WEBINAR:

Who gets counted? Examining representativeness in Canadian COVID-19 serosurveillance studies

Presented by:

Dr. W. Alton Russell
Matthew Knight
Dr. Yuan Yu

Wednesday, September 10

1:00 pm EDT

Agenda

- **Background**
- Representativeness of SARS-Cov-2 serosurveillance studies
- Adjusting for representativeness
- Future directions

Typical objectives of serosurveillance

Measure antibodies in the population to:

- Estimate population-level exposure to infection and/or vaccination
- Understand transmission dynamics
- Identify immune gaps
- Inform public health interventions
  - Vaccination campaigns, Social distancing

Requires that the population sampled **resembles the target population** (at least after adjustments are made)

Source:
https://www.covid19immunitytaskforce.ca/

Dimensions of representativeness

- Age
- Sex, gender, sexual orientation
- Health, comorbidity
- Race, racialization
- Mobility, disability
- Lifestyle, preventative health
- Neighborhood
- Income, material deprivation
- Occupation
- Risk behavior
- Geography
- Urban/rural
- First Nations reserves, military bases, carceral settings
- …

Source: pmlive.com

Adjusting for representativeness

Statistical adjustment can improve generalizability for subgroups if

- Group members in study population
- Characteristic measured
- Distribution for target population available

Source: pmlive.com

- Age
- Sex, gender, sexual orientation
- Health, comorbidity
- Race, racialization
- Mobility, disability
- Lifestyle, preventative health
- Neighborhood
- Income, material deprivation
- Occupation
- Attitudes, risk behavior
- Geography
- Urban/rural
- First Nations reserves, military bases, carceral settings
- …

Match population to pathogen & policy question

CLINIQUE SIDEP+ MONTÉRÉGIE-CENTRE

CanPath

Match cohort to pathogen & policy question

What makes a study population less representative?

- **Selection bias:** Who was invited or sampled?
  - Are geographic, demographic or clinical groups absent or underrepresented?
  - Can be mitigated using probabilistic sampling frames
- **Non-response bias:** Who responds to an invitation?
  - Non-response to surveys has **increased** in recent decades
  - Public health surveillance using residual blood from routinely collected samples can sometimes be done without consent, **eliminating non-response bias**
    - But, **selection bias** may be high, and contextual information is often low

Agenda

- Background
- **Representativeness of SARS-Cov-2 serosurveillance studies**
- Adjusting for representativeness
- Future directions

Knight *et al. BMC Public Health* (2025) 25:2057
https://doi.org/10.1186/s12889-025-22975-y

BMC Public Health

RESEARCH Open Access

Sociodemographic characteristics of SARS-CoV-2 serosurveillance studies with diverse recruitment strategies, Canada, 2020 to 2023

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] and W. Alton Russell[1,4]

Project overview

- Serological testing composed a significant portion of Canada's SARS-CoV-2 surveillance strategy
- Multiple study designs, including convenience samples, probabilistic surveys, and de novo cohorts, were used to gain insights into population immunity and disease dynamics
- Study aim: assess the demographic representativeness of six serosurveillance studies to the general Canadian population

Methodology

- Analyzed demographic data of specimens collected by 6 serosurveillance studies
  - De novo cohort (Canadian COVID-19 Antibody and Health Survey [CCAHS-1])
  - Convenience samples (Canadian Blood Services [CBS], Alberta Precision Laboratories [APL])
  - Probabilistic surveys (Action to Beat Coronavirus (Ab-C), Canadian Longitudinal Study on Aging COVID-19 Study (CLSA), Canadian Partnership for Tomorrow's Health COVID-19 Study (CanPath)

- General population data collected from 2016 Canadian Census
  - Variables: age, sex, self-reported race/ethnicity, urban/rural residence, social and material quintile of deprivation

Methodology

- Analysis restricted to participants who provided age, province/territory of residence, and serology sample
  - Required to calibrate representation assessment

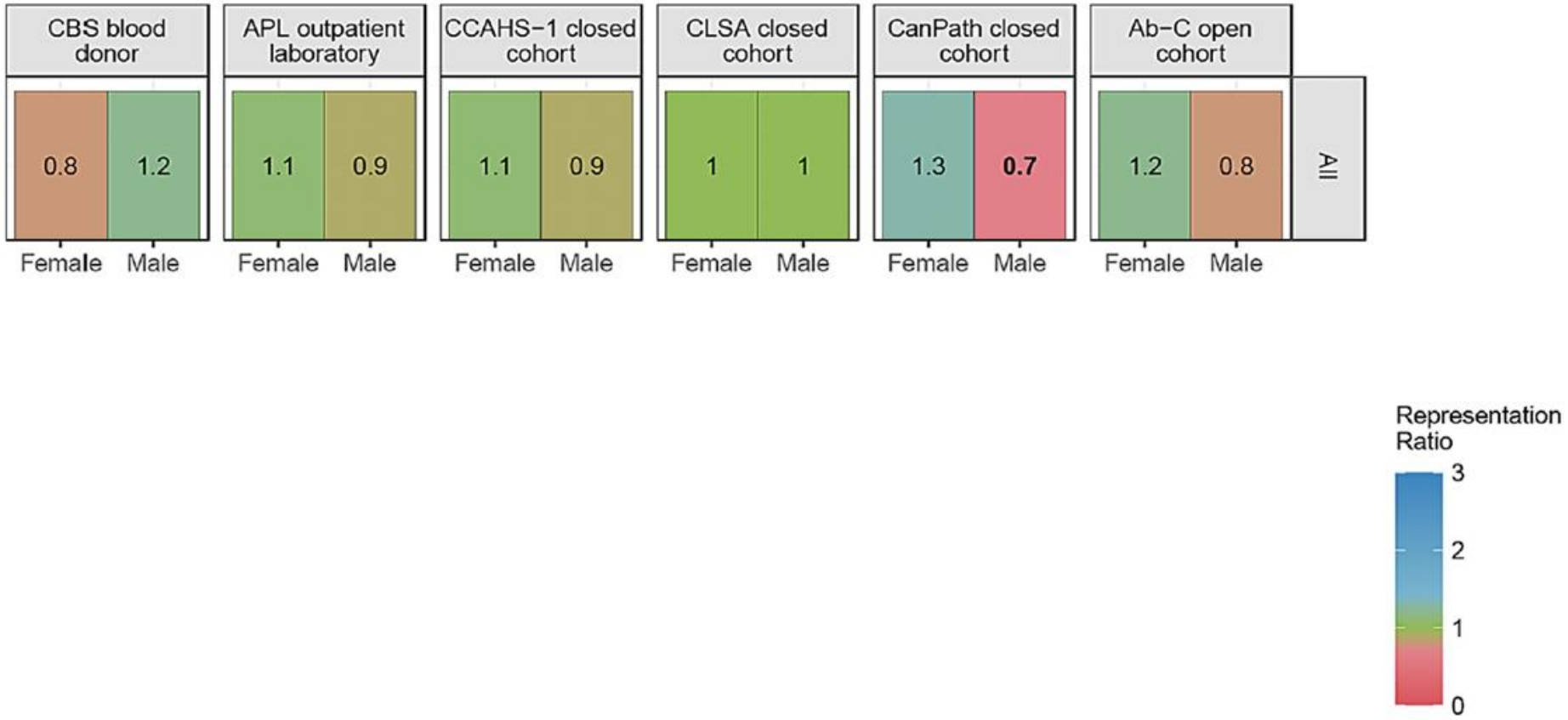
- Representativeness assessed using a representation ratio
  - Representation ratio < 1 indicates subgroup underrepresented compared to census population, while a ratio > 1 indicates overrepresentation

