Estimates of COVID-19 Cases across Four Canadian Provinces

DAVID BENATIA
Center for Research in Economics and Statistics, École nationale de la statistique et de l’administration économique, Institut Polytechnique de Paris, Palaiseau, France

RAPHAEL GODEFROY
Department of Economics, Université de Montréal, Montreal, Quebec

JOSHUA LEWIS
Department of Economics, Université de Montréal, Montreal, Quebec

Les auteurs estiment les taux d’infection par le coronavirus 2019 (COVID-19) de la population de quatre provinces canadiennes, de la fin de mars au début de mai 2020. Dans leur analyse, ils associent les données quotidiennes relatives au nombre de tests effectués et au nombre de cas diagnostiqués au moyen d’une méthodologie grâce à laquelle les données sont corrigées pour tenir compte du caractère non aléatoire des tests. Ils estiment la relation entre l’évolution quotidienne du nombre de tests effectués et de la proportion de cas positifs dans l’échantillon non aléatoire (généralement moins de 1 pour cent de la population) et utilisent ce gradient pour extrapoler une prédiction quant à la proportion de cas positifs qui seraient diagnostiqués si les tests étaient administrés à la population entière. Au cours de la période d’échantillonnage, les taux estimatifs d’infection de la population se situent dans les intervalles suivants : 1,7 à 2,6 pour cent au Québec, 0,7 à 1,4 pour cent en Ontario, 0,5 à 1,2 pour cent en Alberta et 0,2 à 0,4 pour cent en Colombie-Britannique. Dans chaque province, ces estimations se révèlent largement inférieures au taux moyen de cas positifs, ce qui s’explique par l’administration non aléatoire des tests aux populations qui présentent un risque plus élevé. Ces résultats indiquent également que les cas de COVID-19 non diagnostiqués sont largement répandus. Pour chaque cas diagnostiqué jusqu’à la mi-avril, les auteurs estiment à environ 12 le nombre de cas d’infection dans la population.

Mots clés : COVID-19, taux d’infection de la population, tests non aléatoires

This article estimates population infection rates from coronavirus disease 2019 (COVID-19) across four Canadian provinces from late March to early May 2020. The analysis combines daily data on the number of conducted tests and diagnosed cases with a methodology that corrects for non-random testing. We estimate the relationship between daily changes in the number of conducted tests and the fraction of positive cases in the non-random sample (typically less than 1 percent of the population) and apply this gradient to extrapolate the predicted fraction of positive cases if testing were expanded to the entire population. Over the sample period, the estimated population infection rates were 1.7–2.6 percent in Quebec, 0.7–1.4 percent in Ontario, 0.5–1.2 percent in Alberta, and 0.2–0.4 percent in British Columbia. In each province, these estimates are substantially below the average positive case rate, consistent with non-random testing of higher-risk populations. The results also imply widespread undiagnosed COVID-19 infection. For each identified case by mid-April, we estimate there were roughly 12 population infections.

Keywords: COVID-19, non-random testing, population infection rates

Introduction
The first cases of coronavirus disease 2019 (COVID-19) in Canada were documented in late January 2020, and by 5 May more than 63,000 cases had been reported (CSSE 2020). Because testing has been limited to a small fraction of the population and infected individuals with mild or
no symptoms may not seek testing, however, the potential exists for widespread undocumented infections (see Bai et al. 2020; Dong et al. 2020; Hoehl et al. 2020; Lu et al. 2020; Pan et al. 2020).

In this article, we estimate population infection rates for COVID-19 across four Canadian provinces—Quebec, Ontario, Alberta, and British Columbia—from late March to early May. The analysis is based on the methodology developed in Benatia, Godefroy, and Lewis (2020) that corrects observed infection rates among tested individuals for non-random sampling to calculate infection rates in the overall population.1 To implement the procedure, we estimate the relationship between the number of tests and the share of positive tests. This gradient is informative for the severity of selection bias. For example, a negative slope indicates positive selection bias, because individuals who are most frequently tested have the highest probability of infection. If the functional form of this relationship can be consistently estimated, we can compute the population infection rate as a combination of the observed sample infection rate and the estimated selection gradient, which corrects for non-random testing.

