How well do covariates perform when adjusting for sampling bias in online COVID-19 research? Insights from multiverse analyses

Keven Joyal-Desmarais1,2, Jovana Stojanovic2,3, Eric B. Kennedy4, Joanne C. Enticott5,6, Vincent Gosselin Boucher7, Hung Vo8, Urška Košir1,2, Kim L. Lavoie2,9, Simon L. Bacon1,2 on behalf of iCARE Study Team

Abstract
COVID-19 research has relied heavily on convenience-based samples, which—though often necessary—are susceptible to important sampling biases. We begin with a theoretical overview and introduction to the dynamics that underlie sampling bias. We then empirically examine sampling bias in online COVID-19 surveys and evaluate the degree to which common statistical adjustments for demographic covariates successfully attenuate such bias. This registered study analysed responses to identical questions from three convenience and three largely representative samples (total N = 13,731) collected online in Canada within the International COVID-19 Awareness and Responses Evaluation Study (www.icarestudy.com). We compared samples on 11 behavioural and psychological outcomes (e.g., adherence to COVID-19 prevention measures, vaccine intentions) across three time points and employed multiverse-style analyses to examine how 512 combinations of demographic covariates (e.g., sex, age, education, income, ethnicity) impacted sampling discrepancies on these outcomes. Significant discrepancies emerged between samples on 73% of outcomes. Participants in the convenience samples held more positive thoughts towards and engaged in more COVID-19 prevention behaviours. Covariates attenuated sampling differences in only 55% of cases and increased differences in 45%. No covariate performed reliably well. Our results suggest that online convenience samples may display more positive dispositions towards COVID-19 prevention behaviours than would samples drawn using more representative means. Adjusting results for demographic covariates frequently increased rather than decreased bias, suggesting that researchers should be cautious when interpreting adjusted findings. Using multiverse-style analyses as extended sensitivity analyses is recommended.

Keywords Sampling bias · Covariate adjustment · COVID-19 · Multiverse analysis · Selection bias · Collider bias
**Key messages**

- Online convenience samples are susceptible to important sampling bias; however, the nature of this bias, and how it can be best adjusted for analytically, are two questions that research has yet to fully answer.
- One bias in COVID-19 research is that participants may be more concerned about COVID-19 and hold more positive inclinations towards prevention measures (e.g., show higher vaccine acceptance) than those who do not participate.
- Adjusting analyses for demographic variables (e.g., sex, age, education) is a common and theoretically useful strategy to deal with sampling bias (e.g., when informed by causal theory), but most research uses an atheoretical/unstructured approach to select covariates.
- Adjusting analyses for demographic variables in an atheoretical/unstructured way may be unreliable to account for sampling bias; in nearly 17,000 models, we found covariates to reduce bias (i.e., discrepancies between convenience/representative samples) in 55% of cases and to increase bias in 45% of cases.
- Researchers that use convenience samples should consider multiverse-style covariate analyses (i.e., extended sensitivity analyses)—as demonstrated in this paper—to examine how covariate selection impacts findings.

**Sampling bias: a non-technical explanation**

Sampling bias occurs when different members of a population have unequal probabilities of being included in a study. This can occur for many reasons, such as when recruitment strategies have unequal reach for different groups, or when groups, once reached, differ in their response rates. Sampling bias can impact estimates of prevalence/incidence rates as well as of the link between exposure-outcome pairs. To understand sampling bias, and how to counter it, we can represent the phenomenon using causal diagrams such as Panel A of Fig. 1 [11, 12].

Using Fig. 1, consider an illustrative example. Imagine we are conducting research to estimate rates of COVID-19 vaccine acceptance (i.e., our outcome, defined as vaccine receipt or intentions to get vaccinated) in a community, and wish to explore how sex (our exposure) influences acceptance. In this study, we compare two recruitment strategies: a convenience-based and a probability-based sampling method. Importantly, our analyses are restricted to the selection of responses we obtain (we have no data on non-respondents). How might we expect sampling to affect our findings under these circumstances?

First, recruitment strategies will influence the selection of responses we obtain (path p1). In an ideal probability-based design, all population members would have an equal likelihood of being reached and efforts would be made (e.g., using incentives) to ensure high participation rates. In contrast, in convenience-based samples, reach is usually skewed towards certain groups (e.g., social media users for a study advertised on social media) and small/absent incentives can skew participation further [13, 14]. Second, participant characteristics can impact responses, either in conjunction with, or independently from recruitment strategy. For our example, research has shown that female (vs. male) individuals are more likely to volunteer for research (path p2) [15, 16] and we could anticipate that people are more likely to participate in vaccine-related research if they hold favourable attitudes towards vaccines (due to human tendencies to seek information in line with pre-existing beliefs; path p3) [17, 18].

The result of these forces (of paths p1, p2, p3) is that we will observe a series of biased findings if we attempt to use the convenience sample to draw inferences about the population (compared to using the probability sample).
How well do covariates perform when adjusting for sampling bias in online COVID-19 research?…

Specifically, we will overestimate the degree to which participants are female and accepting of vaccines (paths $p_4$ and $p_5$) and will also spuriously find that vaccine acceptance is lower among female (vs. male) participants (path $p_6$)—even if, in the overall population, no such association exists (see Box 1). These three biased findings are spurious and are manifestations of sampling bias.

Fig. 1 Examples of sampling bias and the roles of covariates. The black square around selection indicates that analyses are limited to individuals who participated (either through selection by a study’s design or through self-selection). Selection is a collider, a common effect of other variables. Panel A is an example of sampling bias shown through a causal diagram. Here, recruitment strategy (convenience vs. probability), along with an exposure (sex) and an outcome (vaccine acceptance) each influence a person’s selection into a study. Though the three variables are not causally linked, conditioning on selection (a collider) leads them to be associated through a process known as collider bias. Panel B is a simulated example of the dynamic in Panel A where 50% of a population is accepting of vaccination, and this ratio is equivalent for male and female individuals. However, both vaccine acceptance and sex predict selection. Having data only from people who participate will lead an analyst to overestimate vaccine acceptance and see a spurious association between sex and vaccine acceptance such that female (vs. male) participants show lower levels of acceptance (also see Box 1). Panel C provides example roles covariates can play in an association between an outcome (vaccine acceptance) and selection. Adjusting analyses for a mediator (confirmation seeking) or a confounder (education) can both reduce sampling bias. However, adjusting for another collider (employment) can introduce further collider bias. Thus, analysts must be mindful of the causal role of covariates in relation to their exposure-outcome links of interest.
Box 1 Three ways of understanding how collider bias creates spurious findings

Below are three ways of understanding/depicting how conditioning analyses on selection induces bias. We use the dynamic in Fig. 1’s Panel A as a guiding example throughout.

(1) Deductive Account: When a person participates in a study, we know that something caused them to do so. Learning about the status of one cause allows us to predict the status of other causes. For example, if participation is caused by (a) being recruited through more rigorous methods, (b) being female, and (c) being accepting of vaccines, then learning that a person is male eliminates one possible cause and other causes become more probable (i.e., “if Y is caused by a, b, or c, then, given Y, the absence of b implies the presence of a or c”). Stated another way, given participation, an inverse association exists between the causes, allowing us to predict the presence/absence of each from the presence/absence of the others.

