BMJ Open

Canadians' opinions towards COVID-19 data-sharing: a national cross-sectional survey

Sarah A Savic Kallesoe ![ORCID],[1,2] Tian Rabbani ![ORCID],[1,3] Erin E Gill,[4] Fiona Brinkman,[4] Emma J Griffiths,[1] Ma'n Zawati,[5] Hanshi Liu,[5] Nicole Palmour,[5] Yann Joly ![ORCID],[5] William W L Hsiao[1]

![Check for updates]

[1]Simon Fraser University Faculty of Health Sciences, Burnaby, British Columbia, Canada
[2]Department of Public Health and Primary Care, University of Cambridge School of Clinical Medicine, Cambridge, UK
[3]School of Kinesiology, The University of British Columbia Faculty of Education, Vancouver, British Columbia, Canada
[4]Department of Molecular Biology and Biochemistry, Simon Fraser University Faculty of Sciences, Burnaby, British Columbia, Canada
[5]Department of Human Genetics, McGill University Faculty of Medicine and Health Sciences, Montreal, Québec, Canada

Correspondence to
Dr William W L Hsiao;
wwhsiao@sfu.ca

ABSTRACT

Objectives COVID-19 research has significantly contributed to pandemic response and the enhancement of public health capacity. COVID-19 data collected by provincial/territorial health authorities in Canada are valuable for research advancement yet not readily available to the public, including researchers. To inform developments in public health data-sharing in Canada, we explored Canadians' opinions of public health authorities sharing deidentified individual-level COVID-19 data publicly.

Design/setting/interventions/outcomes A national cross-sectional survey was administered in Canada in March 2022, assessing Canadians' opinions on publicly sharing COVID-19 datatypes. Market research firm Léger was employed for recruitment and data collection.

Participants Anyone greater than or equal to 18 years and currently living in Canada.

Results 4981 participants completed the survey with a 92.3% response rate. 79.7% were supportive of provincial/territorial authorities publicly sharing deidentified COVID-19 data, while 20.3% were hesitant/averse/unsure. Datatypes most supported for being shared publicly were symptoms (83.0% in support), geographical region (82.6%) and COVID-19 vaccination status (81.7%). Datatypes with the most aversion were employment sector (27.4% averse), postal area (26.7%) and international travel history (19.7%). Generally supportive Canadians were characterised as being ≥50 years, with higher education, and being vaccinated against COVID-19 at least once. Vaccination status was the most influential predictor of data-sharing opinion, with respondents who were ever vaccinated being 4.20 times more likely (95% CI 3.21 to 5.48, p=0.000) to be generally supportive of data-sharing than those unvaccinated.

Conclusions These findings suggest that the Canadian public is generally favourable to deidentified data-sharing. Identifying factors that are likely to improve attitudes towards data-sharing are useful to stakeholders involved in data-sharing initiatives, such as public health agencies, in informing the development of public health communication and data-sharing policies. As Canada progresses through the COVID-19 pandemic, and with limited testing and reporting of COVID-19 data, it is essential to improve deidentified data-sharing given the public's general support for these efforts.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The sampling strategy was designed to maximise external validity by matching demographic sample quotas to the 2016 Canadian census data, the most recent population-level data available at the time.

⇒ Health research patient partners, a plain language specialist and community partners were included from the outset in the development of the study design, including questionnaire language, scope and recruitment to improve the accessibility of the survey.

⇒ Detailed questionnaire that allowed for the assessment of public opinions on individual-level COVID-19 data-sharing specific to the Canadian context.

⇒ Despite efforts to make the questionnaire language as accessible as possible and defining all scientific terms (eg, 'anonymised', 'genetic information'), it is possible that respondents did not provide fully informed responses.

⇒ Participants were recruited exclusively through online platforms, which may under-represent groups of people who don't engage with online surveys.

BACKGROUND

Data-sharing in Canada and internationally

Publicly sharing population health and clinical data contributes to evidence-informed preventative public health action, especially during pandemic emergencies. Canada faces systemic barriers to individual-level data-sharing, largely due to its decentralised healthcare structure; instead of a single healthcare system, there are 10 provincial, 3 territorial and 1 federal system, each providing care for different constituent groups, typically based on geographical region.[1] Each system is the sole owner and steward of its health data, including surveillance and patient data.[2] Currently, there is no pan-Canadian standard for health data-sharing,[3] though such a system would help inform public health infrastructure and patient care, as well as maximise the utility of existing data.[4 5] Under the

Personal Information Protection and Electronic Documents Act (PIPEDA), using and disclosing unconsented individual-level data is limited, with exceptions for data-sharing in emergencies.[6] Where obtaining consent may not be practical, deidentification of the data could be an alternative and within the legislative framework.[6 7] During pandemic situations, such exceptions should permit data-sharing with researchers who are qualified, authorised and vetted by the provincial/territorial health authorities who own the health data.[8] Though such legislation allows for the federal government to make regulations of epidemiological data collection, especially in pandemic conditions,[9] it does not obligate provinces and territories to share such data with other health agencies, including the Public Health Agency of Canada.[10] The World Health Organization (WHO) conducted an evaluation of Canada's data-sharing efforts and its compliance with International Health Regulations capacities in 2018. The report concluded that existing legislation does not mandate interjurisdictional data-sharing[9] and that 'informal collegial relationships with provincial/territorial health authorities have been essential for public health surveillance and response to acute public health events across Canada'.(WHO, p.27)[11]

Lack of legal obligations affects the timeliness of data-sharing between provincial, territorial and federal health agencies, as well as data-sharing with the public. An example of a public data-sharing initiative is the Global Initiative on Sharing All Influenza Data (GISAID). GISAID contains the largest international SARS-CoV-2 genomic database with accompanying individual-level contextual data, and is publicly accessible—though users are required to register and agree to the terms of use.[12] The public accessibility of GISAID data has enabled the rapid research and development of diagnostic testing, therapies, vaccines and identification of risk factors across population subgroups.[12 13] The data submitted to GISAID and other public archives such as the NCBI GenBank and Canada's VirusSeq Data Portal, referred to as 'deidentified COVID-19 data', includes the genomic sequence data of SARS-CoV-2 samples and accompanying metadata. Metadata, sometimes known as 'contextual data', can include the date of sample collection, the health authority which conducted the testing, and information about the infected individual, frequently their age, sex, symptoms, travel history, vaccination status and residential province/territory. These metadata improve the quality of the interpretation of sequence data and are pertinent to epidemiological analyses. Throughout the pandemic, countries have contributed deidentified COVID-19 data to GISAID voluntarily. As of May 2021, Canada was in the unenvious position of sharing the least complete contextual data among the top 10 contributors to GISAID.[14] This could be partly attributed to the lack of mandated interjurisdictional data-sharing leading to delays and inconsistencies in sharing COVID-19 data with international repositories.