In practice, our approach faces two main empirical challenges. First, there is concern that the supply of testing may respond endogenously to underlying disease prevalence. For example, if policy-makers expand testing in response to increases in underlying disease prevalence, our estimation strategy would underestimate the selection bias gradient and thus overestimate total population infections. To address this concern, we focus on high-frequency day-to-day changes in the number of completed tests across US states and all Canadian provinces.2 Because there is little scope for evolution in disease prevalence, individual trends in testing should be orthogonal to changes in population infection rates. To further validate this assumption, we estimate models that control for province and state fixed effects, thereby allowing for daily exponential growth in disease prevalence that is specific to each jurisdiction.

The second empirical challenge stems from uncertainty regarding the true functional relationship between the positive test rate and the size of the tested sample. In the empirical implementation, we specify a flexible functional relationship that appears to fit the data well. Nevertheless, the results ultimately depend on an untested assumption that the estimated relationship—based on data from the sample of tested individuals who typically make up less than 1 percent of the population—can be extrapolated to the rest of the population. Despite this limitation, we believe the approach offers significant advantages over existing methods used to estimate population infection rates (described later). In ongoing work, we hope to refine the estimation procedure to address this functional form concern.

We find wide cross-province differences in both the level of population infection rates and their trends. Average population infection rates over the sample period ranged from 0.3 percent in British Columbia to 3.0 percent in Quebec. Infection rates in British Columbia declined modestly over the sample period. In Ontario, infection rates rose from early to mid-April and subsequently declined. Meanwhile, Quebec and Alberta experienced increases in population infection rates over the month of April. These trends need not reflect increases in the number of newly infected individuals because our population infection rates capture both newly infected individuals and those with continued detectable viral load over the sample period.3

Our results also suggest widespread undetected COVID-19 infection across Canadian provinces. We calculate that for every diagnosed case, there were 12 population infections in mid-April. The ratio of population infection to diagnosis ranges from 8.6 in British Columbia to 14.8 in Ontario. These estimates are comparable to recent evidence on the rates of undetected infection in the United States and internationally (Ferguson et al. 2020; Johndrow, Lum, and Ball 2020; Perkins et al. 2020; Verity et al. 2020). For example, Aspelund et al. (2020) estimate that 80–90 percent of COVID-19 cases went undiagnosed in Iceland from late March to early April. Benatia et al. (2020) estimate a ratio of 12 population infections per diagnosed case in the United States by early April. More recently, the CDC (2020a) reported results based on sero-prevalence that suggest that total infections were 10 times higher than the number of confirmed cases.

This article provides new evidence on overall population infection rates for COVID-19 in Canada. Our findings complement evidence for COVID-19 prevalence nationwide. Using survey results for COVID-19 symptoms, Reid (2020) finds that more than 100,000 households reported COVID-like symptoms after adjusting for seasonal influenza rates. The results do not account for potentially large numbers of asymptomatic infections. Meanwhile, Verity et al. (2020) combines assumptions regarding the age-adjusted case fatality rate with COVID-related deaths to estimate total population infections in Canada on 31 March. These estimates indicate that the case detection rate was just 5 percent through March. Our analysis provides the first provincial-level estimates. Given wide cross-province differences in per capita testing, official case counts may mask important geographic differences in the severity of the outbreak. Indeed, whereas the official case count in Quebec was 55 percent higher than in Ontario, our results show that gap in total cases was less than 20 percent.

Our empirical framework complements existing methods used to estimate population infection rates in the United States and internationally (Ferguson et al. 2020; Javan, Fox, and Meyers 2020; Johndrow et al. 2020; Li, Pei, et al. 2020; Perkins et al. 2020; Riou, Hauser, Counotte, and Althaus 2020; Verity et al. 2020). One approach has been...
based on the Susceptible Infectious Removed epidemiological model, which calibrates parameters to the specific characteristics of the severe acute respiratory syndrome coronavirus 2 pandemic to estimate current and future infections. A challenge for this approach is the large uncertainty regarding the relevant parameter values for the virus and the fact that the parameter values will evolve as societies take different measures to reduce transmission. Other research has relied on Bayesian modelling to infer past disease prevalence from observed COVID-19 deaths. Although these models require fewer assumptions regarding the underlying parameter values, because they scale up observed deaths to estimate population infections, small differences in the assumed case fatality will have substantial effects on the estimates. Given considerable uncertainty regarding the true case fatality, which may depend on local sociodemographic and environmental conditions, and the fact that COVID-19–related deaths may be undercounted during the course of the pandemic, these estimates may not capture the overall extent of population infection (see Clay, Lewis, and Severini 2018, 2019; Clay et al. 2020; Han et al. 2020; Katz and Sanger-Katz 2020; Prakash and Hall 2020; Riou, Hauser, Counotte, Margossian, et al. 2020; Wu et al. 2020).