(2) Graphical/Simulated Account: Panel B of Fig. 1 displays how collider bias can create a spurious association between sex and vaccine acceptance. In a simulated population (Panel B, left side), there is no association between sex and vaccine acceptance: 50% of male and 50% of female individuals accept vaccines. However, being female and accepting vaccines both predict study participation (blue dots, shown separately in the upper right side of Panel B). For female participants, there will be a clear overrepresentation of those that accept vaccines, but also a good representation of female vaccine refusers. Among male participants, there will be a good representation of those who accept vaccines, but a clear underrepresentation of those who refuse vaccines. Overall, vaccine acceptance rates will be more inflated for male than female participants, leading us to spuriously conclude that sex impacts rates of vaccine acceptance. A similar misleading conclusion would still occur if analyses were limited to non-participants (lower right side of Panel B).

(3) Mathematical Account: Focusing again on the link between sex and vaccine acceptance, let us posit the following parameters:

- \( P \) = The odds of participating in research for male persons who do not accept vaccination.
- \( a \) = The effect of accepting vaccination on people’s odds of participating in research.
- \( f \) = The effect of being female (vs. male) on people’s odds of participating in research.

We can compare the degree to which a study is expected to represent male participants who accept (\( P_a \)) vs. refuse vaccination (\( P_f \)). Algebraically, if \( a \) and \( f \) are positive, the former ratio will always be larger than the latter.

\[
\frac{P_a}{P_f} > \frac{P_{a+f}}{P_{f+a}}
\]

This becomes clear if we expand the above ratios to have a common denominator:

\[
\frac{(P+a)(P+f)af}{P(a+f)(f+a)} > \frac{(P+a+f)(a+f)af}{P(a+f)(f+a)}
\]

In other words, our sample will overrepresent male acceptance to a greater degree than it overrepresents female acceptance. Once again, collider bias will lead to the spurious conclusion that female individuals have lower acceptance (compared to male individuals).

Generally, sampling bias emerges through a process known as collider bias, whereby an association is induced (or distorted) between two variables because analyses are conditioned on a common outcome of those variables—known as a collider [5, 11, 12, 19]. In Fig. 1, selection is a collider (a common outcome) of recruitment strategy, sex, and vaccine acceptance. Conditioning analyses on selection—by limiting analyses to participants—leaves us at the root of the biased observations (paths \( p_1, p_3, \) and \( p_6 \)). An example of this dynamic is provided in Panel B of Fig. 1, demonstrating how limiting analyses to participants induces a spurious (and inverse) association between sex and vaccine acceptance (path \( p_6 \)). Given the importance of collider bias for understanding sampling bias, Box 1 provides three ways of conceiving/understanding this concept.

How to reduce/eliminate sampling bias

Of central interest to researchers is the question: how can we reduce or eliminate (the effects of) sampling bias? One way is to rely on representative sampling, but this will often be unfeasible and sometimes even undesirable [20, 21]. Alternatively, we can disrupt the dynamic that leads to sampling bias analytically by using covariates within statistical models (e.g., adding covariates to a regression, or by using propensity scoring) [8–10]. For instance, in Fig. 1, the spurious path \( p_6 \) (between sex and vaccine acceptance) occurs because analyses are conditioned on selection (Box 1). If we can analytically keep selection from acting as a collider, we can eliminate this bias. To do so, we can disrupt path \( p_3 \), so that there is no effect from vaccine acceptance to selection (i.e., in the absence of \( p_3 \), selection is no longer a common cause of sex and vaccine acceptance) or disrupt path \( p_2 \) so that there is no effect from sex to selection. Likewise, we can also eliminate the spurious paths \( p_4 \) (or \( p_5 \)) by disrupting the causal effects \( p_1 \) and \( p_2 \) (or \( p_1 \) and \( p_3 \)). Unfortunately, identifying covariates for these tasks is easier said than done. In practice, covariates play a multitude of causal roles, each of which have unique implications for disrupting/amplifying paths leading to selection.

Panel C of Fig. 1 demonstrates this complexity. If a causal link exists between an outcome and a collider (\( p_3 \) from Panel A), adjusting for a variable that accounts for this causal link (a mediator) can reduce sampling bias. In our example, we reasoned that vaccine acceptance would cause self-selection because people seek attitude-confirming information. Thus, we could measure and adjust for confirmation-seeking
behaviour. To fully disrupt the association between an outcome and self-selection, however, we should also adjust for confounders. In Panel C, higher education promotes participation in research [15, 22] and greater vaccine acceptance [23]; education should therefore be adjusted for. That said, one should also avoid adjusting for additional colliders as doing so can introduce further collider bias. For example, if vaccine mandates exist for employment [24, 25] (i.e., vaccination predicts employment) and certain personality factors like conscientiousness facilitate both survey participation [26] and employment [27], then adjusting for employment may increase bias. Consequently, researchers must be very careful in their choice of covariates (and similar cautions could be made for disrupting any causal pathway in Fig. 1; i.e., \( p_1, p_2, \) or \( p_3 \)).

These concerns are not novel, and many articles give guidance on how to use causal theory/diagrams to select covariates [4, 5, 11, 12, 19]. Unfortunately, systematic reviews find that it remains rare for research to adequately justify covariate selection choices, especially by using a causal perspective [28–32]. Instead, researchers frequently rely on heuristics/norms (e.g., always adjusting for demographics variables like sex, age, socioeconomic status), focus on variables for which population-data is readily accessible (also typically demographic variables), use all available covariates in their data, or rely on simple statistical rules such as controlling for any covariate known to relate to either the exposure or the outcome [30–34]—with each of these criteria failing to distinguish between confounders, mediators, and colliders [11, 12].

Researchers also vary widely in their selection of covariates even when examining similar research questions [32–35]. For instance, nutritional epidemiology work studying the same outcomes rarely adjust for the same sets of covariates [32]. This issue was particularly well-captured in two methodological studies [34, 35] which recruited 29 and 120 research teams, respectively, and tasked teams to independently answer the same research question using the exact same dataset. In both studies, most teams opted for unique selections of covariates (distinct from all other teams). Clearly, there is much uncertainty as to which covariates investigators should and shouldn’t include in analyses, and relatedly, as to whether most covariate choices in the literature are useful for attenuating bias.

Goals of the current study

Being able to identify and adjust for sampling bias is an important goal for science. This is particularly true in contexts like the COVID-19 pandemic, when urgency in decision-making can allow biased findings to have undue repercussions on scientific/public discourse and on policy making [5–7]. With this in mind, we set out with two primary goals.

First, we sought to inform future efforts to attenuate sampling bias by qualifying who gets recruited through online convenience sampling in COVID-19 research. Given research on selective-exposure to attitude-congruent information [17, 18], we hypothesised that participants recruited using convenience methods (versus those recruited through more representative means) would display higher levels of concerns about COVID-19, hold beliefs that prevention behaviours are more important, and show greater adherence to behavioural recommendations (e.g., social distancing, mask wearing, vaccination).