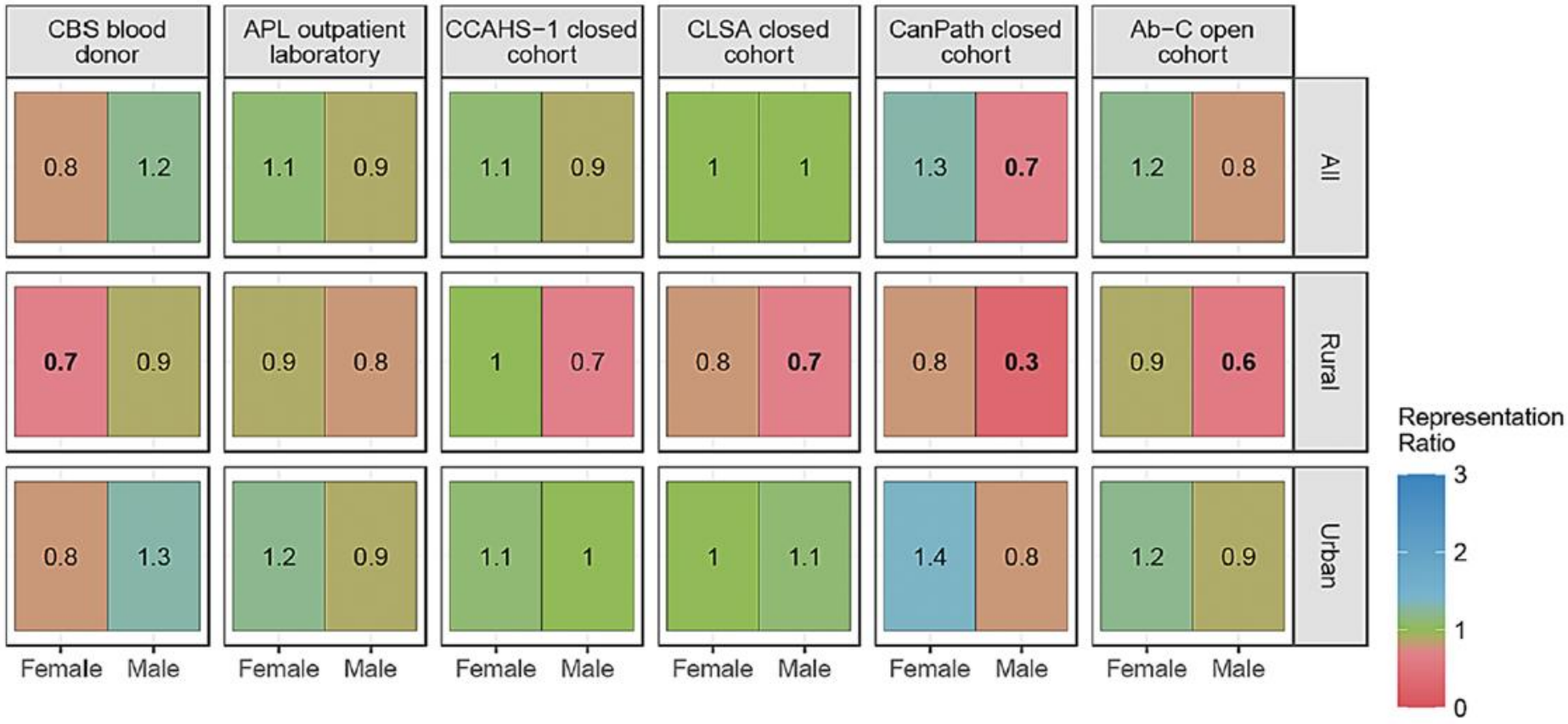
- Bootstrapping used to identify significantly underrepresented subgroups
  - Subgroup significantly underrepresented if > 95% of replicates produced representation ratio < 0.75

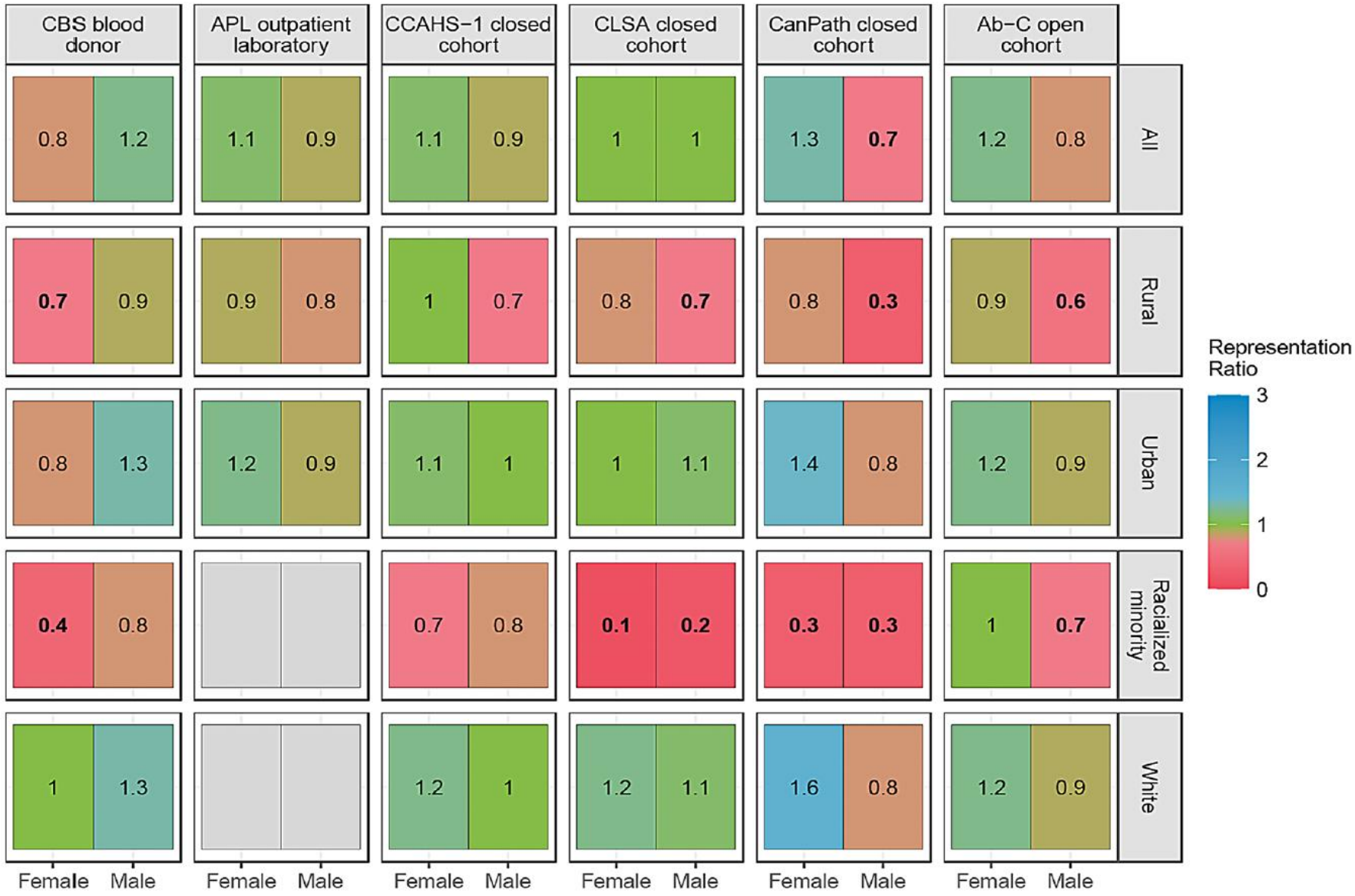
$$Representation\ ratio = \frac{proportion\ of\ study\ specimens\ in\ strata}{proportion\ of\ population\ in\ strata}$$