Most closely related to our article is Manski and Molinari (forthcoming), who use data on the total number of tests and the positive test rate to estimate ranges for population COVID-19 infection rates for Illinois, New York, and Italy in early April. Their approach requires only the imposition of weak monotonicity assumptions for identification. Their estimated bounds for infection rates are 0.1%–51.7% for Illinois, 0.8%–64.5% for New York, and 0.3%–51.0% for Italy. These are wide, model-free bounds. In this article, we estimate much narrower intervals, which are conditional on the accuracy of our model. Although we have developed our method for use during this crisis to employ the available information as fully as possible, policy-makers should be aware that bounds that include model uncertainty would be wider than ours by some unknown amount, which is not uncommon in econometric analyses. The model-free bounds of Manski and Molinari (forthcoming) serve as a reminder of that issue.

Data

Our analysis draws on daily data on total test results (positive plus negative) and positive tests across Canadian provinces and US states for the period 31 March–5 May. Provincial data were obtained from the Epidemiological Data from the COVID-19 Outbreak in Canada project (Berry et al. 2020). This project is conducted by a team of researchers from the Universities of Toronto and Guelph and provides information on cases and testing across provinces based on publicly available information from government reports and news media. We exclude days on which there were identified changes in provincial reporting standards and days on which provincial health authorities did not release information on completed tests. In addition, we use information on the number of positive tests by age group, which is available from provincial health departments, and provincial population estimates from Statistics Canada (2020). We supplement these data with information on total test results and positive cases across US states for the same time period from the COVID Tracking Project, a site launched by journalists from The Atlantic that publishes high-quality data on the outbreak across US states (Meyer, Kissane, and Madrigal 2020).

Figure 1 reports the daily tests and positive cases across the four provinces. Daily testing was fairly stable in Quebec throughout April. In contrast, daily testing substantially increased in both Ontario and Alberta and to a lesser extent in British Columbia.

Methodology

In this section, we present the theoretical framework developed in Benatia et al. (2020) to estimate COVID-19 prevalence. This framework motivates estimating Equation (6).

Theory

To evaluate population disease prevalence, we develop a simple selection model for COVID-19 testing and use the framework to link observed rates of positive tests to population disease prevalence. We consider a stable population, normalized to a population of one, and denote A and B as the number of sick and healthy individuals, respectively. Let \( p_s \) denote the probability that a sick person is tested and \( q_h \) the probability that a healthy person is tested, given a total number of tests, \( n \). Thus, we have \( n = p_s A + q_h B \), and assuming the test is accurate, the number of positive tests is \( s = p_s A \).

This simple framework highlights how non-random testing will bias estimates of the population disease prevalence. Using Bayes’s rule, we can write the relative probability of testing as

\[
\frac{p_n}{q_n} = \frac{Pr(sick|tested,n)}{Pr(healthy|tested,n)} / \frac{Pr(sick,n)}{Pr(healthy,n)}
\]

which is equal to one if tests are randomly allocated, \( Pr(sick|tested,n) = Pr(sick|n) \). When testing is targeted to individuals who are more likely to be sick, we have \( Pr(sick|tested,n) > Pr(sick|n) \), \( Pr(sick\backslash tested,n) < Pr(sick\backslash n) \), and \( Pr(healthy\backslash tested,n) < Pr(healthy\backslash n) \), so the ratio will be greater than one. In this scenario, the ratio of sick to healthy people in the sample, \( p_s A / q_h B \), will exceed the ratio in the overall population, \( A/B \).

We assume that the severity of selection bias can be expressed as a function of the number of tests,

\[
\frac{p_n}{q_n} = f(n; \theta),
\]

(1)
Figure 1: Daily Testing and New Cases across Provinces: (a) Quebec, (b) Ontario, (c) Alberta, and (d) British Columbia

Notes: This figure reports the total daily coronavirus tests and the number of new cases per 100,000 population by province. The trends are based on data from Berry et al. (2020). We exclude days for which there were identified changes in provincial reporting standards and days for which provincial health authorities did not release information on completed tests.