Second, given that adjusting analyses for demographic covariates (e.g., adding variables in a regression) is a common method for addressing sampling bias, we sought to evaluate the frequency with which this technique successfully accounts for and attenuates sampling bias within online surveys. To account for how researchers make different choices on which covariates to adjust for, we made use of multiverse analyses [36, 37], an analytical perspective that urges analysts to evaluate how all plausible study choices can influence their results (i.e., by running and reporting results for all analytic choices they could have justifiably made). In our case, this entailed evaluating the degree to which all combinations of a set of plausible and common demographic covariates (e.g., sex, age, education) were successful in attenuating sampling bias in a set of convenience samples.

Methods

This project (e.g., hypotheses, analyses) was registered a priori on the Open Science Framework (https://osf.io/f2p6j), and a project page hosts supplemental files (https://osf.io/dp9kd/).

Data source

We used three online convenience samples (\( N = 3225; 884; 609 \)) and three largely representative web-panel samples (\( N = 3003; 3005; 3005 \)) of Canadians recruited over three time periods in 2020 (summarized in relation to the pandemic in Fig. 2). These data represent cross-sectional surveys that were deployed as part of the International COVID-19 Awareness and Responses Evaluation (iCARE: www.icarestudy.com) Study [38]. The convenience-based samples consisted of unpaid volunteers recruited using a combination of online advertising (by iCARE team members) and snowball sampling (e.g., encouraging participants to share the survey within their own networks). In
contrast, *web panel* participants were paid and recruited through Léger, a polling and marketing firm that is commonly employed by researchers aiming to recruit representative samples of Canadians [39, 40]. Participants were drawn from Léger’s LEO panel, a panel of over 400,000 Canadians that was predominantly constructed using probability-based sampling methods (e.g., random-digit dialling) [41]. Additional details on the recruitment/sampling used for the current project are available in the supplemental files (Section 1), as well as through other iCARE-related publications [38, 42, 43].

**Measures**

Our predictor variable of interest was the type of sample participants were recruited from (convenience vs. web panel). We analysed differences between samples on the 11 *outcome* variables summarized in Table 1. These were selected and registered in line with the first goal of this article and included participants’: pandemic-related concerns (e.g., about getting infected, losing one’s ability to earn income); adherence to various preventative behaviours (e.g., mask wearing); and intentions to get vaccinated against COVID-19. For our multiverse analyses, we examined the influence of nine covariates that were consistently measured across surveys. These included participants’: province of residence; age; sex; highest education level attained; employment status pre-COVID; student status; parent status; perceived relative household income; and ethnic identity. These were selected as each of these factors has previously been associated to sampling bias in online research [15, 16, 44–47]. A detailed account of how each outcome and covariate was assessed is provided in the supplemental files.

**Analyses**

*Sampling bias* was operationalized as the discrepancy in results between the convenience samples and the web-panel samples. We conducted simple (unadjusted) linear regressions to identify such discrepancies on each outcome variable per time point. An alpha of 0.01 was chosen to be conservative when making inferences (see registration for rationale). Given that some outcomes were assessed using single Likert-type items, we also computed ordered logistic regressions; the results were equivalent to the regression-based models and are reported in the supplemental files. *Change in bias* due to covariate adjustments was operationalized as reductions/increases in the discrepancy between the sample types (convenience vs. web panel) in adjusted models compared to their unadjusted counterparts. We employed *specification curves*, a type of multiverse analysis that use caterpillar plots and other visual tools to examine how data-analytic choices impact estimates of interest [48]. In our case, for each outcome (at each time point), 512 unique models could be specified. These ranged from regressions with no covariate-based adjustments to regressions that adjusted for all covariates. To reflect how using covariates typically operates in practice, we further specified our models according to normative practice in the field: we used an alpha of 0.05 to compute inferential statistics, and refrained from modelling higher-level terms (e.g.,
interactions) between covariates. Although we limit our analyses to regression models, our procedure should generally produce convergent results with other common methods to deal with sampling bias, such as the use of sample weights derived from the same set of covariates (e.g., using raking or propensity score-based methods [49–52]).

All analyses were conducted using R version 4.1.0 [53]. Specification curves used the `specr` and `rifanalysis` packages [54, 55]. Our analysis code is available on our project page.

### Results and interpretations

#### Sample demographics

Table 2 presents demographic information on our samples and compares them to the 2016 Canadian census. Overall, the web panels were generally similar in composition to the Canadian population—unsurprising, as Léger panels were explicitly designed to reflect the Canadian population on attributes like sex, age, and region. However, there was some

| Table 1 | Summary of Outcome Measures Evaluated (Full Measures in Supplemental Files) |
|---------|--------------------------------------------------------------------------------|
| Outcome | Item #a | Example item | Anchors exampleb |
| 01. Perceived importance of prevention measures | 1 | To what extent do you believe that the measures asked of you by your government or local health authority are important to prevent and/or reduce the spread of COVID-19? | 1 = Not at all important; 4 = Very important |
| 02. Concern: Being infected personallyc | 2 | Because of COVID-19, I am concerned [about] being infected myself | 1 = Not at all; 4 = To a great extent |
| 03. Concern: Others Being Infectedc | 5 | Because of COVID-19, I am concerned [about] a family member with whom I do not share my home being infected | 1 = Not at all; 4 = To a great extent |
| 04. Concern: Economic Impactsc | 2 | Because of COVID-19, I am concerned [about] my country going into an economic recession/depression | 1 = Not at all; 4 = To a great extent |
| 05. Concern: Personal Livelihoodc | 3–4 | Because of COVID-19, I am concerned [about] losing my job/family income | 1 = Not at all; 4 = To a great extent |
| 06. Behaviour: Hand Washing | 2 | Please indicate the frequency with which you have adopted each [behaviour] in the previous 7 days: Hand washing with soap and water | 1 = Never; 4 = Most of the time |
| 07. Behaviour: Mask Wearing | 1 | Please indicate the frequency with which you have adopted each [behaviour] in the previous 7 days: Wearing a face mask | 1 = Never; 4 = Most of the time |
| 08. Behaviour: Social Distancing | 1 | Please indicate the frequency with which you have adopted each [behaviour] in the previous 7 days: Staying at least 6 feet or 1–2 [meters] away from other people | 1 = Never; 4 = Most of the time |
| 09. Behaviour: Self-Quarantine | 2 | Please indicate the frequency with which you have adopted each [behaviour] in the previous 7 days: Self-quarantining if you are returning from a trip | 1 = Never; 4 = Most of the time |
| 10. Behaviour: Avoiding Social Gatherings | 1–3 | Please indicate the frequency with which you have adopted each [behaviour] in the previous 7 days: Avoiding all social gatherings (large and small) | 1 = Never; 4 = Most of the time |
| 11. Vaccine Intentions | 1 | If a vaccine for COVID-19 were available today, what is the likelihood that you would get vaccinated? | 1 = Very unlikely; 4 = Extremely likely |

---

*a* When a range is presented, this indicates the number of items varied across time points for this measure

*b* Anchors shown here represent the direction in which responses were coded for analyses. Direction of scoring was standardized across outcomes such that higher scores would indicate higher perceived importance, concerns, rates of behaviours, and intentions

