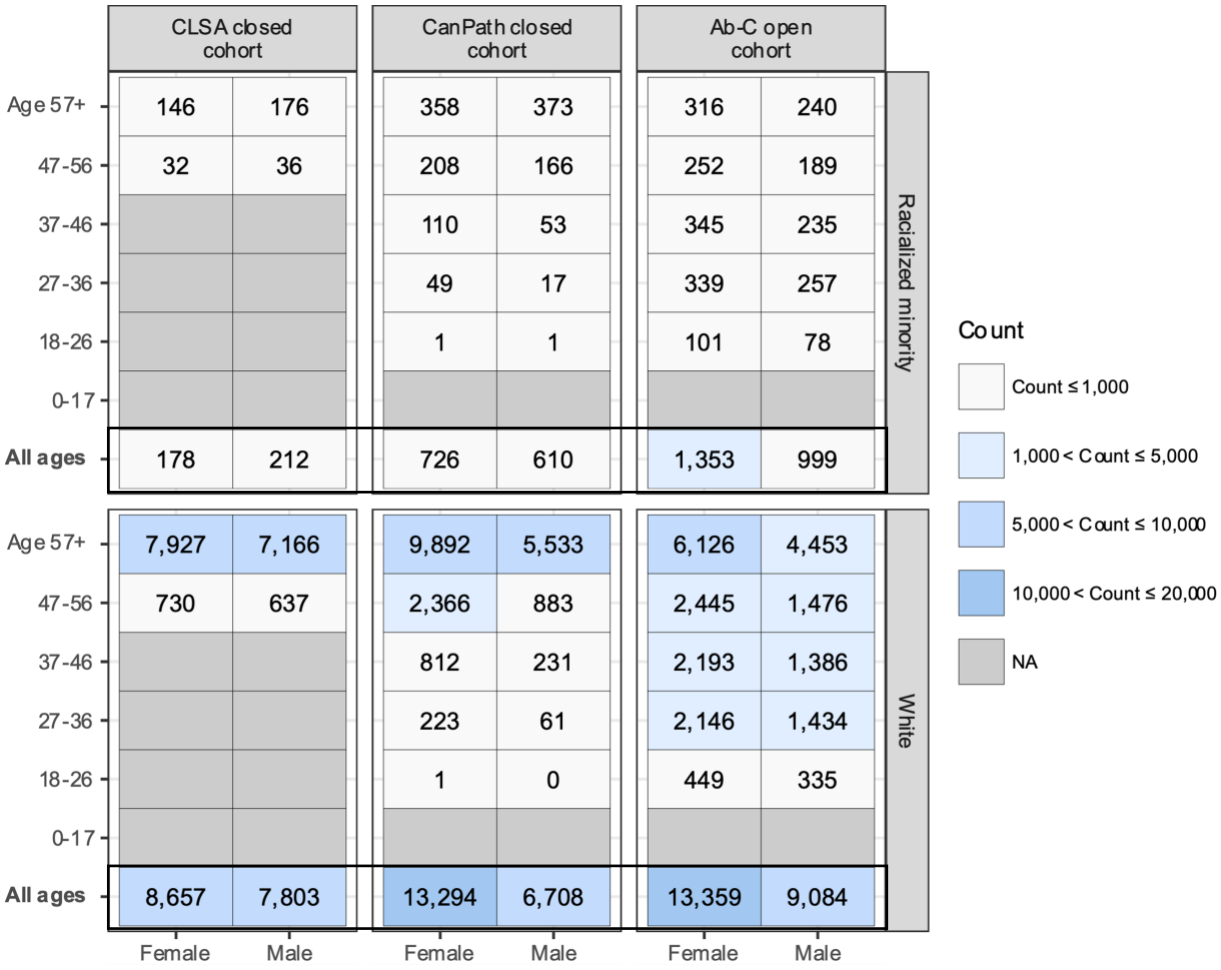


Figure S8: Sensitivity analysis on the demographic composition of Canadian serological studies by participant age, sex, and self-identified race/ethnicity where study participants identifying as mixed race/ethnicity were classified as white. Counts were calculated as the number of serological specimens contributed by each study subgroup. CCAHS-1 counts were not included due to privacy regulations.

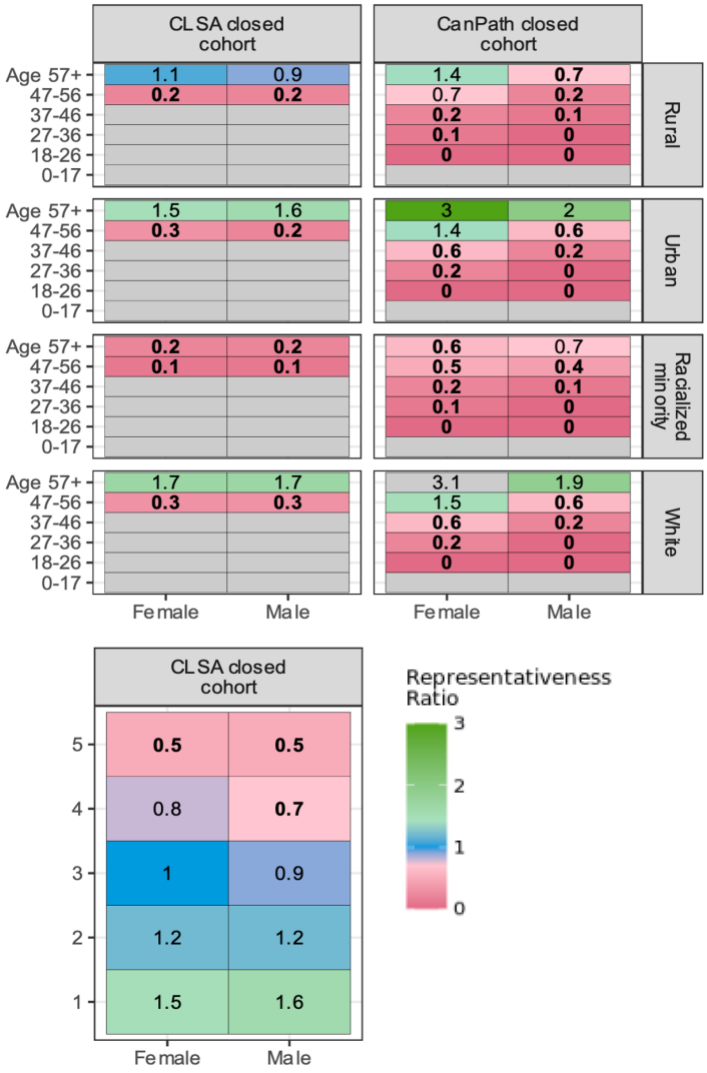


Figure S9: Sensitivity analysis on the demographic representativeness of Canadian serological studies compared to the general population where total population counts were estimated using the 2016 Canadian census and included Indigenous-identifying individuals.[19] Representativeness was assessed by age group, sex, urbanicity, racial/ethnic identity, and material deprivation quintile. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Material deprivation scores were not available for the CanPath study. Bolded representation ratios indicate greater than 95% of subgroup bootstrap replicates produced representation ratios below 0.75.

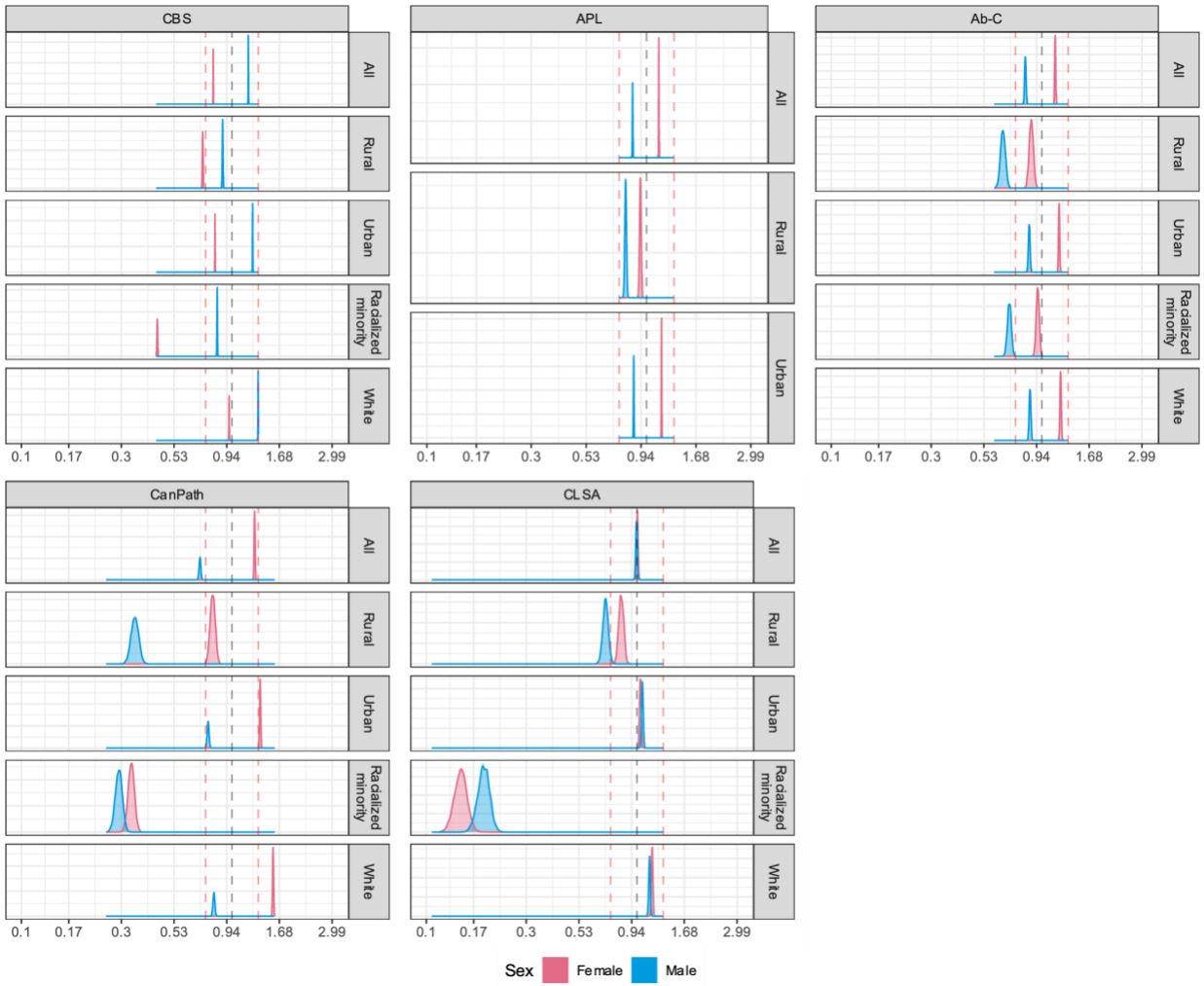


Figure S10: Bootstrap distribution of Canadian serological study log10-transformed representation ratios stratified by sex, urbanicity, and self-identified racial/ethnic identity. Distributions were generated using 5000 bootstrap resamples and plotted with a binwidth of 0.01. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of

population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Dashed red lines indicate representation ratio values of 0.75 and 1.33.