With the increasing role of the omic sciences in public health policy, researchers are now ever more involved in informing public health policies. Whether this is through mathematical modelling, or other methods, making individual-level data widely available is important for both research and public health. As of May 2022, improvements in Canada's data-sharing speed tie Canada with Moldova for 63rd place, out of 147 countries, vis-à-vis timely contributions to GISAID.[12 15 16] Beyond the lack of Canadian legal infrastructure mandating interjurisdictional data-sharing, there are other factors hindering data-sharing leading to sporadic improvements. Provincial/territorial governments have privacy-related concerns with publicly sharing deidentified COVID-19 data, specifically that of reidentification.[17] Reidentification is the process of piecing together deidentified data to identify the individual. The caution to prevent reidentification is important and valid, and provincial/territorial governments take steps, such as aggregation, to mitigate this risk before sharing data with other public health jurisdictions. For example, instead of stating the exact age in years of the individual with COVID-19, their age bracket is provided (eg, 0–9 years, 10–19 years). Despite these mitigation options, there is still hesitancy among the provinces/territories to release individual-level deidentified data, which is partly attributed to a perceived and assumed opposition of the public—that Canadians do not want their deidentified data being shared publicly.[13 18] Regardless of these assumptions, the population's opinion of publicly sharing deidentified individual-level data would likely influence its outcomes[19 20] and thus, there is a need to investigate the public's perceptions. This is especially so as the utility and necessity of enhanced data-sharing practices for navigating health emergencies grow.[21 22]

The existing literature provides a broad perspective of international attitudes towards the sharing of deidentified health data with researchers. The 'Your DNA, Your Say' study investigated public opinions of sharing such data with medical professionals, non-profit researchers (eg, government and academia), and for-profit researchers (eg, commercial sector) across 22 countries in 2017, including Canada.[23] The proportion of participants willing to share their deidentified health data for research ranged between 29.0% (Japan) and 63.7% (Mexico). Canada ranked below the median (50.7%, Belgium and the USA) with 46.0% of respondents willing to share their data for research. Respondents were most likely to be willing to share their data with a medical professional, followed by a non-profit researcher, and least likely to share with for-profit researchers. Another study provided an in-depth investigation of the UK public's willingness to share deidentified health data in 2020.[24] These findings suggest that there is broad, but not universal, support for sharing deidentified data to other clinics for direct care without explicit consent.[24] Net support for sharing deidentified health data for research to the National Health Service, academic and charitable sector, but not to the commercial sector, was observed; this is consistent with the 'Your DNA, Your Say' study findings that participants were most willing to share their health data with other medical professionals and least willing to share with for-profit researchers. A similar

deep-dive into Canadian public opinions of deidentified health data-sharing could prove interesting, especially considering the existing literature data were collected before the COVID-19 pandemic and public opinions may have shifted with the enhancement of data-sharing initiatives to enable pandemic response.

Objectives

In Canada, individual-level data are collected by the health authorities mostly at the provincial level, and are then shared only with public health authorities at the federal level.[8] Canadian privacy laws do not provide for automatic data-sharing for health research and Canada does not presently have an opt-out data-sharing clause for health data. Data protection legislation and privacy laws generally allow for public health authorities to share individual data without informed consent for public health purposes which may or may not explicitly include health.[6] These laws also give the public health authorities significant discretion in deciding whether to share or not to share the data. A particular reason for not sharing these data for research is the fact that such sharing could raise concerns among the Canadian population.

Public opinions have the potential to influence the success of data-sharing initiatives[25] and, consequently, there exists a need to gauge the Canadian public's opinions on COVID-19 data-sharing. To do this, a national study has been conducted by researchers at the Centre for Infectious Disease Genomics and One Health (CIDGOH) at Simon Fraser University (SFU) and McGill University's Centre of Genomics and Policy to survey a 5000-person cross-section of the Canadian population in Spring 2022.

The objective of this study is to examine the question, 'What are Canadians' opinions of provincial/territorial public health authorities publicly sharing deidentified COVID-19 genomic and health data for research?' Specifically, this study aims to understand:

1. The proportions of Canadians that are generally supportive of sharing deidentified COVID-19 data by public health authorities.
2. Canadians' comfort with publicly sharing each specific type of deidentified COVID-19 data.
3. The factors that are associated with increased support for publicly sharing deidentified COVID-19 data, compared with those who are generally hesitant/averse/unsure.

METHODS
Data collection

An online-only survey was constructed with options for both open-ended and close-ended responses. This survey was created iteratively by a team of researchers at SFU and McGill University, and was open to respondents from 1 March 2022 to 17 March 2022. Community members were consulted to help inform the utility and accessibility of the survey and Canadian-owned market research firm, Léger, recruited a representative sample of the Canadian population according to 2016 Census estimates for age,

sex, region and ethnicity.[26] To retain statistical power, less populous regions in geographical proximity were grouped. New Brunswick, Newfoundland, Nova Scotia and Prince Edward Island were grouped as the Atlantic. The Northwest Territories, Nunavut and the Yukon were grouped as the Territories.

The survey was available in English and French, with the translation performed by Léger and verified by the research team. To increase the level of confidence, and ensure robust base sizes for subgroup analysis, oversamples of the Territories and those identifying as Indigenous were conducted by Léger. Participants were recruited using probability sampling via Léger's online research panel, consisting of more than 450 000 members across Canada. The survey could be accessed through Léger's LEO application or any browser. Panel members received 1200 points (effectively CAD$1.20) and two chances in prize draws as remuneration for participation. No personal information, such as name, address, contact information or IP address, was collected at any point. Participants could respond, 'Prefer not to answer' ('PNTA'), or skip questions, as no question in the survey was mandatory, excluding the eligibility question. Léger provided both complete (n=5014) and incomplete response (n=385) datasets; selecting 'PNTA' was considered a response, while skipping was considered incomplete. Only the complete response dataset was used for this high-level analysis.

The survey consisted of 30 multiple-choice questions assessing respondent demographics, COVID-19 experience and data-sharing opinions, with one open-ended question for additional thoughts. Respondents were asked to state their comfort with sharing 16 different datatypes (online supplemental appendix A) by posing the following question: 'Would you be comfortable with the following 'anonymised' COVID-19 data collected from the population by public health authorities, which could potentially include your data, being publicly accessible?' Responses for data-sharing comfort included 'yes', 'no' and 'PNTA'. Willingness to participate in research was assessed by asking the following: 'If you were asked to provide researchers access to your 'anonymised' COVID-19 data for a study, would you agree?'. Responses to assess willingness were on a five-point frequency scale ranging from 'very willing' to 'very unwilling', with a sixth option to select 'PNTA'. To assess perceptions of risk, participants were asked 'How much risk do you think is associated with participating in research that involves your 'anonymised' COVID-19 data?' Responses to assess risk were also on a five-point frequency scale ranging from 'no risk' to 'a lot of risk', with a sixth option to select 'PNTA'.

Sample size calculation

A target sample size of n=5000 Canadian residents (≥18 years) was selected to ensure the sample size was large enough to allow for the smallest stratum to be used in the analysis. The smallest age stratum was the 70–75 year age group with a prevalence of 1.91%, based on the 2016 Canadian census.[26] In addition, the smallest stratum for ethnicity was Oceania representing 0.02% of the Canadian population.[26] With a target confidence level of 99%,

the margin of error of 0.50%, and the smallest stratum of interest having a prevalence of 1.91%, the required sample size was n=4973. This sample size would further decrease to n=2879 if the confidence level is decreased to 95%. Funds allowed for recruiting up to n=5000 participants.

Patient and public involvement

Community partners, primarily from the BC SUPPORT Unit, were consulted to review survey language accessibility and address the potentials for bias in our sample. The BC SUPPORT Unit is funded by the Canadian Institutes of Health Research and reinforces patient-oriented research in British Columbia.[27] Four patient partners were consulted to inform the survey preambles, length of time to complete the survey, the use of ethnicity as a variable, language accessibility, and overall survey structure. Over 10 hours were dedicated to patient partner consultations and all forms of feedback were addressed to produce the final survey. A CAD$50 honorarium was provided to each partner for their time and feedback.

Finally, to ascertain the readability, accessibility and correct comprehension of the survey, so as not to bias respondents towards any answer selection, knowledge mobilisation officer and plain language expert Dr. Lupin Battersby at SFU was consulted. Specifically, the language around 'genomics data-sharing' and the notion of 'anonymised' data were of interest in our discussion. Initially, 'deidentified COVID-19 data' was used instead of 'anonymised', however, during community consultations, we recognised that 'deidentified' was not an accessible term. On feedback, we used 'anonymised COVID-19 data' in the survey and defined it as 'COVID-19 data removed of personal information which could reveal the person's identity'.