*c* The iCARE concerns module was divided to measure four distinct factors based on psychometric analyses (e.g., factor analyses). The results of these analyses are detailed in the supplemental materials.
Table 2  Demographic Distribution of the Samples (Presented as Percentages)

| Variable                        | 2016 Census (%) | Web Panel | Convenience Sample |
|---------------------------------|-----------------|-----------|--------------------|
|                                 |                 | Time 1 N= 3003 | Time 2 N= 3005 | Time 3 N= 3005 | Time 1 N= 3225 | Time 2 N= 884 | Time 3 N= 609 |
| Province/Region                 |                 |             |                   |                 |              |              |                |
| Atlantic Canada                 | 6.66            | 6.66       | 6.66              | 6.66            | 4.85         | 1.47         | 3.33           |
| Quebec                          | 23.30           | 22.98      | 23.03             | 22.96           | 83.83        | 80.41        | 78.20          |
| Ontario                         | 38.38           | 38.63      | 38.60             | 38.74           | 5.84         | 11.19        | 10.15          |
| Canadian Prairies               | 18.39           | 18.38      | 18.34             | 18.34           | 2.13         | 3.24         | 4.49           |
| British Columbia                | 13.27           | 13.35      | 13.38             | 13.31           | 3.35         | 3.68         | 3.83           |
| Sex a                           |                 |             |                   |                 |              |              |                |
| Male                            | 49.11           | 49.42      | 47.93             | 46.89           | 24.85        | 24.54        | 25.55          |
| Female                          | 50.89           | 50.42      | 51.80             | 52.91           | 74.80        | 75.23        | 73.94          |
| Other                           | .17             | .27        | .20               | .20             | .34          | .23          | .51            |
| Age b                           |                 |             |                   |                 |              |              |                |
| 15–24                           | 14.56           | 7.45       | 9.94              | 7.56            | 7.29         | 20.21        | 9.28           |
| 25–54                           | 48.43           | 54.73      | 54.45             | 50.37           | 53.49        | 50.75        | 45.88          |
| 55–64                           | 16.75           | 16.35      | 14.66             | 18.18           | 19.67        | 13.01        | 24.40          |
| 65+                             | 20.25           | 21.46      | 20.96             | 23.89           | 19.55        | 16.03        | 20.45          |
| Language of Choice c            |                 |             |                   |                 |              |              |                |
| English                         | 63.75           | 80.99      | 77.94             | 78.34           | 31.13        | 41.86        | 43.02          |
| French                          | 19.97           | 19.01      | 22.06             | 21.66           | 67.78        | 54.30        | 56.32          |
| Other or Multiple               | 16.28           | .          | .                 | .               | 1.09         | 3.85         | .66            |
| Mother Tongue d                 |                 |             |                   |                 |              |              |                |
| French                          | 20.61           | 20.02      | 23.16             | 21.82           | .            | .            | .              |
| English                         | 55.97           | 62.49      | 60.58             | 61.16           | .            | .            | .              |
| Mix or other                    | 23.42           | 17.49      | 16.26             | 17.02           | .            | .            | .              |
| Highest Education               |                 |             |                   |                 |              |              |                |
| Less than secondary/high school | 18.29           | 1.28       | .94               | 1.24            | .44          | 1.33         | .86            |
| Secondary/high school           | 26.45           | 30.51      | 31.01             | 30.95           | 22.25        | 26.55        | 21.34          |
| College/university: bachelor's or less | 47.52 | 51.18 | 51.85 | 51.43 | 42.03 | 38.20 | 41.59 |
| Graduate, postgraduate           | 7.74            | 17.04      | 16.20             | 16.38           | 35.28        | 33.92        | 36.21          |
| Ethnic Identity/Origin e        |                 |             |                   |                 |              |              |                |
| Aboriginal                      | 4.86            | 2.03       | 2.27              | 2.00            | .80          | .96          | .53            |
| African/Black                   | 3.48            | 1.86       | 2.95              | 2.71            | 1.26         | .19          | .53            |
| Asian                           | 16.44           | 14.68      | 13.86             | 14.58           | 3.86         | 6.69         | 4.26           |
| European descent/White          | 72.87           | 79.87      | 79.74             | 79.25           | 91.31        | 89.29        | 92.29          |
| Latin/South American            | 1.30            | 1.56       | 1.19              | 1.46            | 1.51         | 2.29         | 1.06           |
| Other or mixed                  | 1.06            | .          | .                 | .               | 1.26         | .57          | 1.33           |

N=sample size (using complete responses only); Periods (.) indicate a variable/option was not available in the survey. In the table, category sums (e.g., combining all age groups) within a sample can deviate by up to 0.01% from 100% due to rounding

aCensus asks participants about sex assigned at birth (male/female). The iCARE question varied across samples in asking about sex (male/female/other) or gender (man/woman/other)

bThe web panels only recruited people 18 years or above

cIn the Canadian census, this is the language most spoken at home. In iCARE samples, it is the chosen language in which to answer the survey

dVariable was only collected in the web panel survey, and not in the convenience sample

eThe web panel used a close-ended question similar to the Canadian census, whereas the convenience sample used an open-ended question. See supplemental materials for more details
overrepresentation of individuals that were more educated, English-speaking, and of European descent or White. As expected, discrepancies between the census and the convenience samples were considerably larger. The convenience samples consistently and strongly overrepresented participants from Quebec, that were female, spoke French, were highly educated, and were of European descent or White.

Evaluating overall bias on each outcome

Figure 3 presents a forest plot of our inferential results (i.e., unadjusted regressions), evaluating the overall discrepancy between the convenience and web-panel surveys on each outcome. Overall, 24 of 33 tests (73%) indicated significant discrepancies between the samples. Several outcomes were consistent in the direction of these discrepancies over time, with participants in the convenience sample reporting prevention measures as more important, being less concerned about the economy and their personal livelihood, being more likely to self-quarantine, and having higher intentions to get vaccinated. Other variables shifted in the direction of the discrepancy across time points—e.g., participants in the convenience sample reported wearing masks at a lower frequency at Time 1, but at a higher frequency at times 2 and 3. Section 6 of the supplemental file presents the distribution of responses for each outcome and can be used to contextualize effects from Fig. 3.

How frequently did covariates reduce sampling discrepancies?

Figures 4 and 5 summarize our specification curve analyses and display how discrepancies between the convenience and web panel surveys varied as a function of 512 combinations of covariates. Each plot (i.e., panel) within Figs. 4 and 5 indicates findings for one outcome at a given time point. Each plot also indicates the percent of adjusted models (those that control for covariates) that found smaller estimated discrepancies (i.e., our index of sampling bias) relative to their corresponding unadjusted models. Overall, adjusted models reduced sampling discrepancies 55% of the time, and increased discrepancies 45% of the time. However, there was substantial variation across outcomes. We organize these into three patterns (denoted by circled numbers in Figs. 4 and 5).

Pattern 1 For 33% of cases (i.e., 11 of the 33 panels across Figs. 4 and 5, with each panel indicating a particular outcome at a given time point), fewer than 25% of adjusted models showed smaller sampling discrepancies relative to their unadjusted counterparts. This pattern was most apparent for hand washing across all three time points (Fig. 4). For these cases, a large majority of covariate combinations increased sampling discrepancies (i.e., bias), frequently leading what was initially a non-significant discrepancy (in unadjusted models) to become significant.