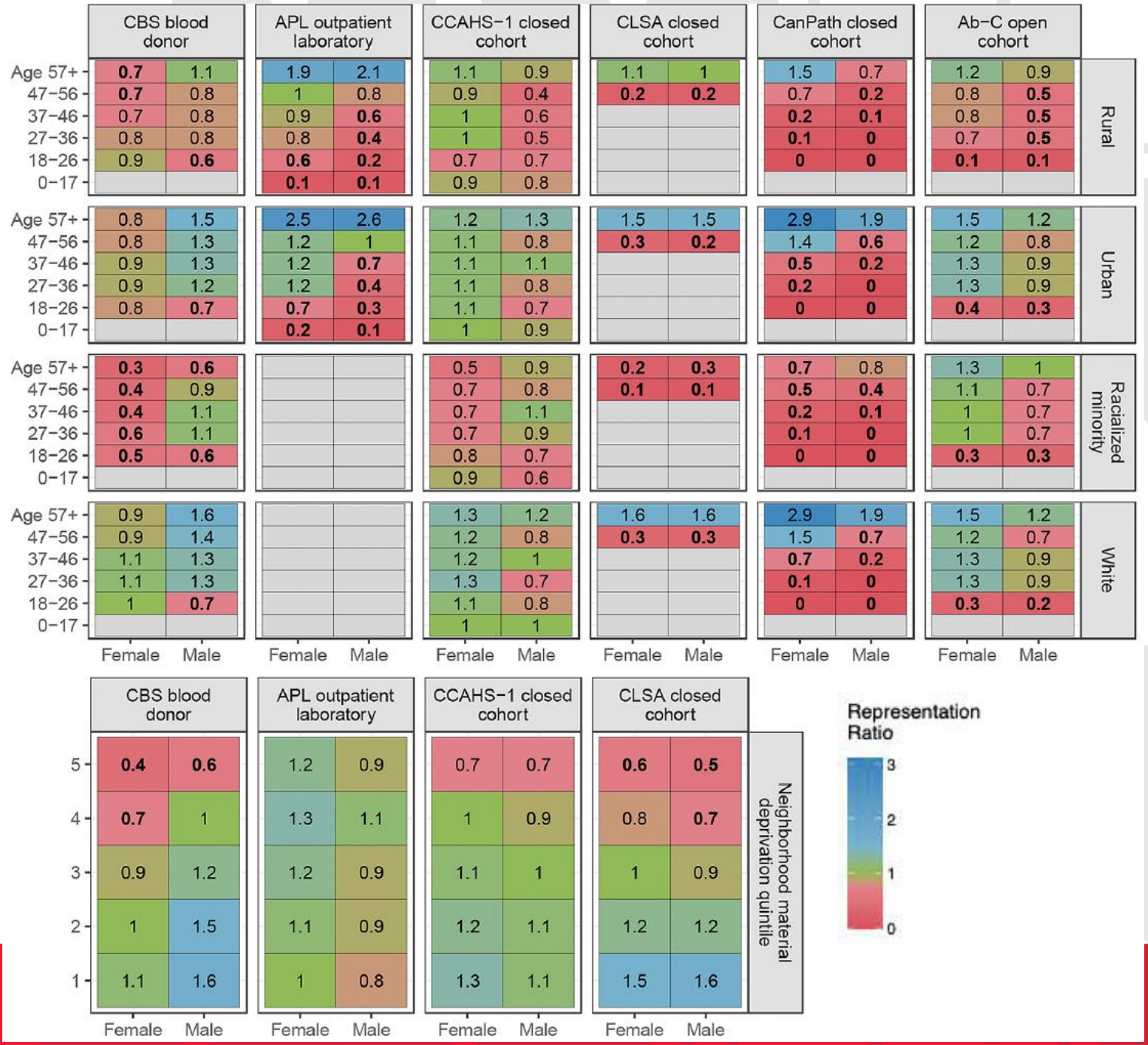
Comparing six serosurveillance studies

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	27,140 specimens from 10,621 participants May 2020 - April 2022	No
Alberta Precision Laboratories (APL)	Serial cross-sectional convenience sample	≥ 0	AB	Heparinized plasma Plasma Serum	210,906 specimens from 187,888 participants 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 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 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 Plasma	17,310 specimens from 17,310 participants 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	25,156 specimens from 25,153 participants September 2020 - November 2021	No


CBS blood donor
APL outpatient laboratory
CCAHS−1 closed cohort
CLSA closed cohort
CanPath closed cohort
Ab−C open cohort
All
0.8
1.2
1.1
0.9
1.1
0.9
1
1
1.3
0.7
1.2
0.8
Female
Male
Representation Ratio
3
2
1
0


CBS blood donor
APL outpatient laboratory
CCAHS−1 closed cohort
CLSA closed cohort
CanPath closed cohort
Ab−C open cohort
All
Rural
Urban
0.8 1.2
1.1 0.9
1.1 0.9
1 1
1.3 0.7
1.2 0.8
0.7 0.9
0.9 0.8
1 0.7
0.8 0.7
0.8 0.3
0.9 0.6
0.8 1.3
1.2 0.9
1.1 1
1 1.1
1.4 0.8
1.2 0.9
Female Male
Representation Ratio
3
2
1
0


CBS blood donor
APL outpatient laboratory
CCAHS−1 closed cohort
CLSA closed cohort
CanPath closed cohort
Ab−C open cohort
All
Rural
Urban
Racialized minority
White
0.8 1.2
1.1 0.9
1.1 0.9
1 1
1.3 0.7
1.2 0.8
0.7 0.9
0.9 0.8
1 0.7
0.8 0.7
0.8 0.3
0.9 0.6
0.8 1.3
1.2 0.9
1.1 1
1 1.1
1.4 0.8
1.2 0.9
0.4 0.8
0.7 0.8
0.1 0.2
0.3 0.3
1 0.7
1 1.3
1.2 1
1.2 1.1
1.6 0.8
1.2 0.9
Female Male
Representation Ratio
3
2
1
0