Consent and ethics

Survey responses contained no direct identifiers, including names, phone numbers, addresses or IP addresses. The informed consent form was presented to participants on Léger's survey platform ('Decipher') at the beginning of the survey. Participants were required to click through the consent form to start the survey. The informed consent form is included in our online supplemental appendix A.

Analysis and data definitions

Eligibility for analysis included providing sex, age (≥18 years) and the region as the research team wanted to ensure external validity of the sample vis-à-vis the 2016 Canadian census estimates. There were three outcomes of interest for this study assessed among a sample of the general Canadian population:

i.   General support (ie, supportive vs hesitant/averse/unsure) regarding the public sharing of deidentified COVID-19 data.
ii. Willingness to participate in research requiring individual-level deidentified COVID-19 data.
iii. The perceived risk associated with participating in research involving individual-level deidentified COVID-19 data.

As seen in figure 1, to assess general comfort a binary variable 'general opinion' was constructed by combining responses to all 16 data-sharing questions, whereby participants who responded with 'yes' to nine or more data-sharing questions were categorised as generally supportive and all others were categorised as generally hesitant/averse/unsure. Regarding willingness, a binary variable was constructed whereby participants who responded with 'very willing' or 'somewhat willing' were categorised as supportive and participants who responded with all other
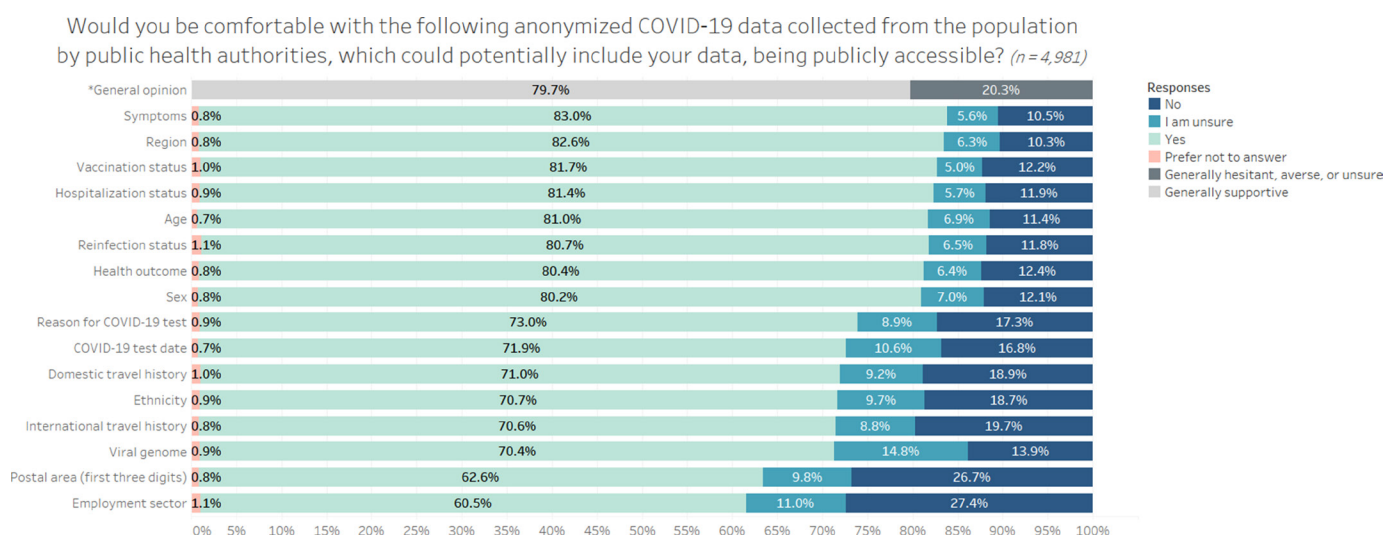
Figure 1 Comfort with data-sharing by datatype. *Participants' responses to the 16 datatypes are summarized in the "General opinion" variable. Participants who responded "Yes, I would be comfortable with this anonymized datatype being publicly shared" to nine or more datatypes were classified as "generally supportive". Those who responded "Yes" to eight or less were classified as "generally hesitant, averse, or unsure" Participants' responses to their comfort on human genome and linked data being shared with authorized researchers are not included in the variable.

options were categorised as hesitant/averse/unsure. The binary variable for risk categorised those that responded with 'no risk' or 'a little risk' as perceived little-to-no risk and all others as perceived risk or unsure. All outcome variables allowed for a comparison of predictor variables whereby the referent category for all outcomes was those who were hesitant/averse/unsure, or in the case of risk, the referent category was those that perceived little-to-no risk.

A total of seven predictor variables were included in the multiple logistic regression analyses, including age, provincial/territorial region, self-reported sex, education, ethnicity, ever testing positive for COVID-19 and ever being vaccinated against COVID-19. Odds ratios (ORs), both crude and adjusted, and 95% confidence intervals (CIs) were used to estimate the strength and significance of the associations at an alpha-level of 0.05 (online supplemental appendix B). Multicollinearity among the predictor variables was assessed using Spearman's coefficient and confirmed by the mean variance inflation factor (VIF). Multicollinearity was not found to be a threat to internal validity with a mean VIF of 1.67. Pairwise correlations were assessed to determine relationships between predictor variables, especially for sex and education. Age and ever testing positive for COVID-19 had the greatest collinearity with a correlation coefficient of −0.1694. Three different logistic regression models were fitted for model comparison, with only the age variable differing in its nature. Age was fitted as either a continuous, binary or aggregated variable and the model with age as an aggregated variable was selected as the best fit. Lasso was used initially to fit a saturated multivariate logistic regression model adjusting for all covariates. Model selection was further confirmed by the Akaike information criterion (AIC) and performances were assessed with R^2 and log likelihood. Analysis for variance (ANOVA) tests were conducted for predictor variables across all three models and various interaction terms were identified as potentially significant. On inclusion of the terms, the original models without the interaction elements were deemed a better fit according to the AIC and the significance values. All statistical analyses were conducted using Stata V.17.0/SE[28]

RESULTS

The original sample (n=5014) was reduced to n=4981 after removal of 33 respondents due to missing data across sex, age and region, which constituted our minimal epidemiological metadata. Removal of 33 respondents constitutes 0.7% lost due to missingness. A total of 5399 individuals opened or engaged with the survey, and 4981 respondents completed the survey in full, constituting a 92.3% response rate. The sample breakdown by age, sex, region, education, ethnicity, COVID-19 experience and COVID-19 vaccination status is detailed in table 1. Respondents who selected more than one ethnicity were categorised as 'mixed', detailed in table 1. Note that

Table 1	Demographics of the sample
Demographic variables	**n (%)**
Overall (N=4981)	
Age (years) (N=4981)	
18–29	822 (16.5)
30–39	878 (17.6)
40–49	879 (17.7)
50–59	818 (16.4)
60–69	911 (18.3)
70–79	577 (11.6)
80 and better	96 (1.9)
Region (N=4981)	
Alberta	522 (10.5)
Atlantic	356 (7.1)
British Columbia	640 (12.8)
Manitoba	213 (4.3)
Ontario	1764 (35.4)
Québec	1194 (24.0)
Saskatchewan	141 (2.8)
Territories	151 (3.0)
Self-reported sex (N=4981)	
Male	2431 (48.8)
Female	2535 (50.9)
Other	15 (0.3)
Education (N=4981)	
No certificate, diploma or degree	123 (2.5)
Secondary school level	1039 (20.9)
Non-university certificate or diploma	1489 (29.9)
Bachelor level	1609 (32.3)
Graduate level	696 (14.0)
Prefer not to answer	25 (0.5)
Ethnicity (N=4981)	
African	50 (1.0)
Caribbean	50 (1.0)
East and Southeast Asian	361 (7.3)
European	1127 (22.6)
Latin, Central and South American	47 (0.9)
Mixed*	441 (8.9)
North American	2206 (44.3)
North American Indigenous	218 (4.4)
Oceania	9 (0.2)
South Asian	151 (3.0)
West Central Asian and Middle Eastern	52 (1.0)
Prefer not to answer	269 (5.4)
COVID-19 (N=4981)	
Never tested positive or not reported	4314 (86.6)
Ever tested positive	667 (13.4)
	Continued