Pattern 2 For 39% of cases (three panels in Fig. 4 and nine panels in Fig. 5), between 25–75% of adjusted models showed reduced sampling discrepancies relative to their unadjusted counterparts. This was especially apparent for vaccine intentions across all three time points (Fig. 5). For these, the inclusion of covariates could frequently reduce or increase sampling discrepancies, but often made little difference in in changing the significance level from that observed in the unadjusted models (e.g., the convenience sample displayed substantially higher vaccine intentions than the web-panel sample regardless of which covariates were adjusted for).

Pattern 3 Finally, for only 27% of cases (two panels in Fig. 4 and eight panels in Fig. 5) did 75% or more of adjusted models lead to reduced sampling discrepancies relative to their unadjusted counterparts. This applied to avoiding social gatherings across all three time points (Fig. 5). For these, covariates could reduce an initially significant discrepancy (in unadjusted models) to nonsignificance, but discrepancies also frequently persisted across many combinations of covariates.

Were there covariates that consistently reduced discrepancies?

In addition to Figs. 4 and 5, our supplemental file (Section 7) provides plots that depict which covariates were adjusted for in any given model. Our project page complements this with tables of results for all 16,896 models computed. Using these, we examined the consistency with which each covariate decreased/increased discrepancies.

When each covariate was adjusted for in isolation, income reduced discrepancies in 73% of cases, but increased discrepancies for the remaining 27%. Other covariates increased estimated discrepancies for 45% (province and education), 48% (ethnicity and sex), 55% (age), 64% (employment status), 67% (parental status), and 73% (student status) of cases. If we consider any combination that includes a given covariate (e.g., income by itself or with any combination of other covariates), each covariate increased discrepancies in between 40–45% of cases. When all nine covariates were adjusted for simultaneously, this performed better, but still increased discrepancies in 33% of cases.

Notably, a given covariate could have drastic and inconsistent effects on estimates across outcomes and time points. For example, if we examine concerns about the economic impact of the pandemic at Time 2 (Fig. 5),
we see a sudden shift in the plot such that half of models showed substantially larger negative estimates than the other half. This is almost entirely attributable to province being adjusted for: the smaller (less negative) half of estimates adjusted for province, whereas the larger (more negative) half did not. Now, consider hand washing at Time 2 (Fig. 4). Here, the reverse pattern occurred: the larger (more negative) half of estimates adjusted for

| Outcome/Time      | N  | Unstandardized Regression Coefficients (and 99% CI) | Est.  | [99% CI] | d   | R²  |
|-------------------|----|----------------------------------------------------|-------|----------|-----|-----|
| Perceived Importance |    |                                                    |       |          |     |     |
| T1                | 6023 |                                                    | 0.055 | [0.031, 0.079] | 0.151 | 0.006 |
| T2                | 3776 |                                                    | 0.076 | [0.016, 0.135] | 0.131 | 0.003 |
| T3                | 3528 |                                                    | 0.068 | [-0.005, 0.141] | 0.112 | 0.002 |
| C: Self Infected  |    |                                                    |       |          |     |     |
| T1                | 5603 |                                                    | -0.112 | [-0.169, -0.055] | -0.136 | 0.005 |
| T2                | 3685 |                                                    | 0.047 | [-0.051, 0.146] | 0.052 | 0.000 |
| T3                | 3446 |                                                    | -0.004 | [-0.117, 0.109] | -0.005 | 0.000 |
| C: Others Infected|    |                                                    |       |          |     |     |
| T1                | 5599 |                                                    | -0.036 | [-0.088, -0.015] | -0.049 | 0.001 |
| T2                | 3692 |                                                    | 0.292 | [0.208, 0.376] | 0.374 | 0.021 |
| T3                | 3455 |                                                    | 0.156 | [0.058, 0.253] | 0.201 | 0.005 |
| C: Economy        |    |                                                    |       |          |     |     |
| T1                | 5499 |                                                    | -0.431 | [-0.498, -0.364] | -0.454 | 0.047 |
| T2                | 3623 |                                                    | 0.198 | [-0.307, -0.090] | -0.301 | 0.006 |
| T3                | 3282 |                                                    | -0.258 | [-0.388, -0.128] | -0.306 | 0.008 |
| C: Personal Livelihood |    |                                                    |       |          |     |     |
| T1                | 5499 |                                                    | -0.431 | [-0.498, -0.364] | -0.447 | 0.047 |
| T2                | 3623 |                                                    | 0.198 | [-0.307, -0.090] | -0.201 | 0.006 |
| T3                | 3282 |                                                    | -0.258 | [-0.388, -0.128] | -0.265 | 0.008 |
| B: Hand Washing   |    |                                                    |       |          |     |     |
| T1                | 5916 |                                                    | -0.224 | [-0.258, -0.190] | -0.440 | 0.046 |
| T2                | 3734 |                                                    | 0.067 | [-0.120, 0.014] | -0.133 | 0.003 |
| T3                | 3512 |                                                    | 0.014 | [-0.042, 0.069] | 0.030 | 0.000 |
| B: Mask Wearing   |    |                                                    |       |          |     |     |
| T1                | 5777 |                                                    | 0.515 | [-0.590, 0.440] | -0.464 | 0.051 |
| T2                | 3685 |                                                    | 0.185 | [0.069, 0.301] | 0.168 | 0.005 |
| T3                | 3493 |                                                    | 0.094 | [0.039, 0.148] | 0.210 | 0.006 |
| B: Social distancing |    |                                                    |       |          |     |     |
| T1                | 5884 |                                                    | 0.083 | [0.059, 0.107] | 0.234 | 0.014 |
| T2                | 3725 |                                                    | 0.038 | [-0.017, 0.093] | 0.073 | 0.001 |
| T3                | 3496 |                                                    | 0.005 | [-0.060, 0.070] | 0.009 | 0.000 |
| B: Self-Quarantining |    |                                                    |       |          |     |     |
| T1                | 3413 |                                                    | 0.058 | [-0.005, 0.121] | 0.083 | 0.002 |
| T2                | 2044 |                                                    | 0.119 | [0.026, 0.212] | 0.212 | 0.005 |
| T3                | 1851 |                                                    | 0.078 | [-0.046, 0.202] | 0.116 | 0.001 |
| B: Avoid Gatherings |    |                                                    |       |          |     |     |
| T1                | 5832 |                                                    | 0.065 | [0.043, 0.088] | 0.193 | 0.009 |
| T2                | 3630 |                                                    | -0.111 | [-0.192, -0.030] | -0.146 | 0.003 |
| T3                | 3415 |                                                    | 0.128 | [0.052, 0.204] | 0.208 | 0.006 |
| Vaccine Intentions |    |                                                    |       |          |     |     |
| T1                | 5465 |                                                    | 0.194 | [0.139, 0.249] | 0.246 | 0.015 |
| T2                | 3552 |                                                    | 0.332 | [0.227, 0.437] | 0.345 | 0.018 |
| T3                | 3316 |                                                    | 0.370 | [0.240, 0.500] | 0.365 | 0.016 |

Fig. 3 Inferential results (unadjusted regression models) evaluating sampling discrepancies between the convenience and web-panel surveys on each outcome (reference group is the web panel). N = Sample size; Est = unstandardized estimate; CI = confidence interval; d = Cohen’s d; R² = R² coefficient of determination; T1 = Time 1; T2 = Time 2; T3 = Time 3; C = concerns; B = Behaviour. Plot created using the forestplot package in R [72].
province, whereas the smaller (less negative) half did not. Importantly, we also see important shifts within outcomes. For instance, when examining concerns about oneself being infected at Time 1, close to half of estimates are non-significant and close to half are significant and negative. The former generally adjust for province and the latter do not. The reverse is generally true at Time 2: the significant models adjust for province and non-significant models do not.