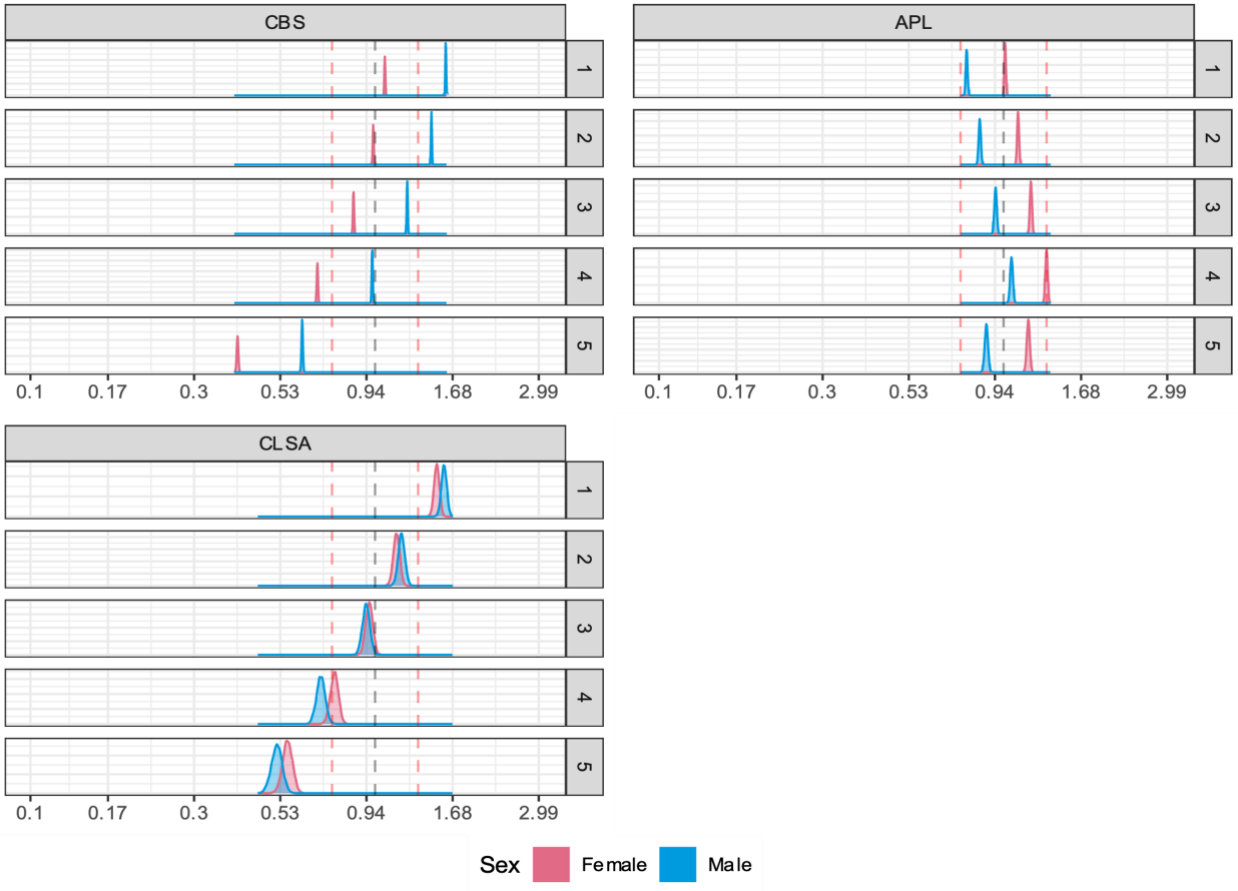


Figure S11: Bootstrap distribution of Canadian serological study log10-transformed representation ratios stratified by sex and material deprivation quintile. Distributions were generated using 5000 bootstrap resamples and plotted with a binwidth of 0.01. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Dashed red lines indicate representation ratio values of 0.75 and 1.33.

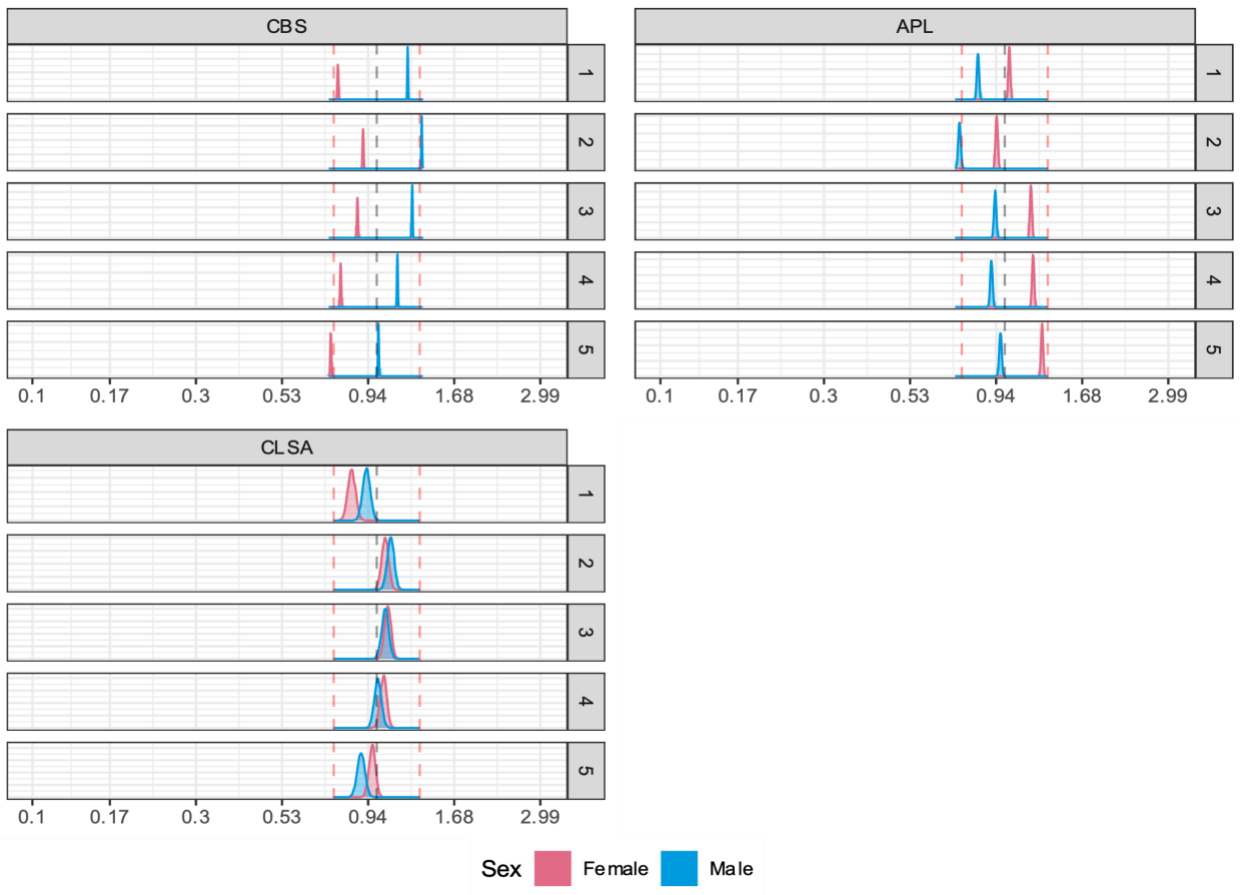


Figure S12: Bootstrap distribution of Canadian serological study log10-transformed representation ratios stratified by sex and social deprivation quintile. Distributions were generated using 5000 bootstrap resamples and plotted with a binwidth of 0.01. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Dashed red lines indicate representation ratio values of 0.75 and 1.33.