Table 1 Continued

Demographic variables	n (%)
Vaccination against COVID-19 (N=4981)	
Never vaccinated or not reported	409 (8.2)
Ever vaccinated	4572 (91.8)
*Ethnicity: mixed (breakdown)	**n (%)**
Overall (N=441)	
North American (N. A.), European	256 (58.0)
N.A. Indigenous, N. A.	45 (10.2)
N.A. Indigenous, European	32 (7.3)
N.A. Indigenous, N. A., European	25 (5.6)
Other unique ethnicity combinations	83 (18.8)
*Mixed category	

Table 2 Logistic regression model: General support of publicly sharing deidentified COVID-19 data

Independent variables	Adjusted OR	SE	P value	95% CI
Age (years)				
18–29	1.00	(Reference)		
30–39	1.25	0.174	0.115	(0.95 to 1.64)
40–49	1.18	0.171	0.257	(0.88 to 1.56)
50–59*	1.75	0.271	<0.000	(1.29 to 2.37)
60–69*	2.25	0.357	<0.000	(1.65 to 3.07)
70–79*	3.13	0.623	<0.000	(2.12 to 4.63)
80 and better* †	13.3	8.440	<0.000	(3.86 to 46.1)
Region				
Ontario	1.00	(Reference)		
Alberta	0.92	0.143	0.584	(0.68 to 1.25)
Atlantic	0.94	0.177	0.737	(0.65 to 1.36)
British Columbia	1.13	0.166	0.388	(0.85 to 1.51)
Manitoba	0.93	0.195	0.730	(0.62 to 1.40)
Québec	1.02	0.130	0.855	(0.80 to 1.31)
Saskatchewan	1.38	0.369	0.226	(0.82 to 2.33)
Territories	1.70	0.700	0.198	(0.76 to 3.80)
Self-reported sex				
Male	1.00	(Reference)		
Female	0.94	0.086	0.500	(0.79 to 1.12)
Other	0.73	0.395	0.563	(0.25 to 2.11)
Education				
No certificate, diploma or degree	1.00	(Reference)		
Secondary school level*	2.00	0.512	0.007	(1.21 to 3.30)
Non-university certificate or diploma*	3.00	0.766	<0.000	(1.81 to 4.94)
Bachelor level*	3.37	0.868	<0.000	(2.04 to 5.58)
Graduate level*	3.12	0.863	<0.000	(1.82 to 5.37)
Prefer not to answer	0.84	0.443	0.735	(0.30 to 2.36)
Ethnicity				
North American	1.00	(Reference)		
African	0.63	0.220	0.184	(0.32 to 1.24)
Caribbean	0.67	0.237	0.255	(0.33 to 1.34)
East and Southeast Asian	0.88	0.145	0.429	(0.64 to 1.21)
European	1.08	0.120	0.621	(0.85 to 1.32)
Latin, Central, and South American	0.89	0.338	0.766	(0.43 to 1.87)
Mixed ‡*	1.61	0.274	0.005	(1.16 to 2.25)
North American Indigenous	0.87	0.162	0.466	(0.61 to 1.26)
Oceania	0.78	0.580	0.743	(0.18 to 3.34)
South Asian*	0.63	0.138	0.035	(0.41 to 0.97)
West Central Asian and Middle Eastern	1.40	0.525	0.370	(0.67 to 2.92)
Prefer not to answer*	0.53	0.078	<0.000	(0.39 to 0.70)
COVID-19				
Never tested positive or not reported	1.00	(Reference)		

Continued

in this heterogeneous group, more than half (58.0%) identified as North American and European. 'Other unique ethnicity combinations' refers to ethnicity identity combinations which were represented by less than 10 respondents.

Comfort of data-sharing by datatype

Figure 1 illustrates respondents' comfort with the 16 different types of deidentified COVID-19 data being publicly accessible, sorted by the proportion of those who responded 'yes'. The first row, 'general opinion', represents a summary of participants' responses to the 16 datatypes. A general trend of support for data-sharing is apparent, with at least 60% approval for sharing all 16 datatypes. The datatypes with the greatest support were symptoms of the individual with COVID-19 (83.0%), region (82.6%) and COVID-19 vaccination status (81.7%). The datatypes with greatest opposition for sharing were employment sector (27.4%), postal area (different from the postal code, postal area is the first three digits of the postal code which identifies a rural region, medium-sized city, or section of a major metropolitan area)[29] (26.7%) and international travel history (19.7%).

Modelling support for public data-sharing

Tables 2–4 present three logistic regression models, predicting general support towards the public sharing of deidentified COVID-19 data (table 2), willingness to participate in research requiring their deidentified COVID-19 data (table 3) and the perceived risk associated with participating in research involving their deidentified COVID-19 data (table 4).

The dependent variable in table 2, general support of deidentified data-sharing, was coded such that 0=hesitant/averse/unsure and 1=supportive. The model suggests that general support for data-sharing is significantly associated with age, education, ethnicity and vaccination status. No trends were found across region, sex or COVID-19 experience.

Table 2 Continued

Independent variables	Adjusted OR	SE	P value	95% CI
Ever tested positive	1.28	0.177	0.069	(0.98 to 1.68)
Vaccination against COVID-19				
Never vaccinated or not reported	1.00	(Reference)		
Ever vaccinated*	4.20	0.572	<0.000	(3.21 to 5.48)
Constant	0.26	0.080	0.000	(0.14 to 0.48)

See online supplemental appendix B for crude ORs.
n=4981; Pseudo R^2=0.0860; Akaike information criterion=4607.1
*Result of interest; Results are reported as odds ratios whereby OR>1.00 means higher likelihood and OR<1.00 means lower likelihood.
†Wide CI; caution when interpreting the point estimate.
‡Please refer to table 1 for details of this heterogeneous group.
CI, Confidence Interval; OR, Odds Ratio; SE, Standard Error.

The odds of being generally supportive of sharing deidentified COVID-19 data with the public significantly increases with age (table 2); those ≥80 years were 13.3 times more likely (95% CI 3.86 to 46.1, p=0.000) to be generally supportive than those aged 18–29 years. Note that while the associated CI is wide, the same trend with narrower margins is found with the preceding age groups: those aged 70–79 years, 60–69 years and 50–59 years were respectively 3.13 times (95% CI 2.12 to 4.63, p=0.000), 2.25 times (95% CI 1.65 to 3.07, p=0.000) and 1.75 times (95% CI 1.29 to 2.37, p=0.000) more likely to be generally supportive than those aged 18–29 years.

A similar significant trend is observed across education: the odds that someone is generally supportive of publicly sharing deidentified COVID-19 data tends to increase with education. The odds of being supportive are twice as likely for those with secondary school-level credentials (95% CI 1.21 to 3.30, p=0.007), three times as likely for a non-university certificate or diploma credentials (95% CI 1.81 to 4.94, p=0.000), 3.37 times as likely for bachelor-level credentials (95% CI 2.04 to 5.58, p=0.000) and 3.12 times as likely for those with graduate-level credentials (95% CI 1.82 to 5.37, p=0.000) compared with those with no certificate, diploma or degree.