Fig. 4 Ordered caterpillar plots summarizing specification curve analyses. Plots were created using the specr package in R [54]. Instructions for reading the plots are provided at the bottom.
Discussion

In this work, we sought to: (1) better understand the effects of sampling bias in online COVID-19 research; and (2) examine the degree to which adjusting analyses for demographic covariates can successfully attenuate such bias. What did we find?

Convenience participants were more favourably disposed towards engaging in COVID-19 prevention behaviours

Significant discrepancies emerged between the online convenience and web-panel surveys on over two thirds of outcomes (averaging $d = 0.21$). For example, vaccine intentions were considerably higher in the convenience sample at all three time points relative to the web-panel, with 13–18% more participants indicating they would be “extremely likely” to get the vaccine. Such discrepancies are of an

Fig. 5 Ordered caterpillar plots summarizing specification curve analyses (continued). Plots were created using the `specr` package in R [54]. Figure 4 provides instructions for reading the plots
important magnitude and are larger than many effects listed as take-away messages from studies using convenience samples (e.g., difference in intentions between subgroups [56, 57]). This highlights the importance of taking care not to overgeneralize when using convenience samples and provides valuable information on how researchers can restrain their inferences (e.g., by recognizing that convenience-sample-based estimates of vaccine intentions could be inflated).

Documenting these descriptive patterns is useful, but how do we make sense of them? At the onset of this project, we reasoned that participants recruited using convenience-based methods would show more positive dispositions towards COVID-19 prevention measures than would participants recruited using more representative means. Indeed, participants in the convenience samples rated prevention measures as more important, engaged in more social distancing, self-quarantining, and avoidance of gatherings, and displayed stronger intentions to get vaccinated. These effects all align well with our hypotheses and the notion that individuals engage in selective exposure when deciding which studies to engage in (e.g., volunteering for topics they approve of [17]). This would suggest that to better reduce sampling bias, studies should assess and account for these associations. This could be by assessing people’s attitudes and adjusting for them statistically, or altering study designs to disrupt selective-exposure effects, such as by having the key topic of a study be less obvious during marketing. In making such choices, researchers should also consider carefully which variables act as mediators and confounds of the link between prevention behaviours and study participation. For example, participants who engage more frequently in prevention behaviours may be healthier and in a better position to engage in research—see the healthy user bias [58].

In contrast to our findings on behavioural outcomes, sampling discrepancies for participants’ concerns towards the pandemic were more varied. Participants in the convenience sample endorsed higher concerns for others being infected (in line with expectations) but fewer concerns about the economy and their personal livelihoods. These latter findings were unexpected, but could have arisen due to unmeasured confounds. For example, those with a neurotic personality may experience greater concerns but be less disposed to engage in research [26]. Additionally, affluent (e.g., White, educated) participants are also more likely to participate in research (as evidenced in our samples), but should generally be less concerned about their finances/livelihood. These possibilities highlight the complex nature of sampling bias, and further emphasize the need for research to think more carefully about confounds, mediators, and colliders when adjusting for sampling bias (i.e., Fig. 1, Panel C).

The performance of demographic covariates in attenuating sampling discrepancies was often poor and variable

The use of demographic covariates in analytic models is a common technique to account for sampling bias. However, across nearly 17,000 models, we found that the inclusion of demographic covariates reduced sampling discrepancies only 55% of the time—barely above chance level. Further, no individual covariate (used either in isolation or in combination with others) consistently reduced discrepancies. In fact, the effects of covariates were highly variable even within outcomes and there were many cases (e.g., vaccine intentions) for which no combination of covariates was sufficient to meaningfully attenuate sampling discrepancies. Certain demographic covariates even increased sampling bias in a systematic way (e.g., student status substantially more frequently increased than decreased sampling discrepancies).

These findings suggest that consistently following rules of thumb for covariate selection (e.g., always adjusting for sex or age) or simply including a subset of demographic characteristics that happen to be measured in a study are likely unreliable strategies for reducing sampling bias. General caution, along with a critical outlook, is therefore advised when using demographic variables as covariate variables.

That said, we do not suggest that efforts to reliably adjust for sampling bias using covariates is a profitless endeavour. Indeed, although we found that including all nine covariates increased sampling discrepancies 33% of the time, this was a better performance than most models adjusting for fewer covariates. Consequently, it is possible that adjusting for demographics could become more successful when a very large number of covariates are included in models. Future research could examine this possibility, along with whether modelling higher order effects (e.g., interaction terms between covariates) could also help attenuate sampling bias. It will also be important for research to examine the degree to which the patterns we report vary when using other types of sampling methods (e.g., in-person recruitment methods) as sampling methods may often interact in unique ways with participant characteristics (e.g., online studies may underrepresent those with less technological expertise, whereas in-person studies may underrepresent individuals with reduced physical mobility). While such studies are underway, researchers can consider several other tools at their disposal to deal with sampling bias.

Recommendations for dealing with sampling bias

One way to reduce sampling bias is through design-based methods. One may, for instance, use probability-based sampling to improve reach within a population. However, as noted in our introduction, such methods are not always
feasible or optimal (e.g., some populations are better reached through non-probability methods [59, 60]), and certain research goals can supersede the need for representativeness (e.g., a researcher may choose purposive sampling when the goal is maximizing diversity of views/experiences) [20, 21]. Other tools may include reducing selective participation through stronger monetary incentives or by mandating participation [13, 14, 61], but both methods can also have barriers and drawbacks to consider [62, 63].

On the analytic side, causal diagrams (e.g., Fig. 1) are a tool that have, over the last few years, emerged as a gold-standard for understanding and determining how to best analytically handle bias in research (including sampling bias) [4, 5, 11, 12, 19]. Importantly, causal diagrams can help researchers pinpoint which covariates can help maximize the validity of inferences, while also helping better plan studies in the design phase. An important insight from the use of causal diagrams is that there is likely no single “correct” set of covariates that can be used across all analyses. Each outcome (and outcome-exposure link) should have its own covariates (and causal diagram) to avoid introducing error and bias (e.g., see discussions on the Table 2 Fallacy, unnecessary adjustment, and overadjustment) [64–66]. To this, our findings further suggest that analysts may also wish to explicitly account for time-specific influences—as we found the role of covariates to differ substantially in their effects across time points even for the same outcome. Adding this type of specificity to causal diagrams could help researchers further reduce the effects of sampling bias.