Source: Berry et al. (2020) and Statistics Canada (2020).

where \( n \) is number of conducted tests and \( \theta \) is a vector of parameters to be estimated.

According to this setup, we can write the fraction of positive tests, \( s/n \), as follows:

\[
\frac{s}{n} = \frac{1}{1 + \frac{q_n}{p_n} \frac{B}{A}}
\]

Equation (2) shows that the log share of positive tests in the sample can be approximated by the sum of the log ratio of the relative probability of testing, \( p_n/q_n \), and the unobserved log ratio of sick to healthy people in the population, \( A/B \).

\[
\log \left( \frac{s}{n} \right) \approx \log \left( \frac{p_n A}{q_n B} \right) = \log \left( \frac{p_n}{q_n} \right) + \log \left( \frac{A}{B} \right).
\]

Equation (3) shows that the log share of positive tests in the sample can be approximated by the sum of the log ratio of the relative probability of testing, \( p_n/q_n \), and the unobserved log ratio of sick to healthy people in the population, \( A/B \).

From Theory to Estimation

To conduct the estimation, we adopt a first difference estimator, using as the dependent variable the difference \( \log \left( \frac{s_{i,t}}{n_{i,t}} \right) - \log \left( \frac{s_{i,t-1}}{n_{i,t-1}} \right) \) on two consecutive days \( t-1 \) and \( t \) in a given province/state \( i \). Given the last equation, this first difference is equal to
log \( \frac{s_{i,t}}{n_{i,t}} - \log \frac{s_{i,t-1}}{n_{i,t-1}} = \log f\left(n_{i,t}; \theta \right) - \log f\left(n_{i,t-1}; \theta \right) + u_{i,t} \)  

where \( u_{i,t} = \log \left( \frac{A_{i,t}}{B_{i,t}} \right) - \log \left( \frac{A_{i,t-1}}{B_{i,t-1}} \right) + \epsilon_{i,t} \) is a mean zero error term that depends on the change in ratio of sick to healthy individuals in the population from \( t-1 \) to \( t \) and an idiosyncratic component, \( \epsilon_{i,t} \).

**Equation (4)** forms the basis of our empirical analysis. Our identifying assumption is strict exogeneity in the error term: \( E(u_{i,t} | n_{i,t}, n_{i,t-1}) \). This assumption ensures that the errors are uncorrelated with any function of changes in the number of tests, \( \Delta n_{i,t} \) and will be violated if changes in the population infection rate were systematically related to testing capacity. This assumption is supported by the short time interval in the daily first difference specification, which limits the scope for disease evolution. Also, in robustness tests, we control for province and state fixed effects, which allow for jurisdiction-specific exponential growth in underlying disease prevalence from one day to the next. These controls do not affect the main coefficient estimates.

Notice that by focusing on a daily first difference estimator, we are able to partial out the unobserved log ratio of sick to healthy people in the population, \( A_{i,t}/B_{i,t} \). As a result, changes in the share of positive tests depend on the number of tests only through a selection channel.

How does this assumption enable us to estimate population infection rates? Using day-to-day changes in the share of positive tests and day-to-day changes in the number of tests, we can recover \( \log \left( \frac{n_{i,t}}{n_{i,t-1}} \right) \) by estimating **Equation (4)**. This term captures how changes in the share of positive tests are predicted to change with \( n \). We can then use this prediction to recover population infection rates. Denote \( \hat{s}_{n_{i,t}} \) as the predicted fraction of positive tests if the entire population in province \( i \) was tested on date \( t \), that is, \( n_{i,t} = \text{pop}_i \). We can rewrite the first difference **Equation (4)** as

\[
\log \hat{s}_{\text{pop},i,t} = \log s_{i,t} + \log \frac{\text{pop}_i}{n_{i,t}} + \log f\left(\text{pop}_i; \hat{\theta} \right) - \log f\left(n_{i,t}; \hat{\theta} \right).
\]

That is, the predicted log fraction of positive tests in the population is equal to the log fraction of positive tests in the sample plus an adjustment factor that corrects for non-random testing. One could also view our exercise as a reduced form estimation of the relationship between the fraction of individuals who test positive and the size of the tested population (holding constant the population share of those who are sick). Once this relationship has been consistently estimated, we can predict the share of positive tests for any value of \( n \), including when \( n = \text{pop}_i \).