Regarding ethnicity, those who identified as South Asian, mixed* (ie, those who identified with more than one ethnicity), or preferred not to answer had significant associations when compared with the most populous group, North American. Those who identified as mixed* were 1.61 times more likely (95% CI 1.16 to 2.25, p=0.005) to be generally supportive, though the heterogeneity of this group ought to be considered. Individuals who identified as South Asian were 37% less likely to be generally supportive (OR 0.63; 95% CI 0.41 to 0.97, p=0.035); however, the upper bound of the CI edges close to the null. Those who did not report their ethnicity were 47% less likely to be generally supportive of data-sharing (OR 0.53; 95% CI 0.39 to 0.70, p=0.000).

COVID-19 vaccination status was the most influential factor of the model, with those who have ever been vaccinated against COVID-19 being 4.20 times more likely

Table 3 Logistic regression model: willingness to share their deidentified COVID-19 data with researchers

Independent variables	Adjusted OR	SE	P value	95% CI
Age (years)				
18–29	1.00	(Reference)		
30–39	1.00	0.134	0.978	(0.77 to 1.30)
40–49	0.87	0.114	0.292	(0.67 to 1.13)
50–59	1.23	0.176	0.146	(0.93 to 1.63)
60–69*	1.46	0.208	0.008	(1.10 to 1.93)
70–79*	1.84	0.314	<0.000	(1.32 to 2.57)
80 and better*	3.36	1.273	0.001	(1.60 to 7.06)
Region				
Ontario	1.00	(Reference)		
Alberta	0.94	0.130	0.640	(0.71 to 1.23)
Atlantic	1.00	0.170	0.991	(0.72 to 1.40)
British Columbia	1.23	0.165	0.116	(0.95 to 1.60)
Manitoba	1.20	0.238	0.355	(0.81 to 1.77)
Québec	0.96	0.110	0.751	(0.77 to 1.20)
Saskatchewan*	2.03	0.509	0.005	(1.25 to 3.32)
Territories	1.73	0.574	0.098	(0.90 to 3.32)
Self-reported sex				
Male	1.00	(Reference)		
Female	0.96	0.079	0.584	(0.81 to 1.12)
Other	0.70	0.350	0.469	(0.26 to 1.86)
Education				
No certificate, diploma or degree	1.00	(Reference)		
Secondary school level	0.83	0.211	0.465	(0.51 to 1.37)
Non-university certificate or diploma	1.13	0.285	0.621	(0.69 to 1.85)
Bachelor level	1.54	0.391	0.092	(0.93 to 2.53)
Graduate level*	1.77	0.482	0.035	(1.04 to 3.02)
Prefer not to answer	1.00	0.526	0.994	(0.36 to 2.80)
Ethnicity				
North American	1.00	(Reference)		
African	0.93	0.300	0.813	(0.49 to 1.75)
Caribbean	0.66	0.208	0.184	(0.35 to 1.22)
East and Southeast Asian*	0.69	0.103	0.013	(0.52 to 0.92)
European	1.08	0.108	0.449	(0.89 to 1.31)
Latin, Central and South American	0.56	0.177	0.065	(0.30 to 1.04)
Mixed*†	1.57	0.239	0.003	(1.16 to 2.11)
North American Indigenous	0.90	0.156	0.561	(0.65 to 1.27)
Oceania	0.62	0.425	0.485	(0.16 to 2.38)
South Asian*	0.66	0.133	0.041	(0.45 to 0.98)
West Central Asian and Middle Eastern	1.47	0.492	0.250	(0.76 to 2.83)
Prefer not to answer*	0.39	0.057	<0.000	(0.30 to 0.52)
COVID-19				
Never tested positive or not reported	1.00	(Reference)		

Continued

Table 3 Continued

Independent variables	Adjusted OR	SE	P value	95% CI
Ever tested positive	1.16	0.147	0.252	(0.90 to 1.48)
Vaccination against COVID-19				
Never vaccinated or not reported	1.00	(Reference)		
Ever vaccinated*	3.87	0.519	<0.000	(2.98 to 5.04)
Constant	0.57	0.168	0.058	(0.32 to 1.02)

See online supplemental appendix B for crude ORs.
n=4981; Pseudo R^2=0.0728; Akaike information criterion=5363.3
*Result of interest; Results are reported as ORs whereby OR>1.00 means higher likelihood and OR<1.00 means lower likelihood
†Please refer to Table 1 for details of this heterogeneous group
CI, Confidence Interval; OR, Odds Ratio; SE, Standard Error.

(95% CI 3.21 to 5.48, p=0.000) to be generally supportive of data-sharing compared with those who have never been vaccinated.

Model: willingness to participate in research

Table 3 details the logistic regression model predicting participants' willingness to participate in research which would require their deidentified COVID-19 data. The dependent variable, willingness to participate, is coded such that 0=hesitant/averse/unsure and 1=supportive. Age and vaccination status were observed to be the most influential predictors. Certain regions, ethnicities and education categories were correlated with willingness to participate in research. No significant associations were found for sex or COVID-19 experience.

Willingness to participate in research requiring deidentified COVID-19 data is significantly positively correlated with age starting at the 60–69 year age group. Compared with the 18–29 years age group, individuals aged 60–69 years, 70–79 years and ≥80 years were 1.46 times (95% CI 1.10 to 1.93, p=0.008), 1.84 times (95% CI 1.32 to 2.57, p=0.000) and 3.36 times (95% CI 1.60 to 7.06, p=0.001) more likely to be willing to participate in research, respectively. Only one region, Saskatchewan, was found to be significantly associated with willingness. Compared with Ontario, the most populous province, respondents in Saskatchewan were 2.03 times as likely (95% CI 1.25 to 3.32, p=0.005) to be willing to participate in research. With regards to education, only the graduate-level category was observed to have a significant association with willingness: compared with those without schooling credentials, graduate-level respondents were 1.77 times more likely (95% CI 1.04 to 3.02, p=0.035) to be willing to participate in research. However, the lower bound of this CI is close to the null.

Certain ethnicity groups were significantly associated with willingness to participate in research. Compared with those identifying as North American, those identifying as East and Southeast Asian were 31% less likely (OR 0.69; 95% CI 0.52 to 0.92, p=0.013) to be willing to participate; those identifying as mixed* were 1.57 times more

Table 4 Logistic regression model: perceived risk of sharing deidentified COVID-19 data with researchers

Independent variables	Adjusted OR	SE	P value	95% CI
Age (years)				
18–29	1.00	(Reference)		
30–39	0.81	0.101	0.088	(0.63 to 1.03)
40–49	0.80	0.100	0.077	(0.63 to 10.2)
50–59*	0.61	0.079	<0.000	(0.48 to 0.79)
60–69*	0.56	0.069	<0.000	(0.44 to 0.71)
70–79*	0.51	0.071	<0.000	(0.38 to 0.67)
80 and better*	0.35	0.104	<0.000	(0.19 to 0.63)
Region				
Ontario	1.00	(Reference)		
Alberta	0.78	0.098	0.053	(0.61 to 1.00)
Atlantic	0.86	0.125	0.299	(0.64 to 1.14)
British Columbia*	0.75	0.086	0.012	(0.60 to 0.94)
Manitoba	0.84	0.152	0.327	(0.59 to 1.20)
Québec	0.96	0.100	0.665	(0.78 to 1.17)
Saskatchewan	1.05	0.238	0.814	(0.68 to 1.64)
Territories	0.65	0.150	0.060	(0.41 to 1.02)
Self-reported sex				
Male	1.00	(Reference)		
Female	1.12	0.081	0.109	(0.97 to 1.30)
Other	1.43	0.800	0.516	(0.48 to 4.26)
Education				
No certificate, diploma, or degree	1.00	(Reference)		
Secondary school level	0.81	0.194	0.369	(0.50 to 1.30)
Non-university certificate or diploma*	0.60	0.142	0.032	(0.37 to 0.96)
Bachelor level*	0.48	0.113	0.002	(0.30 to 0.76)
Graduate level*	0.47	0.117	0.002	(0.29 to 0.77)
Prefer not to answer	0.53	0.293	0.251	(0.18 to 1.56)
Ethnicity				
North American	1.00	(Reference)		
African	1.44	0.460	0.259	(0.77 to 2.69)
Caribbean*	2.80	0.972	0.003	(1.42 to 5.53)
East and Southeast Asian*	1.93	0.265	<0.000	(1.47 to 2.52)
European	0.90	0.767	0.230	(0.76 to 1.07)
Latin, Central and South American*	3.28	1.147	0.001	(1.66 to 6.51)
Mixed*†	0.69	0.086	0.003	(0.54 to 0.88)
North American Indigenous*	1.43	0.230	0.027	(1.04 to 1.96)
Oceania*‡	8.18	8.640	0.047	(1.03 to 64.8)
South Asian*	2.25	0.442	<0.000	(1.53 to 3.31)
West Central Asian and Middle Eastern	1.33	0.412	0.361	(0.72 to 2.44)
Prefer not to answer*	1.75	0.258	<0.000	(1.31 to 2.33)
COVID-19				
Never tested positive or not reported	1.00	(Reference)		
Ever tested positive	1.11	0.123	0.351	(0.89 to 1.38)