CBS blood donor
APL outpatient laboratory
CCAHS−1 closed cohort
CLSA closed cohort
CanPath closed cohort
Ab−C open cohort
Age 57+
47−56
37−46
27−36
18−26
0−17
Rural
Urban
Racialized minority
White
Female
Male
Neighborhood material deprivation quintile
5
4
3
2
1
Representation Ratio
3
2
1
0


CBS blood donor
APL outpatient laboratory
CCAHS–1 closed cohort
CLSA closed cohort
Social deprivation quintile
5
4
3
2
1
0.7
1
0.8
1.2
0.9
1.3
0.9
1.3
0.8
1.2
1.3
1
1.2
0.9
1.2
0.9
0.9
0.7
1
0.8
0.9
0.8
1
0.9
1.1
1
1.2
1
1.1
1
1
0.9
1
1
1.1
1.1
1.1
1.1
0.8
0.9
Female
Male
Female
Male
Female
Male
Female
Male

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%	51%
APL outpatient laboratory (210,906)	27	94%	84%	NA	NA
Ab-C open cohort (27,140)	18	32%	21%	21%	14%
CanPath closed cohort (25,156)	15	40%	31%	33%	26%
CLSA closed cohort (17,310)	11	50%	36%	34%	27%
CCAHS-1 closed cohort (11,050)	6	33%	20%	21%	13%

Notes: Date of sample collection was binned into 2-month intervals for each level of stratification. All specimen counts were unweighted.

Discussion

- Racialized minority and rural subgroups frequently underrepresented (compared to the Census-based adult population)

- Notable patterns in representation across study designs
  - Individuals in regions of high material deprivation underrepresented in blood donor populations (CBS) but not in outpatient populations (APL)
  - Older age groups were adequately represented across probabilistic cohorts (Ab-C, CanPath, CLSA), although younger age groups were better represented in Ab-C

- Residual blood convenience samples were more representative of some population dimensions than other study designs
  - May be a cost-effective avenue to obtain representative data in resource-constrained settings
  - But, probabilistically sampled surveys with sampling frames (**CanPath**) tend to have more contextual data, including non-response, that can be used to quantify and improve generalizability.

Limitations & conclusion

- Availability and differential measurement of participant race/ethnicity limited comparison of studies
  - Type and number of race/ethnicities listed differed between studies
- Representativeness assessment considered only six demographic variables
  - Sociodemographic variables related to health and disability likely important to generate representative estimates
- Ensuring representative estimates is a challenge for all study designs, but is important to consider to effectively detect disease trends between subgroups

Agenda

- Background
- Representativeness of SARS-Cov-2 serosurveillance studies
- **Adjusting for representativeness**
- Future directions

COMPARABILITY OF CANADIAN SARS-COV-2 SEROPREVALENCE ESTIMATES WITH STATISTICAL ADJUSTMENT FOR SOCIO-DEMOGRAPHIC REPRESENTATION

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

Objectives

- Assess population SARS-CoV-2 Anti-N seroprevalence in concurrent studies with statistical adjustment for sociodemographic representativeness

- Determine which adjustment method leads to the greatest concordance in seropositivity between studies

Project Overview

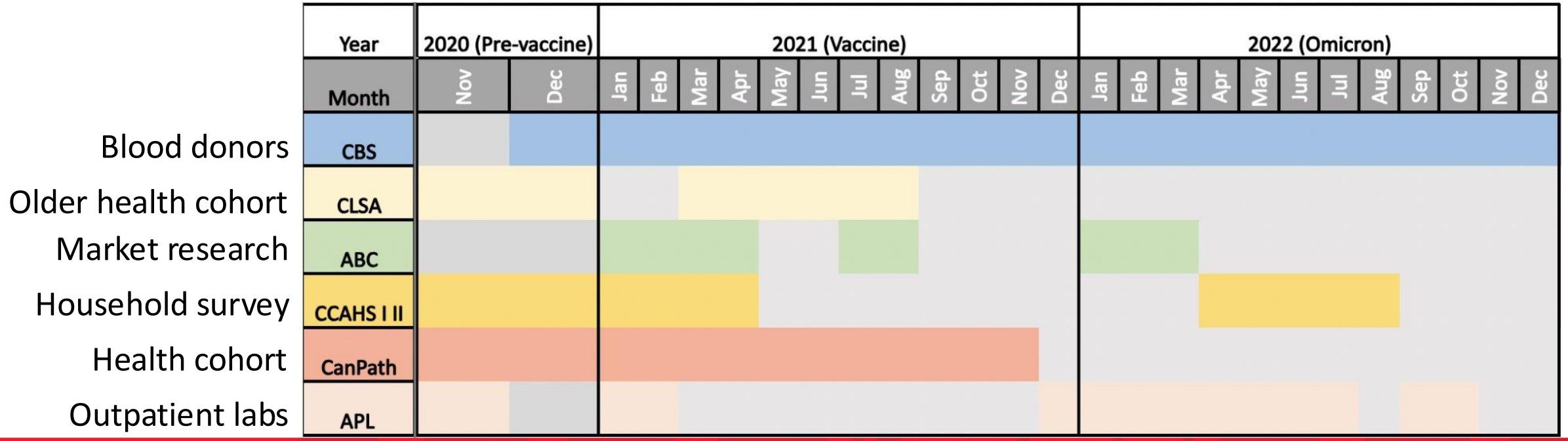
- Comparisons of Multilevel Regression with Poststratification (**MRP**) vs Unadjusted effect for all involved studies
- Simulation study exploring further the representative bias in the largest dataset (Canadian Blood Services)
- Risk factor analysis on involving additional features beyond those used for MRP

Concurrent SARS-CoV-2 serosurveillance studies in Canada

Study	Population/design	Specimen collection
CBS: Canadian blood services	Blood donor convenience sample	Serum
APL: Albert Precision Labs	Provincial lab convenience sample	Plasma, serum, or heparinized plasma
CLSA: Canadian Longitudinal Study on Aging	Pre-existing longitudinal cohort	Dried blood spot or plasma
ABC: Action to beat coronavirus	Pre-existing open market research cohort	Dried blood spot
CCAHS: Canadian COVID-19 Antibodies and Health Survey	Prospective cross-sectional household study	Dried blood spot
CanPath: Canadian Partnership for Health	Pre-existing longitudinal cohort	Dried blood spot

Data collection periods

- Partial overlap in data collection periods
- Defined ~2 – 6-month periods with sufficient overlap to assess methods across province/regions
- For some studies, also compared seropositivity over time using multilevel logistic regression with splines



Methods for seroprevalence estimation

Non-temporal setting

- No adjustment (crude seroprevalence)
- Raking on age and sex
- Raking on age, sex, and race (racialized minority vs. white)
- Post-stratification on age, sex, race, urban vs rural
- Multi-level regression with post-stratification **(MRP)**