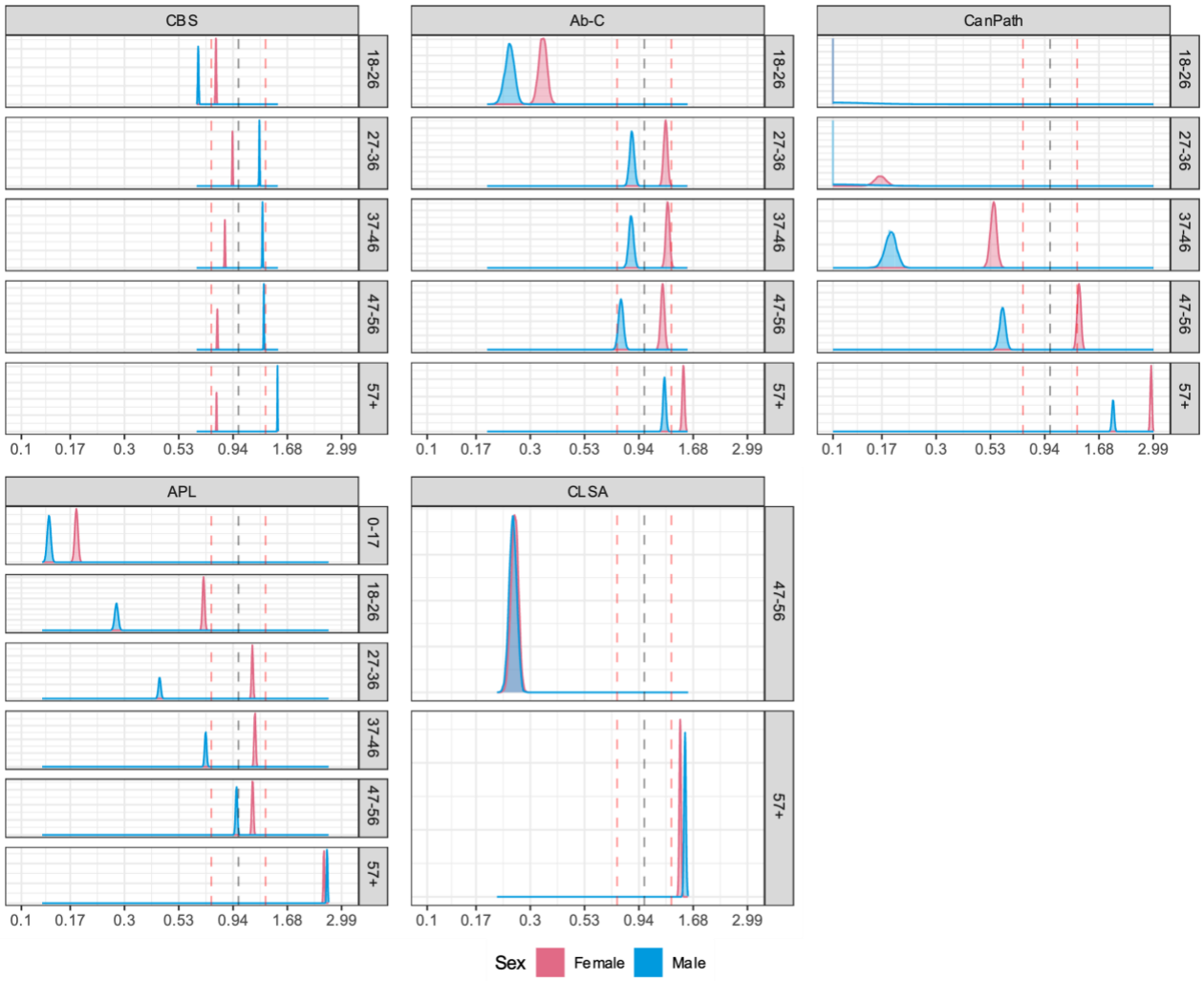


Figure S13: Bootstrap distribution of Canadian serological study log10-transformed representation ratios stratified by age group, sex, and urban residence. Distributions were generated using 5000 bootstrap resamples and plotted with a binwidth of 0.01. Representation ratios between 0.00 – 0.09 were pseudo-adjusted to a value of 0.1 prior to transformation for visualization purposes.

Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Dashed red lines indicate representation ratio values of 0.75 and 1.33.

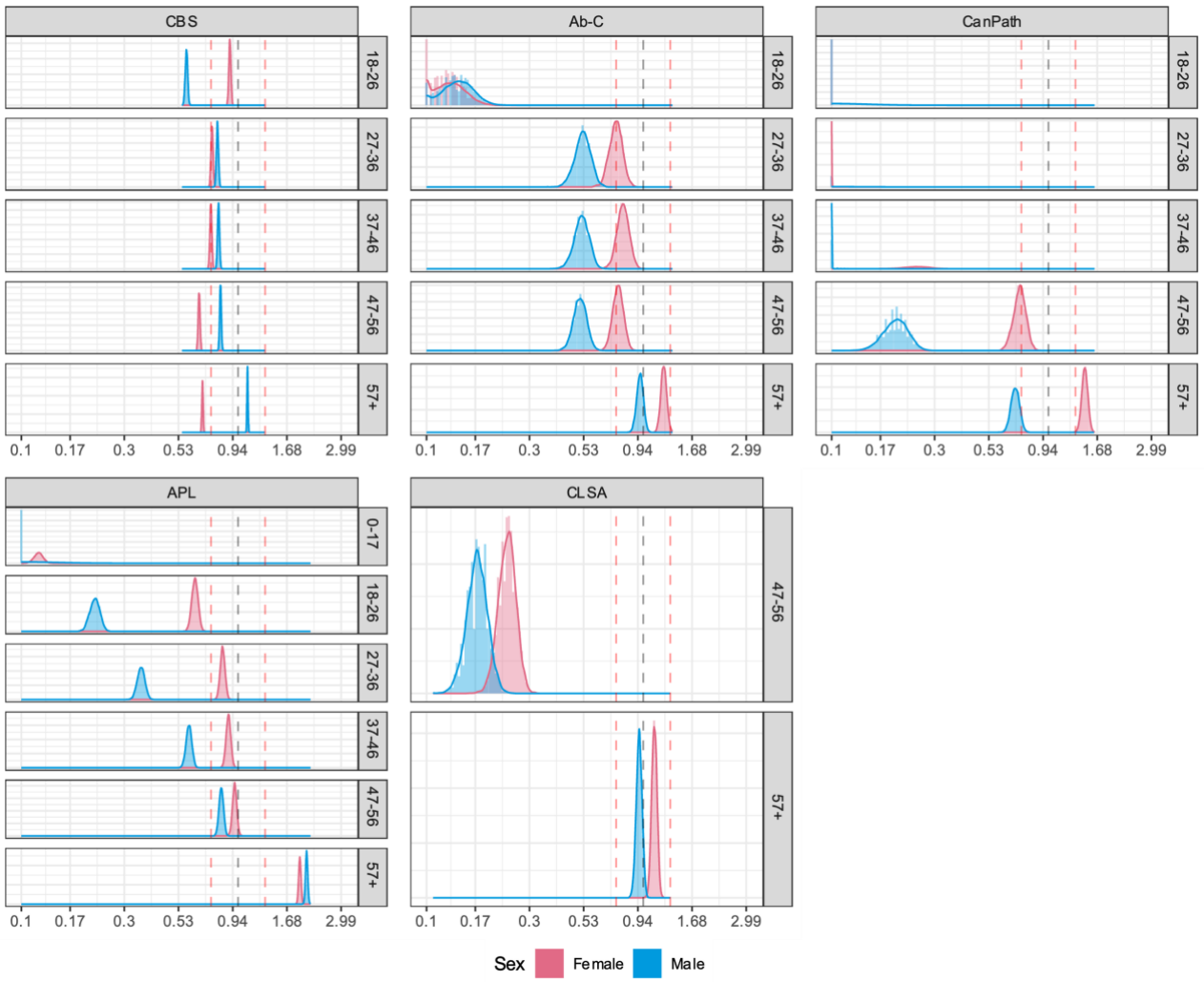


Figure S14: Bootstrap distribution of Canadian serological study log10-transformed representation ratios stratified by age group, sex, and rural residence. Distributions were generated using 5000 bootstrap resamples and plotted with a binwidth of 0.01. Representation ratios between 0.00 – 0.09 were pseudo-adjusted to a value of 0.1 prior to transformation for visualization purposes.

Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Dashed red lines indicate representation ratio values of 0.75 and 1.33.

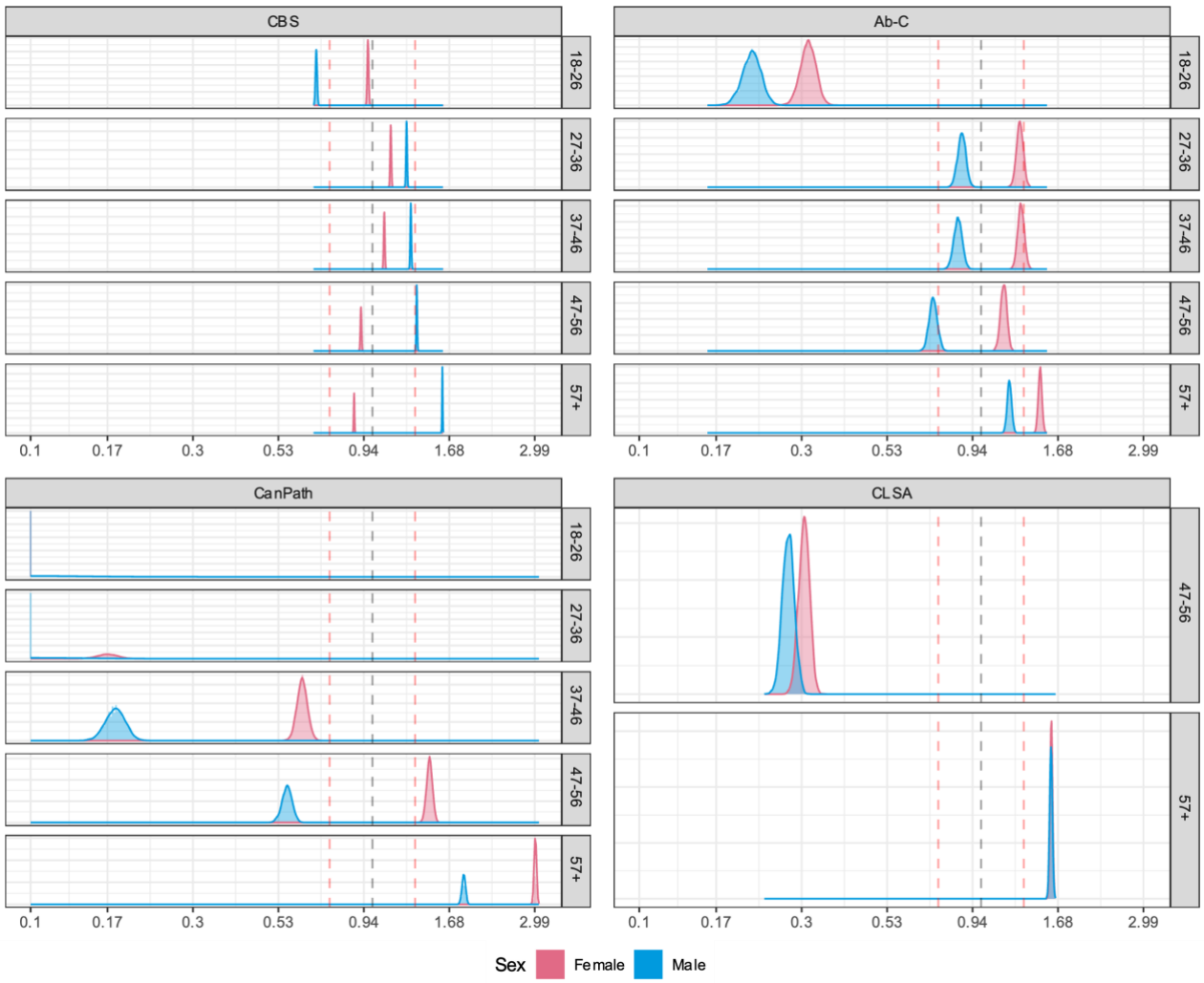


Figure S15: Bootstrap distribution of Canadian serological study log10-transformed representation ratios stratified by age group, sex, and white racial/ethnic identity. Distributions were generated using 5000 bootstrap resamples and plotted with a binwidth of 0.005. Representation ratios between 0.00 – 0.09 were pseudo-adjusted to a value of 0.1 prior to transformation for visualization purposes.

Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Dashed red lines indicate representation ratio values of 0.75 and 1.33.

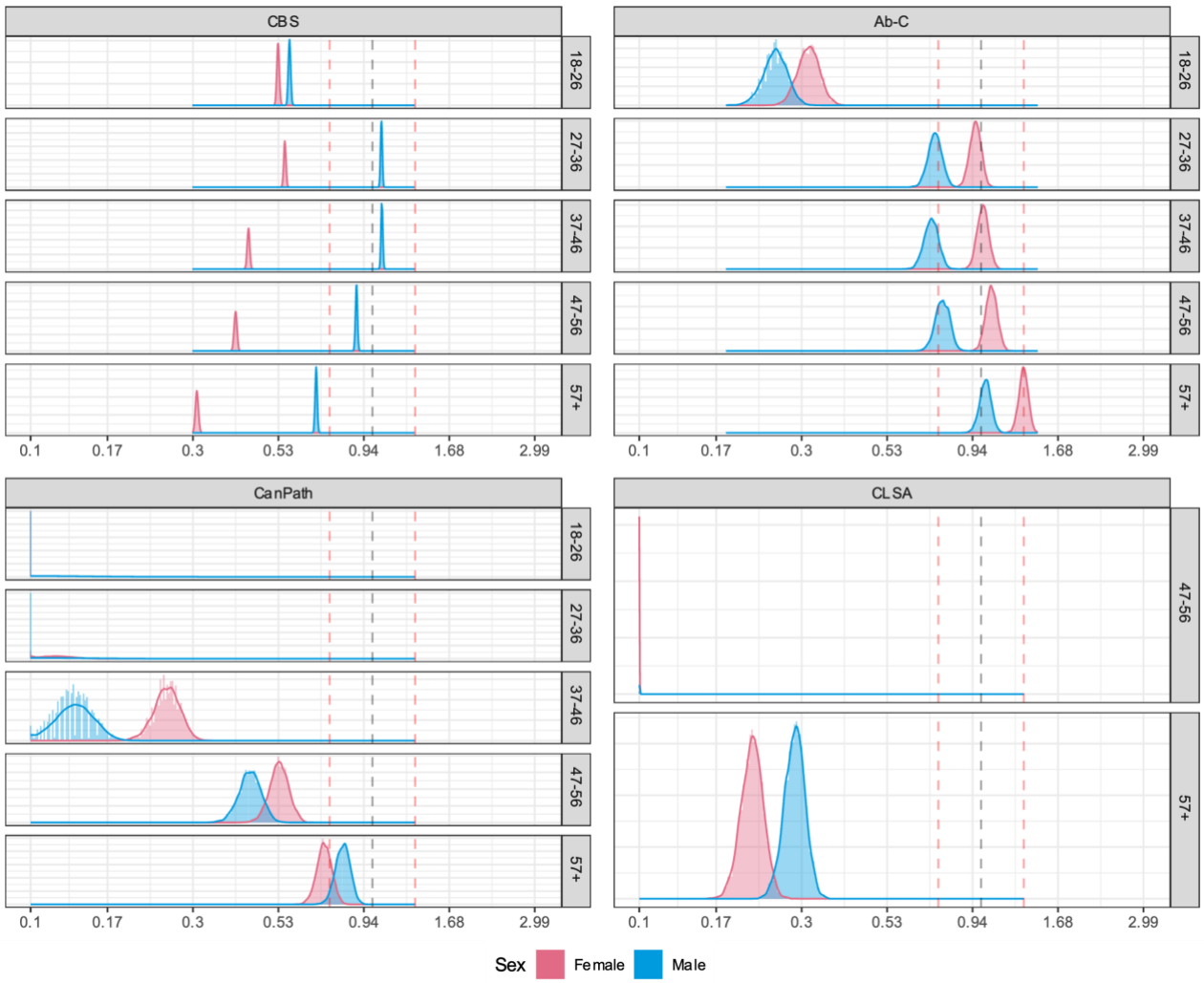


Figure S16: Bootstrap distribution of Canadian serological study log10-transformed representation ratios stratified by age group, sex, and racialized minority racial/ethnic identity. Distributions were generated using 5000 bootstrap resamples and plotted with a binwidth of 0.005. Representation ratios between 0.00 – 0.09 were pseudo-adjusted to a value of 0.1 prior to transformation for visualization

purposes. Representativeness was calculated by dividing the proportion of study specimens collected from a subgroup by the proportion of population in the subgroup. Total population counts were estimated using the 2016 Canadian census.[19] Dashed red lines indicate representation ratio values of 0.75 and 1.33.

## 4.3 Postscript

In this manuscript, I assessed the sociodemographic representativeness of six SARS-CoV-2 serosurveillance studies conducted in Canada. I identified key demographic subgroups, such as racialized minorities and 18-26-year-olds, which were commonly underrepresented across all study designs. I also identified several differences in the representation of studies with similar designs, reflecting the role of recruitment strategy and chosen study population on representativeness to the general Canadian population.

This study emphasizes the need for future research to identify and reduce the barriers impeding racialized minorities' representation in health research. While stratified random sampling and weighting methods may improve study representation, they do not address the personal and societal factors that drive racialized communities' decision to participate in health research. This study also highlights that non-probability data sources may produce samples that are representative of the general population along multiple sociodemographic dimensions, although the lack of a formal sampling frame may make it harder to formally assess generalizability and generate weights for statistical adjustment.[106]

# Chapter 5: Discussion

I observed variability in the sociodemographic representativeness of six Canadian serosurveillance studies that used different recruitment strategies. Racialized minorities were underrepresented (representation ratio < 1) in all serosurveillance studies, regardless of design, although representation improved with age in pre-existing longitudinal cohorts (Action to Beat Coronavirus [Ab-C], Canadian Longitudinal Study on Aging [CLSA], Canadian Partnership for Tomorrow's Health [CanPath]). Individuals residing in materially deprived neighborhoods were underrepresented in all studies except the Alberta Precision Laboratories (APL) convenience sample. Urban regions were consistently better represented relative to their urban counterparts in all studies, although multiple studies (APL convenience sample, CLSA closed longitudinal cohort, Canadian COVID-19 Antibody and Health Survey 1 [CCAHS-1] de novo probabilistic cohort) were generally representative of rural residents aged 57 or older across both sexes. Notably, no single study was sufficiently representative of all sociodemographic subgroups included in this analysis. Some aspects of the study design and findings warrant further discussion beyond those presented in Chapter 4.2.

The underrepresentation of racialized minority groups has also been observed in other severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) serosurveillance studies.[15,16] In some studies, imbalances persisted despite oversampling from regions with a higher proportion of non-white individuals.[16] Because statistical weighting methods may fail to effectively adjust for subgroups largely excluded from the study sample, identifying barriers to recruitment at the sampling phase is key to improve study representativeness.[30] Prior systemic racism and negative experiences from medical and research institutions may influence the decision of individuals from racialized communities to participate in health research.[142,143] Lack of awareness of

research activities, language barriers, and a perceived lack of personal benefit may also contribute to the underrepresentation of certain racialized communities in study samples.[142,143] Community engagement is a common recommendation to improve the recruitment of socially disadvantaged groups in health research and to minimize attrition. Collaborating with community stakeholders in the design, implementation, and ownership of the research data may help increase study awareness and reduce fears regarding data usage.[142] Given racialized minorities experience disproportionately greater SARS-CoV-2 burden compared to self-identified white individuals,[15,28,73,74,99] identifying and reducing the barriers to participation is essential to improve racialized minorities' representation in future serosurveillance studies.

Several other notable differences in sociodemographic representation were observed in this study. Strikingly, the APL convenience sample was the only study that was sufficiently representative across all sex and material deprivation strata (representation ratio 0.8-1.3). The universal need for healthcare and Canada's publicly funded healthcare system may result in a more socioeconomically diverse study population compared to populations who self-select to donate blood or to participate in health research studies. However, provincial representativeness of specimens from 0-17-year-olds was markedly superior in the CCAHS-1 de novo probabilistic cohort compared to the APL convenience sample, suggesting stratified sampling techniques may be better suited to reach younger populations. Among convenience samples, the APL study was equally or more representative across sexes of rural residents aged 47 or older compared to the Canadian Blood Services (CBS) study. This may result from increased healthcare utilization by older populations or reflect easier access to healthcare clinics than blood collection sites, which tend to be in urban areas.[144]

The use of convenience samples for population health research has traditionally been criticized due to the potential biases inherent in non-probability sampling.[106,145] The lack of a sampling frame impedes estimation of sampling biases, while self-selection and recruitment strategies may introduce selection bias if the distribution of characteristics in the sample differs from the target population.[106,118] Characterizing the representativeness of residual blood convenience samples may be further challenged if data providers do not collect data on key variables related to participation, such as vaccination status and prior infection history, or rely on participant self-recall.[14,146] Despite these limitations, this analysis found that residual blood convenience samples can produce sufficiently representative samples along multiple sociodemographic dimensions. Data linkage with supplementary probabilistic population surveys may improve statistical inference from convenience samples. Future studies should assess the utility of statistical methods such as propensity score weighting, inverse probability weighting, and poststratification to improve sample representativeness and adjust for non-probability sampling biases.[20,118]

Several aspects of study design may have influenced the representativeness of the pre-existing longitudinal cohorts included in this study. For example, the Ab-C open longitudinal cohort was recruited from a marketing research panel that provided rewards for participation. Although weighting was performed to correct for higher education levels compared to the general population,[25] latent characteristics inherent to individuals who participate in marketing studies may bias associations if causally related to recruitment and the outcome of interest.[44] Since the CLSA and CanPath closed longitudinal cohorts were conceived in 2010[26] and 2009,[27] respectively, they are not representative of younger age groups. While this demographic profile is a natural consequence of using pre-existing longitudinal cohorts for serosurveillance, it limits

their generalizability compared to probabilistic designs such as CCAHS-1. The representativeness of studies relying on venous collection as part of their specimen collection strategy (CLSA closed longitudinal cohort, APL convenience sample, CBS convenience sample) may have also been limited by participant concerns and hesitancy to travel for specimen collection, particularly for older individuals and other high-risk populations. The use of at-home collection devices may help mitigate these participation barriers, although they may provide reduced test performance compared to venous testing methods.[114]

The relatively limited range of sociodemographic variables used to assess representativeness is a key limitation of this analysis. Numerous additional social and clinical characteristics likely differ between the included study populations and the general population. Participants from the original CLSA and CanPath closed longitudinal cohorts have been shown to have higher education status and/or income compared to the general population.[26,27] This may result in selection bias or confounding if not appropriately adjusted for during analysis,[44] yet this data may be unavailable for blood donor or healthcare study populations. Furthermore, uncharacterized differences in the health status of blood donors or healthcare study populations may produce samples that have generally better or worse health, respectively, than the general population.[13] Health-seeking behaviors, testing practices, and self-selection may all introduce potential biases and limit the generalizability of estimates if not properly acknowledged during analysis.[6,13,107] Future work could use administrative healthcare data to better characterize differences in the distribution of health-related variables between residual blood cohorts and the general population.