Continued

Table 4 Continued

Independent variables	Adjusted OR	SE	P value	95% CI
Vaccination against COVID-19				
Never vaccinated or not reported	1.00		(Reference)	
Ever vaccinated*	0.44	0.062	<0.000	(0.33 to 0.58)
Constant	5.21	1.470	0.000	(2.99 to 9.07)

See online supplemental appendix B for crude ORs.
n=4981; Pseudo R^2=0.0498; Akaike information criterion=6630.6
*Result of interest; Results are reported as ORs whereby OR>1.00 means higher likelihood and OR<1.00 means lower likelihood
†Please refer to table 1 for details of this heterogeneous group.
‡Wide CI: caution when interpreting the point estimate.
CI, Confidence Interval; OR, Odds Ratio; SE, Standard Error.

likely (95% CI 1.16 to 2.11, p=0.003); those identifying as South Asian were 34% less likely (OR 0.66; 95% CI 0.45 to 0.98, p=0.041), though note the upper bound of the CI edges close to the null; and those who did not report their ethnicity were 61% less likely to be willing to participate (OR 0.39; 95% CI 0.30 to 0.52, p=0.000).

Vaccination status was the most strongly correlated. Compared with those who have never been vaccinated against COVID-19, those who have ever been vaccinated were 3.87 times more likely (95% CI 2.98 to 5.04, p=0.000) to be willing to participate.

Model: perceived risk of using data for research

Table 4 details the logistic regression model predicting participants' perceived risk of participating in research that requires their deidentified COVID-19 data. The dependent variable, risk perception, is coded such that 0=little-to-no-risk-perceived and 1=at-least-some-risk-perceived-or-unsure. Certain age groups, vaccination status, regions, ethnicities and education categories were correlated with perceived risk. No significant associations were found for sex or COVID-19 experience.

Perceived risk is significantly correlated with age, with the perception of risk diminishing with age. Compared with the 18–29 year age group, individuals aged 50–59 years were 39% less likely (OR 0.61; 95% CI 0.48 to 0.79, p=0.000) to perceive risk in participating in research; those aged 60–69 years were 44% less likely (OR 0.56; 95% CI 0.44 to 0.71, p=0.000); those aged 70–79 years were 49% less likely (OR 0.51; 95% CI 0.38 to 0.67, p=0.000); and those aged ≥80 years were 65% less likely (OR 0.35; 95% CI 0.19 to 0.35, p=0.000). The province of British Columbia was found to be significantly associated with the perception of risk. Compared with Ontario, respondents in British Columbia were 25% less likely (95% CI 0.60 to 0.94, p=0.012) to perceive risk in research participation.

An inverse trend was observed between the perception of risk and education. Those with a non-university certificate or diploma were 40% less likely (OR 0.60; 95% CI 0.37 to 0.96, p=0.032) to perceive risk with participating in research (note that the upper bound of the CI edges

near the null); those with bachelor-level credentials were 52% less likely to perceive risk (OR 0.48; 95% CI 0.30 to 0.76, p=0.002); and those with graduate-level credentials were 53% less likely to perceive risk (OR 0.47; 95% CI 0.29 to 0.77, p=0.002) compared with those with no certificate, diploma or degree.

Compared with those who self-identified as North American, those identifying as Caribbean were 2.80 times more likely to perceive research participation as risky (95% CI 1.42 to 5.53, p=0.003); East and Southeast Asian identities were 1.93 times more likely (95% CI 1.47 to 2.52, p=0.000); Latin, Central and South American identities were 3.28 times more likely (95% CI 1.66 to 6.51, p=0.001); those identifying as mixed* were 31% less likely to perceive risk (95% CI 0.54 to 0.88, p=0.003); those identifying as North American Indigenous were 1.43 times more likely (95% CI 1.04 to 1.96, p=0.027), though the proximity of the lower bound of the CI to the null is noteworthy; those identifying as Oceanian were 8.18 more likely (95% CI 1.03 to 64.8, p=0.047), though take note of the wide CI; those identifying as South Asian were 2.25 times more likely (95% CI 1.53 to 3.31, p=0.000); and those who did not report their ethnicity were 1.75 times more likely to perceive risk in research participation (95% CI 1.31 to 2.33, p=0.000).

Vaccination status appears to be associated with risk perception: those who have ever been vaccinated against COVID-19 were 56% less likely (OR 0.44; 95% CI 0.33 to 0.58, p=0.000) to perceive risk with research participation than those who have never been vaccinated or did not report their vaccination status.

Synopsis

We found that COVID-19 vaccination status was the strongest predictor for all three outcomes (general opinions on data-sharing (table 2), willingness to participate in research (table 3), and perceived risk (table 4)). Canadians who are generally supportive of the public sharing of deidentified COVID-19 data could be characterised as being ≥50 years, with higher education, and having ever been vaccinated against COVID-19 (table 2). No significant associations between general support and sex, COVID-19 experience or geographical region were found (table 2). Certain ethnicity groups were found to have significant associations with the support of data-sharing. Pairwise correlation analysis was conducted between all outcome and predictor variables. The strongest correlations were found between the outcome variables: r=0.53 between willingness and general support; r=−0.37 between willingness and perceived risk; and r=−0.33 between general support and perceived risk. These coefficients are suggestive of weak-to-moderate correlations, which may be further investigated in future explorations of this dataset.

DISCUSSION

This study investigates predictors of opinions around COVID-19 data-sharing, presents novel findings and serves as a baseline study. These results help inform provincial/territorial policy for deidentified data-sharing for SARS-CoV-2 and potentially future infectious disease outbreaks.

Prior to this study, the existing literature established that approximately 46.0% of Canadians in 2017 were willing to share their data with non-profit researchers.[23] This study observed that 79.7% of sampled Canadians in Spring 2022 were generally supportive of provincial/territorial public health authorities publicly sharing their deidentified COVID-19 data, while 20.3% were generally unsure or unsupportive. While the literature and this study are not perfectly comparable, the difference in the proportions willing to share their data could be a reflection of changing public opinion, especially considering the impact of COVID-19. This study also found that of the 16 deidentified COVID-19 datatypes, the three most supported for being shared publicly were symptoms (83.0% in support), region (82.6%) and vaccination status (81.7%). Finally, the three datatypes with the most aversion for being publicly shared were employment sector (27.4% averse), the first three postal code digits (26.7%) and international travel history (19.7%).