Unfortunately, in many research areas (e.g., in medical and behavioural sciences), it is often difficult for theories to outline causal factors in enough details to delineate complete causal diagrams. In such cases, researchers can consider a final option; that is, examining the robustness of their findings using multiverse-type analyses—as demonstrated within the current works—as a form of extended sensitivity analyses. Multiverse analyses are not only explicitly designed to help researchers handle and understand ambiguities in analysis-based decisions, but the development of new multiverse-type tools/perspectives continues to be an area of burgeoning methodological advancements, and many resources now exist for interested readers to learn more about these approaches [36, 37, 48, 67]. However, in relying on multiverse analyses, it will be important to remember that compared to causal diagrams, multiverse analyses cannot inform which estimate is the most causally valid. Rather, this approach is used to verify that one’s inferences are not limited to only a subset of possible analyses, and to quantify the degree to which largely arbitrary choices (between plausible alternatives) influence inferences.

**Strengths and limitations**

There are a few constraints that warrant consideration when interpreting our findings. First, our study was conducted in a very specific context: Canada during the COVID-19 pandemic. Examining sampling bias in other countries and contexts is therefore warranted. Second, although Léger constructs their web panels using methods such as random-digit dialling, the samples we obtained from these panels were not fully representative of the Canadian population and this could have skewed findings—e.g., if similar but less pronounced biases (as observed in the convenience samples) affected the web panels, our results may generally underestimate bias. Third, we acknowledge that many methods exist to obtain convenience samples and that our analyses were specific to volunteer-based online recruitment methods. Other methods (e.g., in-person, or print-based recruitment techniques) can have idiosyncratic biases [68] such that specific variables (e.g., sex, age, health beliefs) may vary in how they operate to generate (and reduce) bias. Future works will need to parse out such patterns. Fourth, our data was cross-sectional and our findings ultimately still conditioned on self-selection into the study. Consequently, care should be taken when inferring causation; for instance, we cannot infer that vaccine intentions cause self-selection into studies (e.g., as in Fig. 1’s path \( p_3 \)), nor can we infer that the effects of adjusting for covariates operated through causal links. That said, our unadjusted models can still provide good estimates of sampling bias if sampling bias is taken to be entirely spurious associations between sampling and outcomes (akin to path \( p_3 \) in Fig. 1). Lastly, our multiverse analyses treated getting an accurate estimate from a convenience sample (one equal in magnitude/direction to an estimate from a representative sample) as the goal when reducing sampling bias. This was a simplification. Although removing sampling bias would achieve this, so would aggregating divergent biases that so happen to average to the population value. Our analyses cannot tease these scenarios apart.

Finally, our study also has several strengths to consider. Notably, this is the first empirical study to use multiverse style analyses to understand how covariate selection influences estimates produced across sampling methods. Our analyses were also registered a priori and we used large samples across three distinct time points. This contrasts with previous empirical works on sampling bias, which have not been registered, have relied on smaller samples collected over single time points, and have usually examined the influence of a single set of covariates at a time [15, 16, 22, 26, 69–71]. Consequently, our findings are more likely to generalize than past efforts.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1007/s10654-022-00932-y.
University-Hadassah School of Public Health), Ruth Snir (Hebrew University-Hadassah School of Public Health), Yifat Uri (Hebrew University-Hadassah School of Public Health), Mohsen Alyami (University of Auckland), Comfort Sanuade, Olivia Crescenzi, Kyle War- 
kentin, Katya Grinko, Lalita Angne, Jigisha Jain, Nikita Mathur (Syn- 
corp Clinical Research), Anagha Mithe, Sarah Nathan (Community 
Empowerment Lab).

Author contributions The current project was conceptualized by KJD, 
JS, VGB, EBK, and JE. KJD prepared data for analyses, along with 
support from JS and VGB. KJD conducted all analyses, and all authors 
contributed to data interpretation. KJD prepared the first draft. All 
authors contributed to revisions and approved the final draft. Beyond 
the author team, we thank Ms. Mariam Atoui and Ruth Bruno for their 
assistance with data preparation.

Funding This work was supported by the Canadian Institutes of Health 
Research (CIHR: MM1-174903; MS3-173099; SMC-151518); the 
Canada Research Chairs Program (950-232522, Chair holder: Dr. 
Kim L. Lavoie); the Fonds de recherche du Québec—santé (FRQ-S: 
251618; 34757); the Fonds de recherche du Québec – Société et cul-
ture (FRQSC: 2019-SE1-252541); and the Ministère de l’Économie et 
de l’Innovation du Québec (2020-2022-COVID-19-PSOv2a-51754). 
Study sponsors had no role in conducting the research.

Data availability & study materials A project page is available through 
the Open Science Framework (https://osf.io/dp9kq/), which provides 
open access to our supplemental materials, data analysis script files, 
and registration. The data underlying this article can be made available 
by the Montreal Behavioural Medicine Centre upon reasonable request 
(https://mbmc-cmcm.ca/covid19/apl/).

Declarations

Ethical approval The iCARE study was approved by the Comité 
d’étique de recherche du Centre intégré universitaire de santé et 
des services sociaux du Nord-de-l’île-de-Montréal, approval: 2020-
2099/25-03-2020.

Consent to participate Informed consent was obtained from all indi-
vidual participants included in the study.

Competing interests Kim Lavoie has served on the advisory board for 
Schering-Plough, Takeda, AbbVie, Almirall, Janssen, GSK, Novartis, 
Boehringer Ingelheim (B1), and Sojecci Inc, and has received sponsor-
ship for investigator-generated research grants from GlaxoSmithKline 
(GSK) and AbbVie, speaker fees from GSK, Astra-Zeneca, Astellas, 
Novartis, Takeda, AbbVie, Merck, Boehringer Ingelheim, Bayer, Pfizer, 
Xfaco, and Air Liquide, and support for educational materials from 
Merck. Urška Košir has received speaker fees from Merck. None of 
these engagements are related to the current article.

References

1. Tyrer S, Heyman B. Sampling in epidemiological research: issues, 
hazards and pitfalls. BJPsych Bulletin. 2016;40:57–60. https://doi. 
org/10.1192/bp.bp.114.050203.
2. Sarsstedt M, Bengart P, Shaltoni AM, Lehmann S. The use of sam-
pling methods in advertising research: A gap between theory and 
practice. Int J Advert. 2018;37:650–63. https://doi.org/10.1080/ 
02650487.2017.1348329.

3. Kennedy EB, Jensen EA, Jensen AM. Methodological considera-
tions for survey-based research during emergencies and public 
health crises: Improving the quality of evidence & science com-
munication. Front Commun. 2021;6:226.

4. Elwert F, Winship C. Endogenous selection bias: the problem of 
conditioning on a collider variable. Ann Rev Sociol. 2014;40:31– 
53. https://doi.org/10.1146/annurev-soc-071913-043455.

5. Griffith GJ, Morris TT, Tidball MJ, et al. Collider bias under-
mines our understanding of COVID-19 disease risk and severity. 
Nat Commun. 2020;11:1–12. https://doi.org/10.1038/
s41467-020-19478-2.

6. Shen C, VanGennep D, Siegenfeld AF, Bar-Yam Y. Unraveling 
the flaws of estimates of the infection fatality rate for COVID-19. 
J Travel Med. 2021;28:1–3. https://doi.org/10.1093/jtm/taaa239.

7. Zhao Q, Ju N, Bacallado S, Shah RD. BETS: The dangers of selec-
tion bias in early analyses of the coronavirus disease (COVID-
19) pandemic. Ann Appl Stat. 2021;15:363–90. https://doi.org/ 
10.1214/20-AOAS1401.