The study was also constrained by the inconsistent measurement of participant race/ethnicity between studies. Differences in the phrasing and response options necessitated the

use of a binary classification as 'white' or 'racialized minority'. This classification obscured any heterogeneity in representativeness between minority groups and prevented identification of specific racial/ethnic subgroups that were poorly represented in the various study designs. Prior studies have identified minority groups are differentially represented in SARS-CoV-2 study populations,[16,147] yet the lack of a standardized classification for participant race/ethnicity limited the ability to assess this across the included studies. Characterizing the representativeness of Canadian population health studies along racial/ethnic dimensions may also be difficult because race-based data is not routinely available in Canadian administrative healthcare datasets.[148] In contrast, the United States National Institutes of Health provides a standardized framework to collect race-based data,[149] while the United Kingdom's Equality Act mandates the collection of racial/ethnic identity in healthcare records.[150] Implementation of standardized collection and reporting systems for race may aid comparison of racialized communities' representation across population health studies.[148]

There are other limitations in this analysis. The assessment of representativeness did not account for any specific objectives of the included studies. For example, the Ab-C open cohort oversampled participants aged 60 and older to ensure sufficient power for subgroup analyses,[25] but this study assumed representativeness along each stratum was of equal importance to investigators. The representation ratios for the CCAHS-1 de novo probabilistic cohort territorial analysis may be susceptible to coverage error. Because the weighted counts of the CCAHS-1 and 2016 Canadian census were designed to be representative of the three territorial capital cities and the entire territorial population, respectively, bias may be present if the population distribution differs between these two regional subsets.[109,151] Finally, although survey weights were also available for the Ab-C open longitudinal cohort, representativeness was assessed using

unweighted data for all datasets except CCAHS-1. Because the unweighted CCAHS-1 sample was solely designed to contain a sufficient amount of specimens per strata, use of the weighted dataset was required to accurately capture the study design.[109] Unweighted study data was used for the remainder of the analysis to evaluate the influence of recruitment strategy, rather than effectiveness of statistical adjustment, on the representativeness of the included SARS-CoV-2 serosurveillance studies.

Several additional factors should be considered when interpreting the results of this study. Following the framework presented in Rudolph et al., this study estimated the capability of six serosurveillance studies to produce interpretations of effect that are representative of the census-based general population. While this approach assumes the interpretation of effect will be similar between the study and target populations with similar demographic characteristics, it does not imply representativeness of estimated population parameters.[43] The mechanism of selection for residual blood cohorts also directly influences the forms of research that are appropriate. For example, blood donors are an accessible, convenient population for public health surveillance, but are an impractical study population for surveillance of sexually transmitted infections since participants undergo screening prior to donation.[30] However, given their low operational costs, established testing infrastructure, and capability to repeat specimen collection,[7,104] residual blood cohorts may be better suited to monitor changes in general population health status over time compared to the other serosurveillance study designs assessed here.

In summary, this analysis identified considerable differences in the sociodemographic representativeness of six SARS-CoV-2 serosurveillance studies conducted in Canada. Racialized minorities and young individuals were generally underrepresented across study designs. Future

studies should investigate the utility of targeted recruitment methods to improve the representation of these high-risk subgroups. This study also identified that no single recruitment strategy produced a sufficiently representative sample along all sociodemographic dimensions. This highlights the need to investigate multi-tiered surveillance approaches that may help overcome gaps in representation, although data harmonization is needed to pool data across studies.[152,153] Such multi-tiered approaches could integrate a combination of wastewater testing, population-based serology testing, and genomic sequencing to rapidly detect outbreaks, identify subgroup differences in infection, and classify emerging variants of concern.[6,33]

# Chapter 6: Conclusion

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic created a dire need for efficient population surveillance. Public health restrictions, resource limitations, and a need for near real-time surveillance data forced studies to employ a range of strategies to recruit participants, which likely influenced the characteristics of the resulting study populations. Thus, this thesis aimed to assess the sociodemographic representativeness of six Canadian SARS-CoV-2 serosurveillance studies relative to the general Canadian population. Racialized minorities and young adults were identified to be underrepresented (representation ratio < 1) amongst multiple sociodemographic strata across all recruitment strategies. Notably, nonprobability recruitment strategies were adequately or more representative of the census-based general population along multiple sociodemographic dimensions compared to traditional probability-based approaches.

Public health surveillance is an important tool to monitor and protect the health of the general population. To be effective, surveillance systems should ensure study populations are representative of the general population among demographic, geographic, clinical, or other dimensions relevant to the surveillance objective.[8] While representativeness is not necessary for valid causal inference,[10] representative sampling is necessary when attempting to describe the health status of a larger population that cannot be feasibly measured.[9,43,44] This study supports the conclusions of prior works that racialized minorities are underrepresented in SARS-CoV-2 serosurveillance study populations.[7,23,24] Representativeness tended to improve with increasing age, urban residence, and residence in neighborhoods with low amounts of material deprivation. These observations may be useful considerations when planning future population serosurveillance studies.

The findings from this study emphasize the need to further investigate the barriers impeding underrepresented subgroups' participation in population health research studies. Random sampling and statistical adjustment are often recommended to correct for imbalances in study population representativeness, yet the results of the current study suggest further efforts at the recruitment stage may produce more representative samples. Direct engagement with underrepresented population subgroups is necessary to facilitate trust and develop community engagement.[142] This study also highlighted that residual blood convenience samples may produce representative samples along multiple sociodemographic strata. Given advantages of this sampling approach, such as pre-established testing infrastructures and capability to perform repeated specimen collection, future studies should investigate the utility of residual blood convenience samples for public health surveillance.[154] Development of methods to estimate pseudo-weights or linkage to supplementary administrative data sources is necessary to better characterize the representativeness of nonprobability samples and correct for sampling biases.

# Chapter 7: References


1. Thacker SB, Berkelman RL. Public health surveillance in the United States. *Epidemiol Rev*. 1988;10(1):164-190. doi:10.1093/oxfordjournals.epirev.a036021

2. Mobula LM, Samaha H, Yao M, et al. Recommendations for the COVID-19 Response at the National Level Based on Lessons Learned from the Ebola Virus Disease Outbreak in the Democratic Republic of the Congo. *Am J Trop Med Hyg*. 2020;103(1):12-17. doi:10.4269/ajtmh.20-0256

3. Schrag SJ, Brooks JT, Beneden CV, et al. SARS Surveillance during Emergency Public Health Response, United States, March–July 2003. *Emerg Infect Dis*. 2004;10(2):185-194. doi:10.3201/eid1002.030752

4. Sullivan PS, Johnson AS, Pembleton ES, et al. Epidemiology of HIV in the USA: epidemic burden, inequities, contexts, and responses. *The Lancet*. 2021;397(10279):1095-1106. doi:10.1016/S0140-6736(21)00395-0

5. Long S, Loutfi D, Kaufman JS, Schuster T. Limitations of Canadian COVID-19 data reporting to the general public. *J Public Health Policy*. 2022;43(2):203-221. doi:10.1057/s41271-022-00337-x

6. Berry I, Brown KA, Buchan SA, et al. A better normal in Canada will need a better detection system for emerging and re-emerging respiratory pathogens. *CMAJ*. 2022;194(36):E1250-E1254. doi:10.1503/cmaj.220577

7. Reedman CN, Drews SJ, Yi QL, Pambrun C, O'Brien SF. Changing Patterns of SARS-CoV-2 Seroprevalence among Canadian Blood Donors during the Vaccine Era. *Microbiol Spectr*. 2022;10(2):e00339-22. doi:10.1128/spectrum.00339-22

8. Groseclose SL, Buckeridge DL. Public Health Surveillance Systems: Recent Advances in Their Use and Evaluation. *Annu Rev Public Health*. 2017;38:57-79. doi:10.1146/annurev-publhealth-031816-044348

9. Kukull WA, Ganguli M. Generalizability. *Neurology*. 2012;78(23):1886-1891. doi:10.1212/WNL.0b013e318258f812

10. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol*. 2013;42(4):1012-1014. doi:10.1093/ije/dys223

11. Murray J, Cohen AL. Infectious Disease Surveillance. *Int Encycl Public Health*. 2017:222-229. doi:10.1016/B978-0-12-803678-5.00517-8

12. Kanji JN, Nguyen LT, Plitt SS, et al. Seropositivity to SARS-CoV-2 in Alberta, Canada in a post-vaccination period (March 2021–July 2021). *Infect Dis*. 2022;54(9):666-676. doi:10.1080/23744235.2022.2080250

13. Charlton CL, Nguyen LT, Bailey A, et al. Pre-Vaccine Positivity of SARS-CoV-2 Antibodies in Alberta, Canada during the First Two Waves of the COVID-19 Pandemic. *Microbiol Spectr*. 2021;9(1):e00291-21. doi:10.1128/Spectrum.00291-21

14. O'Brien SF, Caffrey N, Yi QL, Pambrun C, Drews SJ. SARS-CoV-2 Seroprevalence among Canadian Blood Donors: The Advance of Omicron. *Viruses*. 2022;14(11):2336. doi:10.3390/v14112336

15. Lewin A, De Serres G, Grégoire Y, et al. Seroprevalence of SARS-CoV-2 antibodies among blood donors in Québec: an update from a serial cross-sectional study. *Can J Public Health*. 2022;113(3):385-393. doi:10.17269/s41997-022-00622-y

16. Jones JM, Stone M, Sulaeman H, et al. Estimated US Infection- and Vaccine-Induced SARS-CoV-2 Seroprevalence Based on Blood Donations, July 2020-May 2021. *JAMA*. 2021;326(14):1400-1409. doi:10.1001/jama.2021.15161

17. Assefa N, Regassa LD, Teklemariam Z, et al. Seroprevalence of anti-SARS-CoV-2 antibodies in women attending antenatal care in eastern Ethiopia: a facility-based surveillance. *BMJ Open*. 2021;11(11):e055834. doi:10.1136/bmjopen-2021-055834

18. Franco A, Malhotra N, Simonovits G, Zigerell LJ. Developing Standards for Post-Hoc Weighting in Population-Based Survey Experiments. *J Exp Polit Sci*. 2017;4(2):161-172. doi:10.1017/XPS.2017.2

19. Cornesse C, Blom AG, Dutwin D, et al. A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research. *J Surv Stat Methodol*. 2020;8(1):4-36. doi:10.1093/jssam/smz041

20. Beaumont JF. Are probability surveys bound to disappear for the production of official statistics? *Surv Methodol*. 2020;46(1). http://www.statcan.gc.ca/pub/12-001-x/2020001/article/00001-eng.htm.