Viral genomic data

Respondents' comfort with sharing viral genomic data, as illustrated in figure 1, tells an interesting story. COVID-19 viral genomic data, once de-hosted as per standard practice, is considered by most jurisdictions as non-identifiable.[30] Yet it ranks quite low, even below more sensitive datatypes, such as ethnicity. This might be because 'viral genomic data' is an unfamiliar concept to the public. Despite the research team's efforts to define viral genomic data in plain language and implementing a page-advancing delay to encourage respondents to read the definition prior to answering questions, it is possible that this was not sufficient. Hence, the results for comfort with sharing viral genomic data are difficult to interpret without further exploring the respondents' reasoning.

Differential outcomes across ethnicity

Our findings indicate significant differences in opinions on data-sharing, willingness to participate in research and perceived risk across ethnicity. COVID-19 has impacted racialised communities disproportionately worse than their non-racialised counterparts.[31 32] These racial inequities in healthcare access have been particularly exacerbated throughout the COVID-19 pandemic.[31] Such differences in experience, fueled by ongoing systemic racism, may contribute to the variable data-sharing opinions across ethnicity groups. This is further supported by the homogeneity in opinions across those self-identifying as European and North American in this study. Due to the historical and ongoing effects of colonialism and systemic racism, ethnicity itself may not provide a causal mechanism for the observed differences in our findings. Rather than participants' ethnicity, racism and discrimination could provide a better and more appropriate explanation of these results and differing opinions.

Public health implications and recommendations with publicly accessible data

Our findings suggest a general trend of support for publicly sharing deidentified health data in Canada across demographic subgroups and regardless of COVID-19 experience. Despite the majority of respondents being supportive of publicly sharing all 16 datatypes presented in figure 1, some concerns with data-sharing, beyond the potential for reidentification, were observed by the authors. Some caveats to publicly accessible deidentified data involve the potential for the misinterpretation, misuse and conclusion of spurious associations of the data, especially when metadata regarding data collection and analytical methodologies are missing from data files.[13 33] There is also concern that parties involved in data collection might not be credited for their efforts, which could tarnish relations between those who collect and those who use the data, as well as impede the chain of data custody used to verify proper data usage. Though the findings suggest there is strong support for public accessibility, and publicly accessible data provides ease of access during pandemic or emergencies, there are methods to ensure its proper custody, use and accreditation.

The following recommendations have been developed based on the study findings. These recommendations further support the data-sharing guidance outlined by expert Canadian public health advisory committees.[4 34] The outlined suggestions aim to enhance data-sharing practices in manners that are trustworthy, transparent, encourage public engagement and uphold data utility. They include:

1. Establishing enhanced interjurisdictional data-sharing practices with a clear chain of command between provincial/territorial and federal public health authorities.
2. Creating a pan-Canadian centralised individual-level data repository with standardised and harmonised methods for data entry using ontologies (ie, standardised terminology and definitions).
3. Including accompanying metadata in data repositories in such a way that allows for the data and metadata to be findable, accessible, interoperable and reusable.[35] Examples of what metadata could include are data collection methods, sampling strategy, collection date and place, participant eligibility and devices used for collection, among others.
4. Practising greater transparency with the Canadian public in the incidental findings of surveillance data collection and analysis by public health authorities.
5. Establishing a Canada-wide emergency response data-sharing standard approved for continual use to eliminate the need to recreate interagency agreements for each emergency. Within this agreement, define:

- The minimum datatypes to be shared between agencies.
- Terms to create standardised data fields.
- A time frame within which provincial/territorial public health authorities ought to share data by.

Strengths and limitations

This study was conducted on a large sample of n=4981 participants to assess public opinions of the general Canadian population. Efforts were made to procure a robust and representative sample of the Canadian population to afford external validity; during sampling, key 2016 Canadian Census demographics were matched and weighted including age, sex, region and ethnicity. Further strengths of this study include patient/public involvement in the study design; assessing data-sharing opinion on myriad COVID-19 datatypes; the only such known study to be conducted on the Canadian population; and assessing COVID-19 disease and vaccination experiences and behaviours as potential predictors of data-sharing opinion.

The limitations of this study may offer pathways for future research. Cross-sectional studies pose difficulty in establishing causation and thus, any measures of association must be interpreted as correlational. Furthermore, the survey was only available in the English and French languages which could have led to population sub-groups being overlooked systematically due to language barriers. Though translations in other languages would have increased the accessibility of the survey, it is unclear how the measures of association would change in direction and magnitude. Self-reported sex options were limited to mitigate the risk of reidentification. Targeted follow-up studies on this topic, and future iterations of this survey, should prioritise the inclusion of typically under-represented groups such as sexual and ethnic minorities. The research team spent 6 months improving the survey's language accessibility by consulting with community partners, defining uncommon words, and implementing a 10-second page-advancing delay on the preamble pages. This encouraged respondents to thoroughly read definitions before answering questions. Despite this, there still exists plausible ambiguity regarding the public's knowledge of concepts such as genetic information and anonymised data. This is evidenced by the increased number of 'Unsure' and 'PNTA' responses for complex datatypes such as 'viral genome'. Though the magnitude is speculative, we suspect these gaps may have biased the results towards the null, indicating a weaker measure of association due to partial understanding. Older age groups in our sample may also be less representative of the adult Canadian population than younger age groups because respondents were recruited from a large online panel. This may imply higher technological literacy, and access to devices, was required to be eligible for and complete the survey, as has been typical of online and telephone surveys.[36 37] We hypothesise that older age groups involved in Léger's online panel may have higher technological literacy than their peers. Though we cannot say with certainty the magnitude or direction of this bias, we believe that this impacts the generalisability of the results seen among older age groups.

CONCLUSIONS

This survey explored Canadians' opinions towards deidentified data-sharing and uncovered predictors which may shape public opinions. The findings indicate a general trend of support for deidentified data-sharing among the adult Canadian population, significantly predicted by age, educational attainment and COVID-19 vaccination status. Second, these findings suggest that vaccination status is the most influential predictor of support for data-sharing. These findings add to existing literature surrounding Canadian data-sharing and pandemic preparedness and may be of interest to public health authorities and stakeholders in data-sharing initiatives. Subsequent investigations of this dataset may include further subgroup or datatype analysis alongside qualitative analysis of open-text responses. In addition, further investigation in partnership with Indigenous communities and other minorities should be conducted on their lived experiences of data-sharing, to better understand how systemic racism and trust in institutions, such as the public health system, relate to data-sharing initiatives.

ORCID iDs
Sarah A Savic Kallesoe http://orcid.org/0000-0003-1329-8275
Tian Rabbani http://orcid.org/0000-0002-9661-4749
Yann Joly http://orcid.org/0000-0002-8775-2322