8. Steiner PM, Cook TD, Shadish WR, Clark MH. The importance of 
covariate selection in controlling for selection bias in observa-
tional studies. Psychol Methods. 2010;15:250. https://doi.org/10. 
1037/a0018719.

9. Starks H, Diehr P, Curtis JR. The challenge of selection bias 
and confounding in palliative care research. J Palliat Med. 
2009;12:181–7. https://doi.org/10.1089/jpm.2009.0672.

10. Wirth KE, Tchetgen EJT. Accounting for selection bias in 
association studies with complex survey data. Epidemiology. 
2014;25:444. https://doi.org/10.1097/EDE.0000000000000371.

11. Greenland S, Pearl J, Robins JM. Causal diagrams for epide-
miologic research. Epidemiology. 1999;10:37–48.

12. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach 
to selection bias. Epidemiology. 2004;15:615–25. https://doi.org/
10.1097/01.ede.0000135174.63482.43.

13. Smith MG, Witte M, Rocha S, Basner M. Effectiveness of incen-
tives and follow-up on increasing survey response rates and partic-
ipation in field studies. BMC Med Res Methodol. 2019;19:1–13. 
https://doi.org/10.1186/s12874-019-0868-8.

14. Barón JD, Breunig RV, Cobb-Clark DA, Gorgens T, Sarthayeva 
A. Does the effect of incentive payments on survey response rates 
differ by income support history? J Off Stat. 2009;25:483–507.

15. Ganguli M, Lytle ME, Reynolds MD, Dodge HH. Random versus 
volunteer selection for a community-based study. J Gerontol Ser 
A: Biol Sci Med Sci. 1998;53:M39–46. https://doi.org/10.1093/
aje/53a.1.m39.

16. Owen JE, Bantum EOC, Criswell K, Bazzio J, Gorlick A, Stan-
ton AL. Representativeness of two sampling procedures for an 
internet intervention targeting cancer-related distress: a com-
parison of convenience and registry samples. J Behav Med. 
2014;37:630–41. https://doi.org/10.1007/s10865-013-9509-6.

17. Hart W, Albarracin D, Eagly AH, Brechan I, Lindberg MJ, Mer-
rell L. Feeling validated versus being correct: a meta-analysis of 
selective exposure to information. Psychol Bull. 2009;135:555– 
88. https://doi.org/10.1037/a0015701.

18. Meppelink CS, Smit EG, Fransen ML, Diviani N. “I was right about 
vaccination”: confirmation bias and health liter-
acy in online health information seeking. J Health Commun. 
2019;24:129–40. https://doi.org/10.1080/10810730.2019.15837 
01.

19. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due 
to conditioning on a collider. Int J Epidemiol. 2010;39:417–20. 
https://doi.org/10.1093/ije/dyp334.

20. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness 
should be avoided. Int J Epidemiol. 2013;42:1012–4. https://doi.
org/10.1093/ije/dys223.
How well do covariates perform when adjusting for sampling bias in online COVID-19 research?...
multivariable methods. J Clin Epidemiol. 2006;59:437.e1-e24. https://doi.org/10.1016/j.jclinepi.2005.07.004.

53. R Core Team. R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2017. https://www.R-project.org/

54. Masur PK, Scharkow M. Specr: conducting and visualizing specification curve analyses R Package. 2020. https://research.vu.nl/en/publications/specr-conducting-and-visualizing-specification-curve-analyses-r-p

55. Gassen J. Rdfanalysis: researcher degrees of freedom analysis, a package to explore and document your degrees of freedom. 2020. https://joachim-gassen.github.io/rdfanalysis/

56. Ali KF, Whitebridge S, Jamal MH, Alsafy M, Atkin SL. Perceptions, knowledge, and behaviors related to COVID-19 among social media users: Cross-sectional study. J Med Internet Res. 2020;22:e19913. https://doi.org/10.2196/19913.

57. Yahia AIO, Alshahrani AM, Alsulmi WG, et al. Determinants of COVID-19 vaccine acceptance and hesitancy: a cross-sectional study in Saudi Arabia. Hum Vaccines Immunother. 2021. https://doi.org/10.1080/21645515.2021.1950506.

58. Shrank WH, Patrick AR, Brookhart MA. Healthy user and related biases in observational studies of preventive interventions: a primer for physicians. J Gen Intern Med. 2011;26:546–50. https://doi.org/10.1007/s11606-010-1609-1.

59. Hequembourg AL, Panagakis C. Maximizing respondent-driven sampling field procedures in the recruitment of sexual minorities for health research. SAGE Open Med. 2019;7:2050312119829983.

60. Piperato SM. Comparative Effectiveness of Conventional and Novel Sampling Methods for the Recruitment of Sexual Minority Identified Women. (Doctoral dissertation). 2018. Retrieved from https://scholarcommons.sc.edu/etd/4800

61. Cheung KL, ten Klooster PM, Smit C, de Vries H, Pieterse ME. The impact of non-response bias due to sampling in public health studies: a comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. BMC Public Health. 2017;17:276. https://doi.org/10.1186/s12889-017-4189-8.

62. Gelinas L, Largent EA, Cohen IG, Kornetsky S, Bierer BE, Fernandez LH. A Framework for Ethical Payment to Research Participants. 2018;378:766–71. https://doi.org/10.1056/NEJMsb1710591.

63. Rothstein MA, Shoben AB. Does consent bias research? Am J Bioeth. 2013;13:27–37. https://doi.org/10.1080/15265161.2013.767955.

64. VanderWeele TJ. On the relative nature of overadjustment and unnecessary adjustment. Epidemiology. 2009;20:496–9. https://doi.org/10.1097/EDE.0b013e3181a8f12.

65. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology. 2009;20:488–95. https://doi.org/10.1097/EDE.0b013e3181a19a1.

66. Westreich D, Greenland S. The Table 2 Fallacy: Presenting and interpreting confounder and modifier coefficients. Am J Epidemiol. 2013;177:292–8. https://doi.org/10.1093/aje/kws412.

67. Patel CJ, Burford B, Ioannidis JP. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J Clin Epidemiol. 2015;68:1046–58. https://doi.org/10.1016/j.jclinepi.2015.05.029.

68. Morley KC, Teesson M, Sannibale C, Haber PS. Sample bias from different recruitment strategies in a randomised controlled trial for alcohol dependence. Drug Alcohol Rev. 2009;28:222–9. https://doi.org/10.1111/j.1465-3362.2008.00022.x.

69. Arnett B, Rikli R. Effects of method of subject selection (volunteer vs random) and treatment variable on motor performance. Res Q Exerc Sport. 1981;52:433–40. https://doi.org/10.1080/02701367.1981.10607888.

70. Dollinger SJ, Leong FTL. Volunteer bias and the five-factor model. J Psychol. 1993;127:29–36. https://doi.org/10.1080/00223980.1993.9915540.

71. Prucho RA, Brill JE, Shands Y, et al. Convenience samples and caregiving research: how generalizable are the findings? Gerontologist. 2008;48:820–7. https://doi.org/10.1093/geront/48.6.820.

72. Gordon M, Lumley T. Advanced Forest Plot Using ‘grid’ Graphics. 2021. https://cran.r-project.org/web/packages/forestplot/forestplot.pdf

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.