21. European Centre for Disease Prevention and Control. *Technical Protocol for Hepatitis C Prevalence Surveys in the General Population: SPHERE C Project.* LU: Publications Office; 2020. https://data.europa.eu/doi/10.2900/30190. Accessed June 19, 2024.

22. Degtiar I, Rose S. A Review of Generalizability and Transportability. *Annu Rev Stat Its Appl*. 2023;10:501-524. doi:10.1146/annurev-statistics-042522-103837

23. Routledge I, Epstein A, Takahashi S, et al. Citywide serosurveillance of the initial SARS-CoV-2 outbreak in San Francisco using electronic health records. *Nat Commun*. 2021;12(1):3566. doi:10.1038/s41467-021-23651-6

24. Powell AA, Ireland G, Leeson R, et al. National and regional prevalence of SARS-CoV-2 antibodies in primary and secondary school children in England: the School Infection Survey, a national open cohort study, November 2021SARS-CoV-2 antibody prevalence in school children. *J Infect*. 2023;86(4):361-368. doi:10.1016/j.jinf.2023.02.016

25. Tang X, Sharma A, Pasic M, et al. Assessment of SARS-CoV-2 Seropositivity During the First and Second Viral Waves in 2020 and 2021 Among Canadian Adults. *JAMA Netw Open*. 2022;5(2):e2146798. doi:10.1001/jamanetworkopen.2021.46798

26. Raina P, Wolfson C, Kirkland S, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *Int J Epidemiol*. 2019;48(6):1752-1753j. doi:10.1093/ije/dyz173

27. Dummer TJB, Awadalla P, Boileau C, et al. The Canadian Partnership for Tomorrow Project: a pan-Canadian platform for research on chronic disease prevention. *Can Med Assoc J*. 2018;190(23):E710-E717. doi:10.1503/cmaj.170292

28. Chen X, Chen Z, Azman AS, et al. Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis. *Lancet Glob Health*. 2021;9(5):e598-e609. doi:10.1016/S2214-109X(21)00026-7

29. Downes M, Carlin JB. Multilevel regression and poststratification as a modeling approach for estimating population quantities in large population health studies: A simulation study. *Biom J*. 2020;62(2):479-491. doi:10.1002/bimj.201900023

30. O'Brien SF, Drews SJ, Lewin A, et al. How do we decide how representative our donors are for public health surveillance? *Transfusion*. 2022;62(12):2431-2437. doi:10.1111/trf.17140

31. Ziegler T, Mamahit A, Cox NJ. 65 years of influenza surveillance by a World Health Organization-coordinated global network. *Influenza Other Respir Viruses*. 2018;12(5):558-565. doi:10.1111/irv.12570

32. Garcia-Abreu A, William H, Danel I. *Public Health Surveillance Toolkit: A Guide for Busy Task Managers*. World Bank; 2002.

33. Tancredi S, Chiolero A. Designing Surveillance at a Population Level. *Am J Public Health*. 2023;113(11):1143-1145. doi:10.2105/AJPH.2023.307425

34. Nsubuga P, White ME, Thacker SB, et al. Public Health Surveillance: A Tool for Targeting and Monitoring Interventions. In: Jamison DT, Breman JG, Measham AR, et al., eds. *Disease Control Priorities in Developing Countries*. 2nd ed. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2006. http://www.ncbi.nlm.nih.gov/books/NBK11770/. Accessed June 5, 2024.

35. Lix L, Puchtinger R, Reimer K, et al. The Canadian Chronic Disease Surveillance System: A model for collaborative surveillance. *Int J Popul Data Sci*. 2018;3(3). doi:10.23889/ijpds.v3i3.433

36. Oloye FF, Xie Y, Asadi M, et al. Rapid transition between SARS-CoV-2 variants of concern Delta and Omicron detected by monitoring municipal wastewater from three Canadian cities. *Sci Total Environ*. 2022;841:156741. doi:10.1016/j.scitotenv.2022.156741

37. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *J Travel Med*. 2020;27(2):taaa008. doi:10.1093/jtm/taaa008

38. Marques-Toledo C de A, Degener CM, Vinhal L, et al. Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level. *PLoS Negl Trop Dis*. 2017;11(7):e0005729. doi:10.1371/journal.pntd.0005729

39. Boulanger V, Poirier É, MacLaurin A, Quach C. Divergences between healthcare-associated infection administrative data and active surveillance data in Canada. *Can Commun Dis Rep*. 2022;48(1):4-16. doi:10.14745/ccdr.v48i01a02

40. Scharff RL, Besser J, Sharp DJ, Jones TF, Peter GS, Hedberg CW. An Economic Evaluation of PulseNet: A Network for Foodborne Disease Surveillance. *Am J Prev Med*. 2016;50(5):S66-S73. doi:10.1016/j.amepre.2015.09.018

41. Belzak L, Halverson J. Evidence synthesis - The opioid crisis in Canada: a national perspective. *Health Promot Chronic Dis Prev Can Res Policy Pract*. 2018;38(6):224-233.

42. Crain J, McFaull S, Thompson W, et al. The Canadian Hospitals Injury Reporting and Prevention Program: a dynamic and innovative injury surveillance system. *Health Promot Chronic Dis Prev Can Res Policy Pract*. 2016;36(6):112-117.

43. Rudolph JE, Zhong Y, Duggal P, Mehta SH, Lau B. Defining representativeness of study samples in medical and population health research. *BMJ Med*. 2023;2(1):e000399. doi:10.1136/bmjmed-2022-000399

44. Griffith GJ, Morris TT, Tudball MJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun*. 2020;11(1):5749. doi:10.1038/s41467-020-19478-2

45. Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. *PLoS ONE*. 2015;10(9):e0137203. doi:10.1371/journal.pone.0137203

46. Arriaga MB, Amorim G, Queiroz ATL, et al. Novel stepwise approach to assess representativeness of a large multicenter observational cohort of tuberculosis patients: The example of RePORT Brazil. *Int J Infect Dis*. 2021;103:110-118. doi:10.1016/j.ijid.2020.11.140

47. Yee MH, Kravitz RL. Racial and Ethnic Representativeness of California Medical Schools, 2021. *JAMA Netw Open*. 2023;6(1):e2251192. doi:10.1001/jamanetworkopen.2022.51192

48. Weng C, Li Y, Ryan P, et al. A Distribution-based Method for Assessing The Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records. *Appl Clin Inform*. 2017;05:463-479. doi:10.4338/ACI-2013-12-RA-0105

49. He Z, Ryan P, Hoxha J, et al. Multivariate analysis of the population representativeness of related clinical studies. *J Biomed Inform*. 2016;60:66-76. doi:10.1016/j.jbi.2016.01.007

50. Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol*. 2022;23(1):3-20. doi:10.1038/s41580-021-00418-x

51. World Health Organization. *Transmission of SARS-CoV-2: Implications for Infection Prevention Precautions*.; 2020. https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions. Accessed May 23, 2024.

52. Ghinai I, McPherson TD, Hunter JC, et al. First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. *The Lancet*. 2020;395(10230):1137-1144. doi:10.1016/S0140-6736(20)30607-3

53. Li Qun, Guan Xuhua, Wu Peng, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med*. 2020;382(13):1199-1207. doi:10.1056/NEJMoa2001316

54. Shang J, Wan Y, Luo C, et al. Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci*. 2020;117(21):11727-11734. doi:10.1073/pnas.2003138117

55. Alimohamadi Y, Sepandi M, Taghdir M, Hosamirudsari H. Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis. *J Prev Med Hyg*. 2020;61(3):E304-E304. doi:10.15167/2421-4248/jpmh2020.61.3.1530

56. Berlin David A., Gulick Roy M., Martinez Fernando J. Severe Covid-19. *N Engl J Med*. 2020;383(25):2451-2460. doi:10.1056/NEJMcp2009575

57. Velásquez García HA, Adu PA, Harrigan S, et al. Risk factors for COVID-19 hospitalization after COVID-19 vaccination: a population-based cohort study in Canada. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis*. 2023;127:116-123. doi:10.1016/j.ijid.2022.12.001

58. Turtle L, Thorpe M, Drake TM, et al. Outcome of COVID-19 in hospitalised immunocompromised patients: An analysis of the WHO ISARIC CCP-UK prospective cohort study. *PLOS Med*. 2023;20(1):e1004086. doi:10.1371/journal.pmed.1004086

59. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. *PLOS Med*. 2021;18(9):e1003773. doi:10.1371/journal.pmed.1003773

60. He Z, Ren L, Yang J, et al. Seroprevalence and humoral immune durability of anti-SARS-CoV-2 antibodies in Wuhan, China: a longitudinal, population-level, cross-sectional study. *The Lancet*. 2021;397(10279):1075-1084. doi:10.1016/S0140-6736(21)00238-5

61. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet*. 2020;396(10250):535-544. doi:10.1016/S0140-6736(20)31483-5

62. Menachemi N, Yiannoutsos, Constantin T., Dixon, Brian E., et al. Population Point Prevalence of SARS-CoV-2 Infection Based on a Statewide Random Sample — Indiana, April 25–29, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:960-964. doi:10.15585/mmwr.mm6929e1

63. Murphy TJ, Swail H, Jain J, et al. The evolution of SARS-CoV-2 seroprevalence in Canada: a time-series study, 2020–2023. *CMAJ*. 2023;195(31):E1030-E1037. doi:10.1503/cmaj.230249