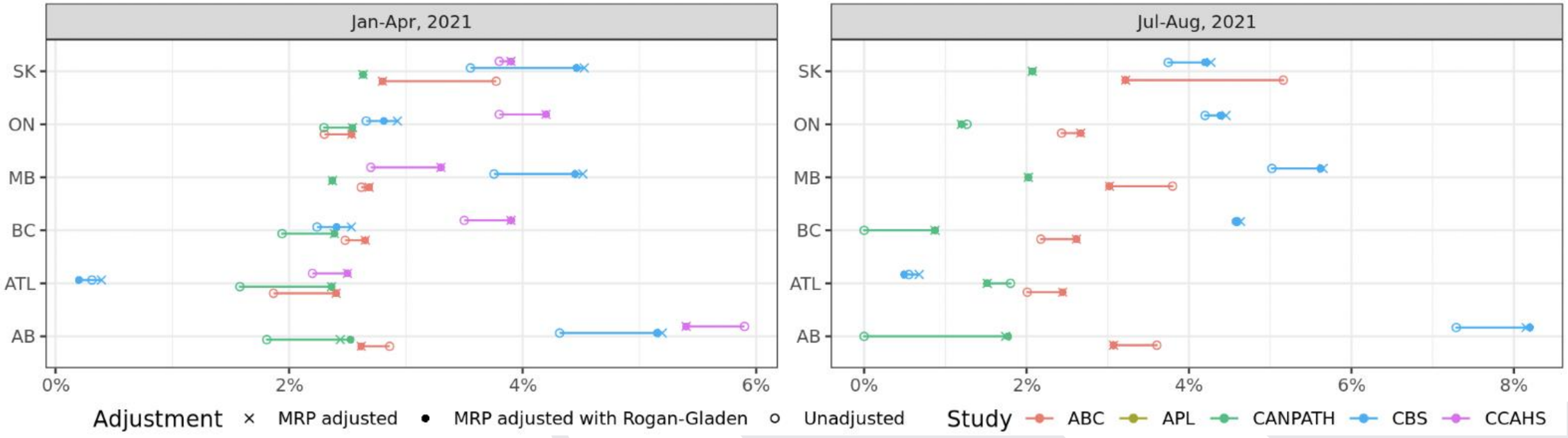
Temporal regression with spline

- Crude monthly seropositivity
- Multilevel regression without post-stratification
- Multilevel regression with post-stratification **(MRP)**

Use Canadian Census for distribution of characteristics in target population

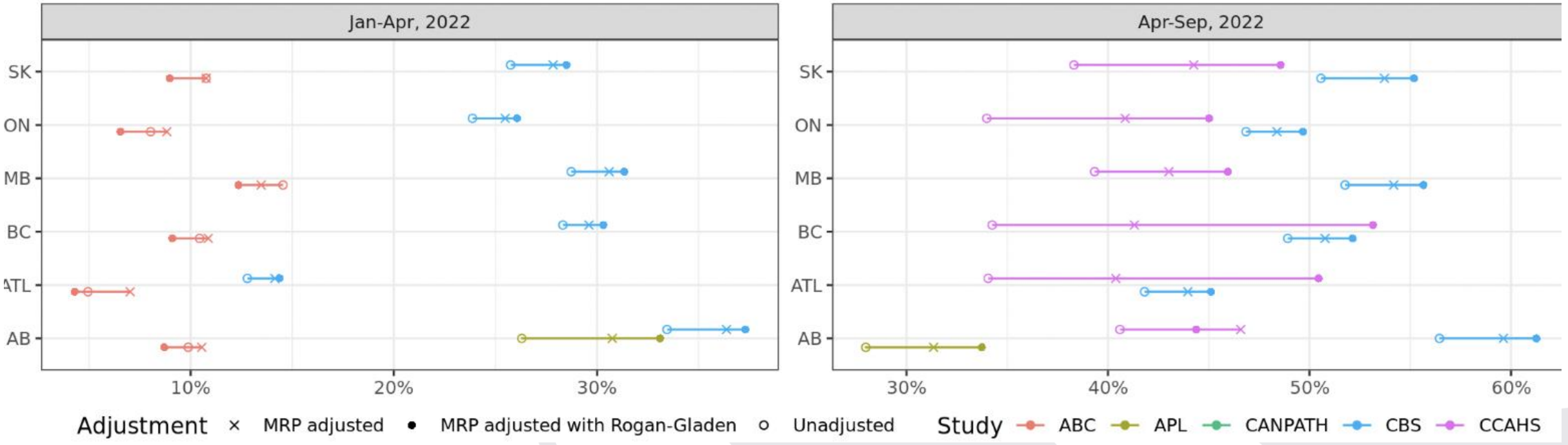
2021: Seroprevalence by region & period

- Seroprevalence ~1% to 8%
- MRP shifted study different directions in different regions by up to 2%
- Rogen-Gladen assay performance adjustment made little difference