REFERENCES

1. Government of Canada. Canada's health care system [Internet], 2016. Available: https://www.canada.ca/en/health-canada/services/canada-health-care-system.html [Accessed 19 Apr 2022].
2. National Human Genome Research Institute. Genomics and Medicine [Internet], 2020. Available: https://www.genome.gov/health/Genomics-and-Medicine [Accessed 05 Aug 2021].
3. Azarm-Daigle M, Kuziemsky C, Peyton L. A review of cross organizational healthcare data sharing. *Procedia Comput Sci* 2015;63:425–32.
4. Government of Canada. Public Health Agency of Canada. Expert Advisory Group Report 3: Toward a world-class health data system [Internet], 2022. Available: https://www.canada.ca/en/public-health/corporate/mandate/about-agency/external-advisory-bodies/list/pan-canadian-health-data-strategy-reports-summaries/expert-advisory-group-report-03-toward-world-class-health-data-system.html [Accessed 02 Jun 2022].
5. Institute of Medicine (US). Chapter 5: standardization to enhance data sharing. In: *Sharing Clinical Research Data: Workshop Summary [Internet]. Sharing Clinical Research Data: Workshop Summary*. US: National Academies Press, 2013: 43–55. https://www.ncbi.nlm.nih.gov/books/NBK137818/
6. Office of the Privacy Commissioner of Canada. Consent and privacy - A discussion paper exploring potential enhancements to consent under the Personal Information Protection and Electronic Documents Act [Internet], 2016. Available: https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2016/consent_201605/#fn15 [Accessed 02 Sep 2022].
7. Information & Privacy Commissioner of Ontario, CHEO Research Institute and University of Ottawa. Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy [Internet], 2011. Available: https://www.ipc.on.ca/wp-content/uploads/2016/11/anonymization.pdf [Accessed 09 Feb 2022].
8. Office of the Privacy Commissioner of Canada. Privacy and the COVID-19 outbreak [Internet], 2020. Available: https://priv.gc.ca/en/privacy-topics/health-genetic-and-other-body-information/health-emergencies/gd_covid_202003/ [Accessed 02 Sep 2022].
9. Attaran A, Houston A. Pandemic Data Sharing: How the Canadian Constitution Turned Into a Suicide Pact [Internet]. Rochester, NY: Social Science Research Network, 2020. Available: https://papers.ssrn.com/abstract=3612825 [Accessed 19 Apr 2022].
10. House of Commons. Debates (Hansard) No. 39 - June 13, 2006 (39-1) - House of Commons of Canada [Internet], 2006. Available: https://www.ourcommons.ca/DocumentViewer/en/39-1/house/sitting-39/hansard [Accessed 19 Apr 2022].
11. World Health Organization, Pan American Health Organization. Joint external evaluation of IHR core capacities of Canada: mission report, 11-20 June 2018 [Internet], 2019. Report No.: WHO/WHE/CPI/2019.62. Available: https://apps.who.int/iris/handle/10665/326138 [Accessed 19 Apr 2022].
12. GISAID. GISAID - Initiative [Internet]. Available: https://www.gisaid.org/ [Accessed 24 May 2022].
13. Song L, Liu H, Brinkman FSL, *et al*. Addressing privacy concerns in sharing viral sequences and minimum contextual data in a public Repository during the COVID-19 pandemic. *Front Genet* 2022;12.
14. Art Poon [@art_poon]. With the near-complete release of sample collection dates from Japan, Canada now has the worst record among top 10 contributors to GISAID, 2021. Available: https://twitter.com/art_poon/status/1368412572559548426 [Accessed 19 Apr 2022].
15. Ritchie H, Mathieu E, Rodés-Guirao L. Coronavirus (COVID-19) Cases. Our World in Data [Internet], 2020. Available: https://ourworldindata.org/covid-cases [Accessed 02 Jun 2022].
16. World Bank. GDP per capita, PPP (current international $) | Data [Internet]. World Bank. Available: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD [Accessed 02 Jun 2022].
17. Lake E. When Tracking the Virus Means Tracking Your Citizens. Bloomberg.com [Internet], 2020. Available: https://www.bloomberg.com/opinion/articles/2020-03-30/covid-19-tracking-threatens-civil-liberties-after-coronavirus [Accessed 19 Apr 2022].
18. van Panhuis WG, Paul P, Emerson C, *et al*. A systematic review of barriers to data sharing in public health. *BMC Public Health* 2014;14:1144.
19. Keller J. Ottawa, provinces must create agency to reform how health data is collected and used, report says [Internet]. The Globe and Mail, 2021. Available: https://www.theglobeandmail.com/canada/article-ottawa-provinces-must-create-agency-to-reform-how-health-data-is/ [Accessed 20 Apr 2022].
20. Aitken M, Cunningham-Burley S, Darlington A. Why the public need a say in how patient data are Used for Covid-19 Responses. International Journal of Population Data Science [Internet], 2020. Available: https://ijpds.org/article/view/1357 [Accessed 20 Apr 2022].
21. Public Health Agency of Canada. Expert Advisory Group Report 2: Building Canada's Health Data Foundation [Internet]. Government of Canada, 2021. Available: https://www.canada.ca/en/public-health/corporate/mandate/about-agency/external-advisory-bodies/list/pan-canadian-health-data-strategy-reports-summaries/expert-advisory-group-report-02-building-canada-health-data-foundation.html [Accessed 02 Jun 2022].
22. Aitken M, Tully MP, Porteous C, *et al*. Consensus statement on public involvement and engagement with data intensive health research. *Int J Popul Data Sci* 2019;4:586.
23. Middleton A, Milne R, Almarri MA, *et al*. Global public perceptions of genomic data sharing: what shapes the willingness to donate DNA and health data? *Am J Hum Genet* 2020;107:743–52.
24. Jones LA, Nelder JR, Fryer JM, *et al*. Public opinion on sharing data from health services for clinical and research purposes without explicit consent: an anonymous online survey in the UK. *BMJ Open* 2022;12:e057579.
25. Savić-Kallesøe S, Middleton A, Milne R. Public trust and genomic medicine in Canada and the UK. *Wellcome Open Res* 2021;6:124.
26. Government of Canada, Statistics Canada. Census of Canada: Data tables – Age (in Single Years) and Average Age (127) and Sex (3) for the Population of Canada, Provinces and Territories, Census Divisions, Census Subdivisions and Dissemination Areas, 2016 Census - 100% Data [Internet], 2016. Available: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Rp-eng.cfm?Lang=e&apath=3&detail=0&dim=0&fl=a&free=0&gc=0&gid=0&gk=0&grp=1&pid=109525&prid=0&ptype=109445&s=0&showall=0&sub=0&temporal=2016&theme=115&vid=0&vnamee=&vnamef [Accessed 20 Apr 2022].
27. British Columbia Academic Health Science Network. Michael Smith Health Research BC. About BC SUPPORT Unit | BC AHSN [Internet]. British Columbia Academic Health Science Network. Available: https://www.bcahsn.ca/our-units/bc-support-unit/about-bc-support-unit [Accessed 20 Apr 2022].
28. Stata. Stata 17 [Internet]. Available: https://www.stata.com/new-in-stata/faster-stata-speed-improvements/ [Accessed 30 May 2022].
29. Government of Canada. Forward Sortation Area—Definition [Internet]. Government of Canada. Innovation, Science and Economic Development Canada, 2015. Available: https://www.ic.

gc.ca/eic/site/bsf-osb.nsf/eng/br03396.html [Accessed 24 May 2022].

30 Luheshi L, Raza S, Moorthie S. Pathogen Genomics Into Practice [Internet], 2015. University of Cambridge. Available: https://www.phgfoundation.org/report/pathogen-genomics-into-practice [Accessed 02 Jun 2022].

31 Khazanchi R, Evans CT, Marcelin JR. Racism, not race, drives inequity across the COVID-19 continuum. *JAMA Netw Open* 2020;3:e2019933.

32 Mamuji AA, Lee C, Rozdilsky J, *et al*. Anti-Chinese stigma in the greater Toronto area during COVID-19: aiming the spotlight towards community capacity. *Soc Sci Humanit Open* 2021;4:100232.

33 Heeney C, Hawkins N, de Vries J, *et al*. Assessing the privacy risks of data sharing in genomics. *Public Health Genomics* 2011;14:17–25.

34 Naylor D, Basrur S, Bergeron MC. Learning from SARS: renewal of public health in Canada: a report of the National Advisory Committee on SARS and Public Health [Internet]. Ottawa: National Advisory Committee on SARS and Public Health, 2003. Available: https://www.canada.ca/content/dam/phac-aspc/migration/phac-aspc/publicat/sars-sras/pdf/sars-e.pdf

35 Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al*. The fair guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.

36 Goyder J, Warriner K, Miller S. Ses bias in survey nonresponse. *Journal of Official Statistics* 2002;18:1–11 https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/evaluating-socio-economic-status-ses-bias-in-survey-nonresponse.pdf

37 Singer E, Van Hoewyk J, Maher MP. Experiments with incentives in telephone surveys. *Public Opin Q* 2000;64:171–88.