64. Murhekar MV, Bhatnagar T, Selvaraju S, et al. Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May-June 2020. *Indian J Med Res*. 2020;152(1-2):48. doi:10.4103/ijmr.IJMR_3290_20

65. Ward H, Atchison C, Whitaker M, et al. SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nat Commun*. 2021;12(1):905. doi:10.1038/s41467-021-21237-w

66. Stringhini S, Wisniak A, Piumatti G, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *The Lancet*. 2020;396(10247):313-319. doi:10.1016/S0140-6736(20)31304-0

67. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*. 2020;395(10228):931-934. doi:10.1016/S0140-6736(20)30567-5

68. Cascini F, Failla G, Gobbi C, et al. A cross-country comparison of Covid-19 containment measures and their effects on the epidemic curves. *BMC Public Health*. 2022;22:1765. doi:10.1186/s12889-022-14088-7

69. Talic S, Shah S, Wild H, et al. Effectiveness of public health measures in reducing the incidence of covid-19, SARS-CoV-2 transmission, and covid-19 mortality: systematic review and meta-analysis. *The BMJ*. 2021;375:e068302. doi:10.1136/bmj-2021-068302

70. Atkinson A, Albert A, McClymont E, et al. Canadian SARS-CoV-2 serological survey using antenatal serum samples: a retrospective seroprevalence study. *Can Med Assoc Open Access J*. 2023;11(2):E305-E313. doi:10.9778/cmajo.20220045

71. Liu L, Iketani S, Guo Y, et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature*. 2022;602(7898):676-681. doi:10.1038/s41586-021-04388-0

72. World Health Organization. *COVID-19 Epidemiological Update – 19 January 2024*. World Health Organization; 2024. https://www.who.int/publications/m/item/covid-19-epidemiological-update---19-january-2024. Accessed May 25, 2024.

73. Price-Haywood Eboni G., Burton Jeffrey, Fort Daniel, Seoane Leonardo. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. *N Engl J Med*. 2020;382(26):2534-2543. doi:10.1056/NEJMsa2011686

74. Dodd RY, Xu M, Stramer SL. Change in Donor Characteristics and Antibodies to SARS-CoV-2 in Donated Blood in the US, June-August 2020. *JAMA*. 2020;324(16):1677-1679. doi:10.1001/jama.2020.18598

75. Hoebel J, Grabka MM, Schröder C, et al. Socioeconomic position and SARS-CoV-2 infections: seroepidemiological findings from a German nationwide dynamic cohort. *J Epidemiol Community Health*. 2022;76(4):350-353. doi:10.1136/jech-2021-217653

76. Sundaram ME, Calzavara A, Mishra S, et al. Individual and social determinants of SARS-CoV-2 testing and positivity in Ontario, Canada: a population-wide study. *CMAJ*. 2021;193(20):E723-E734. doi:10.1503/cmaj.202608

77. Starke KR, Reissig D, Petereit-Haack G, Schmauder S, Nienhaus A, Seidler A. The isolated effect of age on the risk of COVID-19 severe outcomes: a systematic review with meta-analysis. *BMJ Glob Health*. 2021;6(12):e006434. doi:10.1136/bmjgh-2021-006434

78. Brockman MA, Mwimanzi F, Lapointe HR, et al. Reduced Magnitude and Durability of Humoral Immune Responses to COVID-19 mRNA Vaccines Among Older Adults. *J Infect Dis*. 2022;225(7):1129-1140. doi:10.1093/infdis/jiab592

79. Kwon S, Joshi AD, Lo CH, et al. Association of social distancing and face mask use with risk of COVID-19. *Nat Commun*. 2021;12(1):3737. doi:10.1038/s41467-021-24115-7

80. Orlowski EJW, Goldsmith DJA. Four months into the COVID-19 pandemic, Sweden's prized herd immunity is nowhere in sight. *J R Soc Med*. 2020;113(8):292-298. doi:10.1177/0141076820945282

81. Moline HL. Effectiveness of COVID-19 Vaccines in Preventing Hospitalization Among Adults Aged ≥65 Years — COVID-NET, 13 States, February–April 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70. doi:10.15585/mmwr.mm7032e3

82. Polack Fernando P., Thomas Stephen J., Kitchin Nicholas, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med*. 2020;383(27):2603-2615. doi:10.1056/NEJMoa2034577

83. Bar-On Yinon M., Goldberg Yair, Mandel Micha, et al. Protection of BNT162b2 Vaccine Booster against Covid-19 in Israel. *N Engl J Med*. 2021;385(15):1393-1400. doi:10.1056/NEJMoa2114255

84. Havers FP, Reed C, Lim T, et al. Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. *JAMA Intern Med*. 2020;180(12):1576-1586. doi:10.1001/jamainternmed.2020.4130

85. Ibrahim NK. Epidemiologic surveillance for controlling Covid-19 pandemic: types, challenges and implications. *J Infect Public Health*. 2020;13(11):1630-1638. doi:10.1016/j.jiph.2020.07.019

86. Manuel DG, Delatolla R, Fisman DN, et al. *The Role of Wastewater Testing for SARS-CoV-2 Surveillance*. Ontario COVID-19 Science Advisory Table; 2021. doi:10.47326/ocsat.2021.02.40.1.0

87. Dougherty BP, Smith BA, Carson CA, Ogden NH. Exploring the percentage of COVID-19 cases reported in the community in Canada and associated case fatality ratios. *Infect Dis Model*. 2021;6:123-132. doi:10.1016/j.idm.2020.11.008

88. Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environ Sci Technol Lett*. 2020;7(7):511-516. doi:10.1021/acs.estlett.0c00357

89. Pang X, Gao T, Ellehoj E, et al. Wastewater-Based Surveillance Is an Effective Tool for Trending COVID-19 Prevalence in Communities: A Study of 10 Major Communities for 17 Months in Alberta. *ACS EST Water*. 2022;2(11):2243-2254. doi:10.1021/acsestwater.2c00143

90. Hosseini A, Pandey R, Osman E, et al. Roadmap to the Bioanalytical Testing of COVID-19: From Sample Collection to Disease Surveillance. *ACS Sens*. 2020;5(11):3328-3345. doi:10.1021/acssensors.0c01377

91. Platten M, Hoffmann D, Grosser R, et al. SARS-CoV-2, CT-Values, and Infectivity—Conclusions to Be Drawn from Side Observations. *Viruses*. 2021;13(8):1459. doi:10.3390/v13081459

92. Maya S, Kahn JG. COVID-19 testing protocols to guide duration of isolation: a cost-effectiveness analysis. *BMC Public Health*. 2023;23(1):864. doi:10.1186/s12889-023-15762-0

93. Polechová J, Johnson KD, Payne P, et al. SARS-CoV-2 rapid antigen tests provide benefits for epidemic control – observations from Austrian schools. *J Clin Epidemiol*. 2022;145:14-19. doi:10.1016/j.jclinepi.2022.01.002

94. Pray IW. Performance of an Antigen-Based Test for Asymptomatic and Symptomatic SARS-CoV-2 Testing at Two University Campuses — Wisconsin, September–October 2020. *MMWR Morb Mortal Wkly Rep*. 2021;69. doi:10.15585/mmwr.mm695152a3

95. Vandenberg O, Martiny D, Rochas O, van Belkum A, Kozlakidis Z. Considerations for diagnostic COVID-19 tests. *Nat Rev Microbiol*. 2021;19(3):171-183. doi:10.1038/s41579-020-00461-z

96. Patel EU, Bloch EM, Clarke W, et al. Comparative Performance of Five Commercially Available Serologic Assays To Detect Antibodies to SARS-CoV-2 and Identify Individuals

with High Neutralizing Titers. *J Clin Microbiol*. 2021;59(2):10.1128/jcm.02257-20. doi:10.1128/jcm.02257-20

97. Moreira VM, Mascarenhas P, Machado V, et al. Diagnosis of SARS-Cov-2 Infection by RT-PCR Using Specimens Other Than Naso- and Oropharyngeal Swabs: A Systematic Review and Meta-Analysis. *Diagnostics*. 2021;11(2):363. doi:10.3390/diagnostics11020363

98. Guthmiller JJ, Stovicek O, Wang J, et al. SARS-CoV-2 Infection Severity Is Linked to Superior Humoral Immunity against the Spike. *mBio*. 2021;12(1):10.1128/mbio.02940-20. doi:10.1128/mbio.02940-20

99. Ward H, Atchison C, Whitaker M, et al. Design and Implementation of a National Program to Monitor the Prevalence of SARS-CoV-2 IgG Antibodies in England Using Self-Testing: The REACT-2 Study. *Am J Public Health*. 2023;113(11):1201-1209. doi:10.2105/AJPH.2023.307381

100. Yu Y, Knight MJ, Gibson D, O'Brien SF, Buckeridge DL, Russell WA. Temporal trends in disparities in COVID-19 seropositivity among Canadian blood donors. *Int J Epidemiol*. 2024;53(3):dyae078. doi:10.1093/ije/dyae078

101. Haq MA, Roy AK, Ahmed R, et al. Antibody longevity and waning following COVID-19 vaccination in a 1-year longitudinal cohort in Bangladesh. *Sci Rep*. 2024;14(1):11467. doi:10.1038/s41598-024-61922-6

102. Seow J, Graham C, Merrick B, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nat Microbiol*. 2020;5(12):1598-1607. doi:10.1038/s41564-020-00813-8

103. Abu-Raddad LJ, Chemaitelly H, Ayoub HH, et al. Characterizing the Qatar advanced-phase SARS-CoV-2 epidemic. *Sci Rep*. 2021;11(1):6233. doi:10.1038/s41598-021-85428-7

104. Saeed S, Drews SJ, Pambrun C, Yi QL, Osmond L, O'Brien SF. SARS-CoV-2 seroprevalence among blood donors after the first COVID-19 wave in Canada. *Transfusion*. 2021;61(3):862-872. doi:10.1111/trf.16296

105. Price R, Cho J, Nelson S. SARS-CoV-2 Seroprevalence at an Urban Hospital in Haiti. *Cureus*. 2022;14(8):e27690. doi:10.7759/cureus.27690

106. Shook-Sa BE, Boyce RM, Aiello AE. Estimation Without Representation: Early Severe Acute Respiratory Syndrome Coronavirus 2 Seroprevalence Studies and the Path Forward. *J Infect Dis*. 2020;222(7):1086-1089. doi:10.1093/infdis/jiaa429

107. Clapham H, Hay J, Routledge I, et al. Seroepidemiologic Study Designs for Determining SARS-COV-2 Transmission and Immunity. *Emerg Infect Dis*. 2020;26(9):1978-1986. doi:10.3201/eid2609.201840

108.Lindan CP, Desai M, Boothroyd D, et al. Design of a population-based longitudinal cohort study of SARS-CoV-2 incidence and prevalence among adults in the San Francisco Bay Area. *Ann Epidemiol*. 2022;67:81-100. doi:10.1016/j.annepidem.2021.11.001

109.Statistics Canada. Canadian COVID-19 Antibody and Health Survey (CCAHS). https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1287991. Published October 13, 2020. Accessed May 20, 2024.