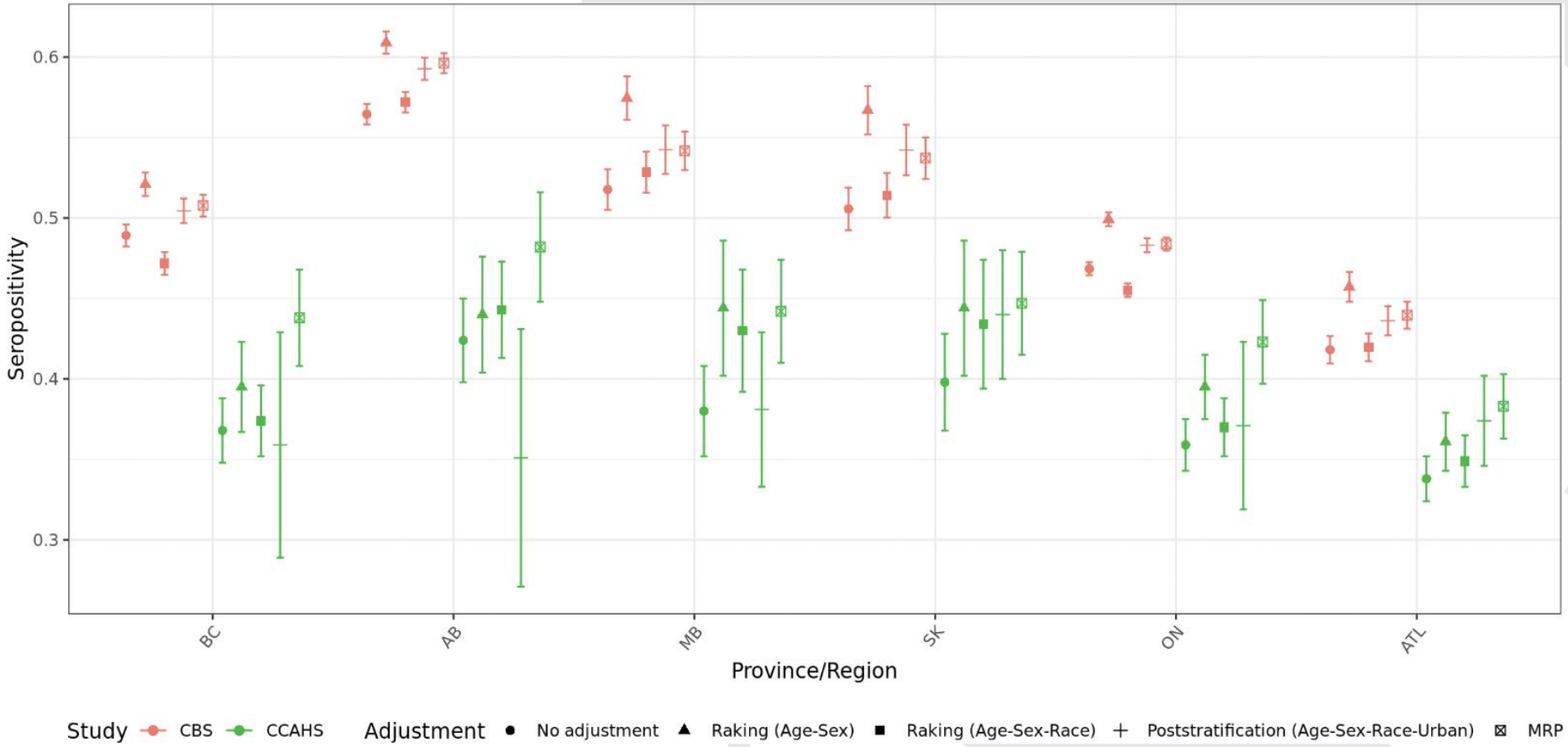


2022: Seroprevalence by region & period

- Seroprevalence ranged from ~10% to ~60%
- MRP estimates higher (up to 1.5X unadjusted) but did not always improve concordance between studies
- Adding Rogen-Gladen adjustment decreased estimate when seroprevalence above ~15%

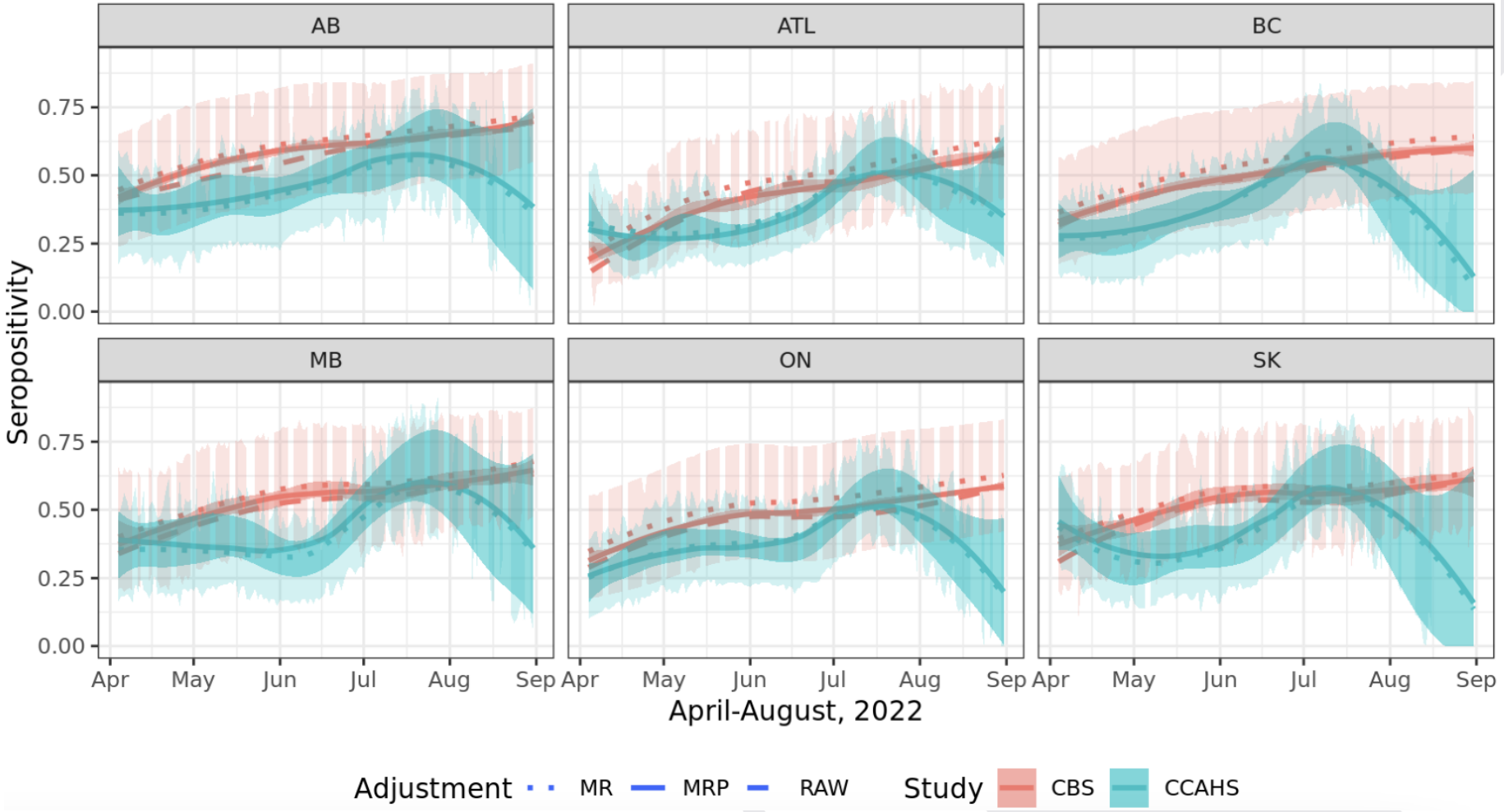


Omicron seropositivity: CBS blood donors vs. CCAHS household survey



Regardless of method, adjustment only **partially improved concordance** between studies **some of the time**

Temporal seropositivity estimation during Omicron period (April to August 2022)



In temporal regression, post-stratification **had little impact** on seropositivity over time and on concordance between study populations

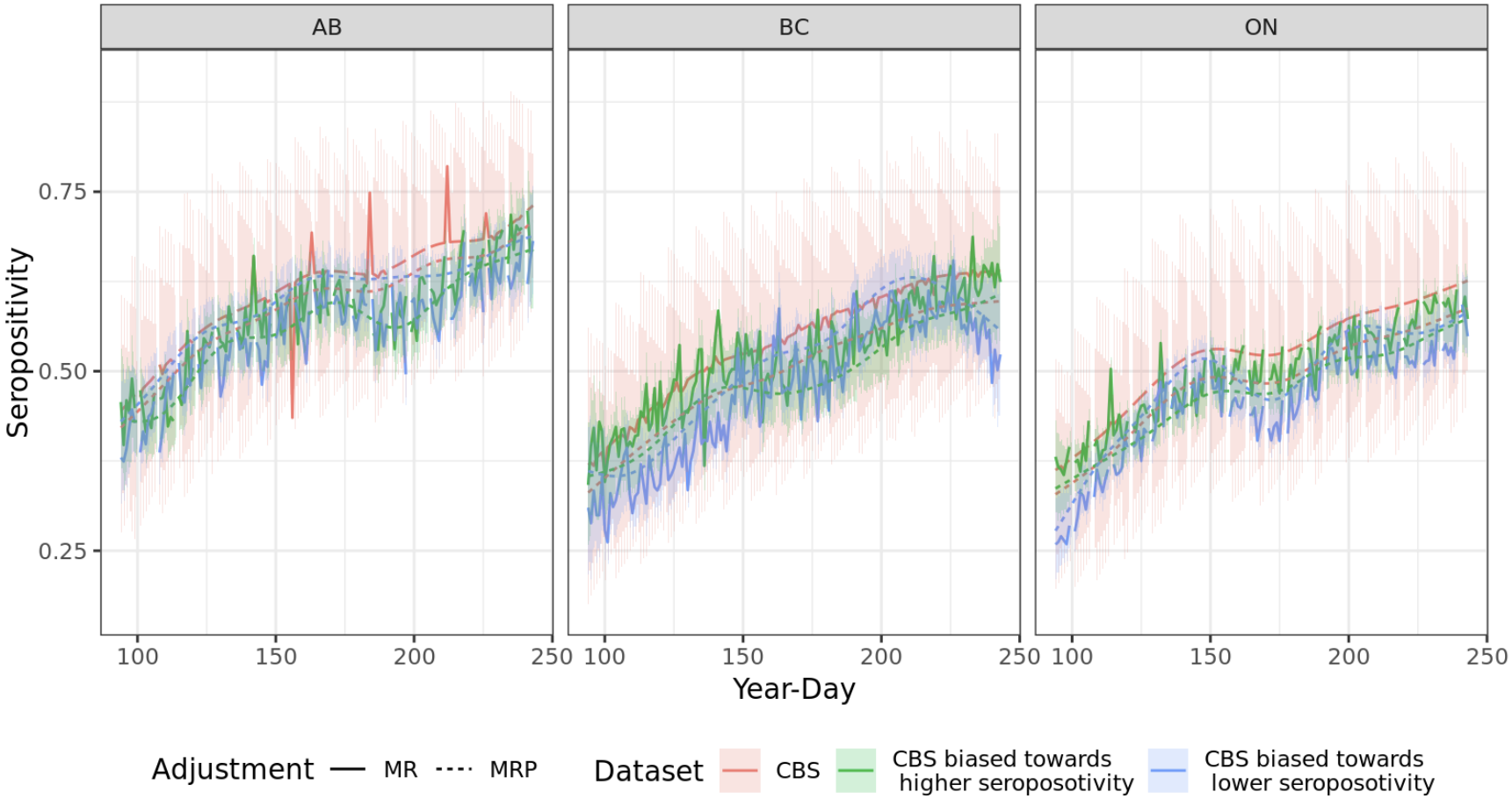
Method: MRP in CBS sub-populations

- **Goal**: Assess MRP in a synthetic setting with identical data collection but different distributions of characteristics associated with seropositivity

- Sub-sampled three sub-populations of CBS blood donors
  - CBS unbiased – completely random sub-sample
  - CBS biased high – oversampled younger, racialized minority
  - CBS biased low – oversampled older, white

- Assessed concordance in regression-based seropositivity over time with or without adjustment

Result: MRP in CBS sub-populations



- Unadjusted seropositivity closer than between-study estimates
- Distance between studies with post-stratification lower by 39% to 77% in every region

Method: Factors increasing odds of seropositivity

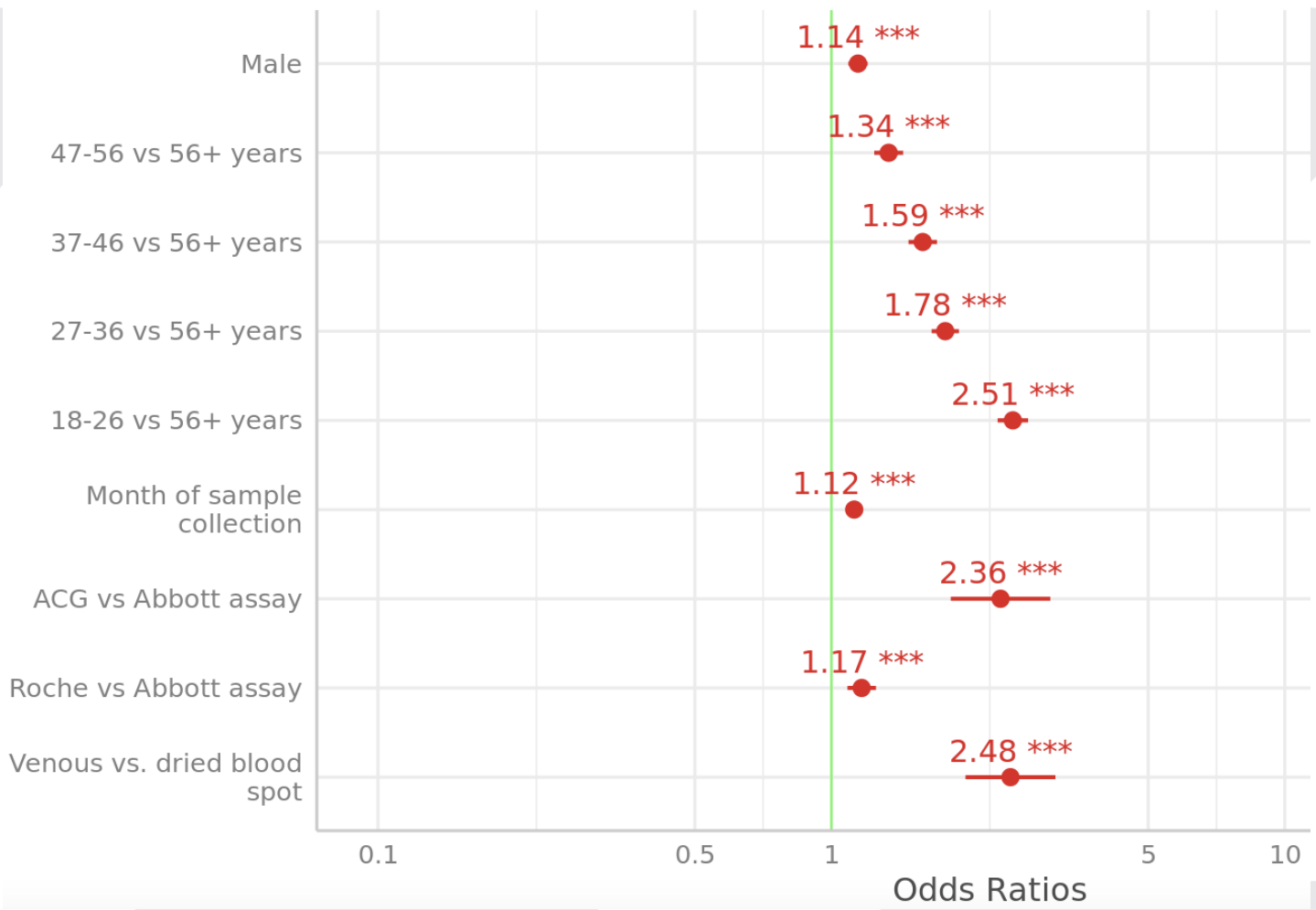
- **Goal:** Compare the affect of demographics to other study design features that could impact seropositivity estimates
- Combined individual-level data from CBS, APL, CLSA, ABC, CanPath
- Regressed seropositivity on:
  - Sex, age group
  - Assay
  - Sample type (serum vs. dried blood spot)
- Assess association between each characteristic and odds of seropositivitiy

Result: Factors increasing odds of seropositivity

Odds ratio of 2–3 for:

- Age 18-26 (vs. 56+)
- ACG assay vs. Abbott
- Sample collection with venous draw (vs. dried blood spot)

Assay and specimen type have a large impact on seropositivity



Agenda

- Background
- Representativeness of SARS-Cov-2 serosurveillance studies
- Adjusting for representativeness
- **Future directions**