110.Pastor-Barriuso R, Pérez-Gómez B, Oteo-Iglesias J, et al. Design and Implementation of a Nationwide Population-Based Longitudinal Survey of SARS-CoV-2 Infection in Spain: The ENE-COVID Study. *Am J Public Health*. 2023;113(5):525-532. doi:10.2105/AJPH.2022.307167

111.Wu DC, Jha P, Lam T, et al. Predictors of self-reported symptoms and testing for COVID-19 in Canada using a nationally representative survey. *PLoS ONE*. 2020;15(10):e0240778. doi:10.1371/journal.pone.0240778

112.Caruana EJ, Roman M, Hernández-Sánchez J, Solli P. Longitudinal studies. *J Thorac Dis*. 2015;7(11):E537-E540. doi:10.3978/j.issn.2072-1439.2015.10.63

113.International Vaccine Access Center, Johns Hopkins University. Serosurveys using Residual Specimens from a Biorepository or Health Facility. Serosurvey Tools. https://serosurveytools.org/residual-specimens/. Accessed June 19, 2024.

114.Meyers E, Heytens S, Formukong A, et al. Comparison of Dried Blood Spots and Venous Blood for the Detection of SARS-CoV-2 Antibodies in a Population of Nursing Home Residents. *Microbiol Spectr*. 9(2):e00178-21. doi:10.1128/Spectrum.00178-21

115.Atchison C, Pristerà P, Cooper E, et al. Usability and Acceptability of Home-based Self-testing for Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Antibodies for Population Surveillance. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2020;72(9):e384-e393. doi:10.1093/cid/ciaa1178

116.Carcelen AC, Limaye RJ, Mutembo S, et al. Acceptability of serosurveys in southern Zambia: data collector and caregiver perspectives. *Discov Soc Sci Health*. 2023;3(1):3. doi:10.1007/s44155-023-00032-6

117.Vehovar V, Toepoel V, Steinmetz S. The SAGE Handbook of Survey Methodology. In: *The SAGE Handbook of Survey Methodology*. 55 City Road: SAGE Publications Ltd; 2016:329-345. doi:10.4135/9781473957893

118.Wu C. Statistical inference with non-probability survey samples. *Surv Methodol*. 2022;8(2). http://www.statcan.gc.ca/pub/12-001-x/2022002/article/00002-eng.htm.

119.Patel EU, Bloch EM, Tobian AAR. Severe Acute Respiratory Syndrome Coronavirus 2 Serosurveillance in Blood Donor Populations. *J Infect Dis*. 2022;225(1):1-4. doi:10.1093/infdis/jiab517

120. Atsma F, Veldhuizen I, Verbeek A, de Kort W, de Vegt F. Healthy donor effect: its magnitude in health research among blood donors. *Transfusion*. 2011;51(8):1820-1828. doi:10.1111/j.1537-2995.2010.03055.x

121. Froom P, Melamed S, Kristal-Boneh E, Benbassat J, Ribak J. Healthy Volunteer Effect in Industrial Workers. *J Clin Epidemiol*. 1999;52(8):731-735. doi:10.1016/S0895-4356(99)00070-0

122. Taylor AE, Jones HJ, Sallis H, et al. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2018;47(4):1207-1216. doi:10.1093/ije/dyy060

123. Kolenikov S. Calibrating Survey Data using Iterative Proportional Fitting (Raking). *Stata J*. 2014;14(1):22-59. doi:10.1177/1536867X1401400104

124. Khandelwal, Aditi. Impact of COVID-19 on blood donation in Canada. Canadian Blood Services. https://professionaleducation.blood.ca/en/transfusion/publications/impact-covid-19-blood-donation-canada. Published 2021. Accessed July 3, 2024.

125. Kirsh VA, Skead K, McDonald K, et al. Cohort Profile: The Ontario Health Study (OHS). *Int J Epidemiol*. 2023;52(2):e137-e151. doi:10.1093/ije/dyac156

126. Dhalla A, McDonald TE, Gallagher RP, et al. Cohort Profile: The British Columbia Generations Project (BCGP). *Int J Epidemiol*. 2019;48(2):377-378k. doi:10.1093/ije/dyy160

127. Sweeney E, Cui Y, DeClercq V, et al. Cohort Profile: The Atlantic Partnership for Tomorrow's Health (Atlantic PATH) Study. *Int J Epidemiol*. 2017;46(6):1762-1763i. doi:10.1093/ije/dyx124

128. Ye M, Robson PJ, Eurich DT, Vena JE, Xu JY, Johnson JA. Cohort Profile: Alberta's Tomorrow Project. *Int J Epidemiol*. 2017;46(4):1097-1098l. doi:10.1093/ije/dyw256

129. Awadalla P, Boileau C, Payette Y, et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *Int J Epidemiol*. 2013;42(5):1285-1299. doi:10.1093/ije/dys160

130. Canadian Partnership for Tomorrow's Health. COVID-19 Initiative. https://canpath.ca/wp-content/uploads/2020/12/CanPath-CPHA-Webinar-Presentation-FINAL.pdf. Accessed June 1, 2024.

131. Canadian Longitudinal Study on Aging. Data Support Document SARS-CoV-2 Antibodies. April 2024.

132. Statistics Canada. Canadian Community Health Survey - Healthy Aging (CCHS). https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5146. Published March 26, 2008. Accessed June 6, 2024.

133.Raina PS, Wolfson C, Kirkland SA, et al. The Canadian Longitudinal Study on Aging (CLSA). *Can J Aging Rev Can Vieil*. 2009;28(3):221-229. doi:10.1017/S0714980809990055

134.Statistics Canada. COVID-19 infection in the Canadian household population. https://www150.statcan.gc.ca/n1/pub/82-003-x/2022004/article/00003-eng.htm. Published April 20, 2022. Accessed October 30, 2022.

135.Statistics Canada. *Sampling and Weighting Technical Report*.; 2018. https://www12.statcan.gc.ca/census-recensement/2016/ref/98-306/98-306-x2016001-eng.pdf.

136.Pampalon R, Hamel D, Gamache P, Philibert MD, Raymond G, Simpson A. An area-based material and social deprivation index for public health in Québec and Canada. *Can J Public Health Rev Can Sante Publique*. 2012;103(8 Suppl 2):S17-22. doi:10.1007/BF03403824

137.Pampalon R, Hamel D, Gamache P, Raymond G. A deprivation index for health planning in Canada. *Chronic Dis Can*. 2009;29(4):178-191. doi:10.24095/hpcdp.29.4.05

138.Statistics Canada. Census Dictionary, 2016. https://www12.statcan.gc.ca/census-recensement/2016/ref/dict/index-eng.cfm. Published November 16, 2016. Accessed June 7, 2024.

139.Lee KJ, Tilling KM, Cornish RP, et al. Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework. *J Clin Epidemiol*. 2021;134:79-88. doi:10.1016/j.jclinepi.2021.01.008

140.Efron B. Bootstrap Methods: Another Look at the Jackknife. *Ann Stat*. 1979;7(1):1-26. doi:10.1214/aos/1176344552

141.R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2023. https://www.R-project.org/.

142.Bonevski B, Randell M, Paul C, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Med Res Methodol*. 2014;14(1):42. doi:10.1186/1471-2288-14-42

143.Etti M, Fofie H, Razai M, Crawshaw AF, Hargreaves S, Goldsmith LP. Ethnic minority and migrant underrepresentation in Covid-19 research: Causes and solutions. *eClinicalMedicine*. 2021;36. doi:10.1016/j.eclinm.2021.100903

144.Globerman S. *Aging and Expenditures on Health Care*. Fraser Institute; 2021:1-12.

145.Andrade C. The Inconvenient Truth About Convenience and Purposive Samples. *Indian J Psychol Med*. 2021;43(1):86-88. doi:10.1177/0253717620977000

146.Lewin A, Therrien R, De Serres G, et al. SARS-CoV-2 seroprevalence among blood donors in Québec, and analysis of symptoms associated with seropositivity: a nested case-control study. *Can J Public Health*. 2021;112(4):576-586. doi:10.17269/s41997-021-00531-6

147.Oliveira J. e Silva L, Gerberi DJ, Cummins NW, Shah AS, Bellolio F. Representativeness of Racial and Ethnic Groups in COVID-19 Outpatient Trials in the United States. *Mayo Clin Proc*. 2022;97(1):184-186. doi:10.1016/j.mayocp.2021.10.016

148.Khan MM, Kobayashi K, Vang ZM, Lee SM. Are visible minorities "invisible" in Canadian health data and research? A scoping review. *Int J Migr Health Soc Care*. 2017;13(1):126-143. doi:10.1108/IJMHSC-10-2015-0036

149.NIH Policy and Guidelines on The Inclusion of Women and Minorities as Subjects in Clinical Research | grants.nih.gov. https://grants.nih.gov/policy/inclusion/women-and-minorities/guidelines.htm. Accessed June 14, 2024.

150.Government of the United Kingdom. Equality Act 2010. https://www.legislation.gov.uk/ukpga/2010/15/contents. Accessed June 14, 2024.

151.Government of Canada. Guide to the Census of Population, 2016. 2017;(98):1-77.

152.Cohen MZ, Thompson CB, Yates B, Zimmerman L, Pullen CH. Implementing common data elements across studies to advance research. *Nurs Outlook*. 2015;63(2):181-188. doi:10.1016/j.outlook.2014.11.006

153.Gleason JL, Tamburro R, Signore C. Promoting Data Harmonization of COVID-19 Research in Pregnant and Pediatric Populations. *JAMA*. 2023;330(6):497-498. doi:10.1001/jama.2023.10835

154.O'Brien SF, Goldman M, Drews SJ. An expanded role for blood donor emerging pathogens surveillance. *CMAJ*. 2023;195(1):E16-E16. doi:10.1503/cmaj.147635-l