Other dimensions of representativeness

To analyze more dimensions of representativeness, we can:


Map postal code to area-level indicators


Administer questionnaires


Link to administrative data

CHALLENGE: Concerns of privacy/consent; alignment across data holders

Linkage to administrative data

Residual sample cohort

General population comparators

Canadian Blood Services

ALBERTA PRECISION LABORATORIES

Linked administrative data

Demographics, dissemination area, COVID PCR (tests date, results), vaccination, discharge data for all hospitalizations, billing data for clinical visits, long-term disability claims

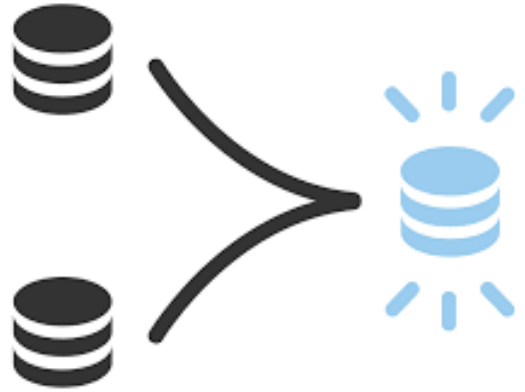
CHALLENGE: Data linkage took >1.5 years!

IC/ES

Alberta Health Services

Multi-model data synthesis

- Ideally, multiple data streams could be integrated for enhanced surveillance
  - Multiple residual blood sources
  - Surveillance in pre-existing cohorts (CanPath)
  - Prospectively sampled cohorts
- Actionable surveillance requires near-real-time data sharing
- Federated approach could address privacy concerns
  - Statistical adjustments applied locally then aggregate estimates generated centrally



CHALLENGE: also need to adjust for lab/assay/sample differences

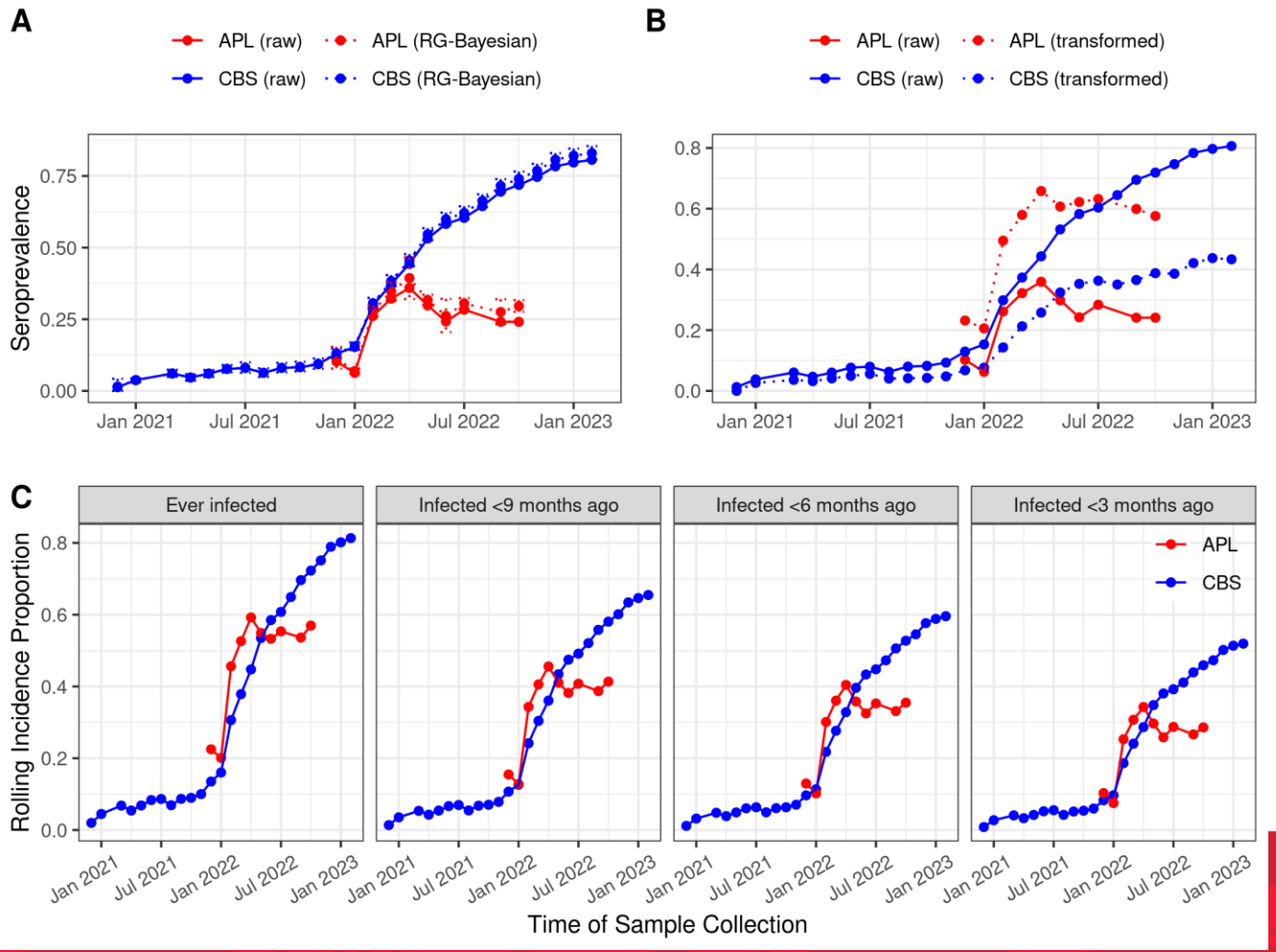
The impact of statistical adjustment for assay performance on inferences from SARS-CoV-2 serological surveillance studies

Jiacheng Chen[1], Yuan Yu[1], Sheila F. O'Brien[2,3], Carmen L. Charlton[4,5,6], Steven J. Drews[4,5], Jane M. Heffernan[7], Amber M. Smith[8], Yu Nakagama[9], Yasutoshi Kido[9], David L. Buckeridge[1], W. Alton Russell*[,1]

American Journal of Epidemiology, 2025, 00, 1–9
https://doi.org/10.1093/aje/kwaf157
Advance access publication date July 22, 2025
Novel Methodology

Regression-based transformation methods improved concordance in serosurveillance estimates between:

- **Blood donor population using Roche Anti-N**
- **Outpatient lab population using Abbott Anti-N**



McGill

Take-home messages

1. Representativeness is a challenge for **all** study designs

   Certain residual data streams well-suited for certain use cases; good for monitoring over time

2. For SARS-CoV-2 serology studies, population sampled seems to not be the main cause of differences across studies

   Hard to attribute, but differences in assay, threshold, sample, timing, etc. are likely more important

3. We can only account for measured variables

   Capturing race/ethnicity data, questionnaires, and **timely** linkage to administrative data needed to assess, improve and adjust for (under)representativeness

Thank you!

- Yuan Yu, McGill
- Matthew Knight, McGill
- Jiacheng Chen, CITF/McGill
- David Buckeridge, McGill
- Sheila O’Brien, Canadian Blood Services

alton.russell@mcgill.ca

Source: pmlive.com