### Empirical Implementation

To implement the procedure described in the preceding section, we specify the following functional form for the selection process into testing, \( f(n; \theta) \):

\[
p_n = f\left(n; \theta \right) = 1 + \exp\left(\beta_n \gamma_n \right).
\]

The term \( \exp\left(\beta_n \gamma_n \right) \geq 0 \) reflects the fact that testing has been targeted toward higher-risk populations, with the intercept, \( \gamma_n \), capturing the severity of selection bias when testing is limited. Meanwhile, the coefficient \( \beta > 0 \) identifies how selection bias decreases with \( n \) as the ratio \( p_{n}/q_{n} \) approaches one. Intuitively, as testing expands, the sample will become more representative of the overall population, and the selection bias will diminish.

We substitute this function into the first difference regression model, taking a third-order power series approximation of the log function, which yields the following estimating equation:

\[
\left( \begin{array}{c}
\log \left( \frac{s_{i,t}}{n_{i,t}} \right) - \log \left( \frac{s_{i,t-1}}{n_{i,t-1}} \right) \\
\log \left( \frac{s_{i,t}}{n_{i,t}} \right) - \log \left( \frac{s_{i,t-1}}{n_{i,t-1}} \right)
\end{array} \right) = \alpha_1 \left( \begin{array}{c}
E_{i,t} \left( \frac{n_{i,t}}{n_{i,t-1}} \right) - E_{i,t} \left( \frac{n_{i,t}}{n_{i,t-1}} \right) + \alpha_2 \left( \begin{array}{c}
E_{i,t} \left( \frac{n_{i,t}}{n_{i,t-1}} \right) - E_{i,t} \left( \frac{n_{i,t}}{n_{i,t-1}} \right)
\end{array} \right)
\end{array} \right) + \alpha_3 \left( \begin{array}{c}
E_{i,t} \left( \frac{n_{i,t}}{n_{i,t-1}} \right) - E_{i,t} \left( \frac{n_{i,t}}{n_{i,t-1}} \right)
\end{array} \right) + \epsilon_{i,t}.
\]

We estimate **Equation (6)** by non-linear least squares, allowing for heteroskedastic errors. After estimation, we derive predicted values for population infection rates based on **Equation (5)**, using the delta method to construct confidence intervals.

Before turning to the main results, several caveats should be highlighted. First, the estimates of population infection rates depend on a correctly specified functional relationship between the positive test rate and the size of the tested sample. Although the model fits the data well (see Figure 2), an important assumption underlying our analysis is that this observed relationship in the tested sample—who typically make up less than 1 percent of the population—would continue to hold if testing were expanded out to the broader population. The accuracy of this extrapolation depends on a smoothness condition on the functional form and may be violated if, for example, some segments of the population can easily be tested and others cannot.

We also constrain the population coefficients \((\alpha, \beta)\) to be the same across jurisdictions. This assumption requires that decisions regarding how to prioritize tests were made similarly across provinces and US states. Although states had latitude to implement their own diagnostic testing procedures, the guidance laid out for testing prioritization by the CDC (2020b) was broadly similar to the policies implemented across Canadian provincial health departments. We also estimate the model for three distinct
Results

Table 1, Panel A, reports the estimates for Equation (1) across three time periods: 31 March–7 April, 14–21 April, and 28 April–5 May. We estimate the model separately for all ages (Columns [1], [3], and [5]) and excluding cases among individuals aged older than 70 years (Columns [2], [4], and [6]). Consistent with the theoretical framework, we find large estimates of $\beta$ ranging from $-1,093$ to $-1,391$, which implies that the sample selection in testing approaches zero as the number of tests approaches the total population size. We also find alternating signs on coefficient $\alpha$, consistent with the estimates of the power series approximation developed in Benatia et al. (2020).

Figure 2 presents scatterplots of the relationship between daily changes in per capita testing and the share of positive tests across states and provinces for the three time periods. The downward-sloping relationships imply that one-week intervals, 31 March–7 April, 14–21 April, and 28 April–5 May, to allow for the possibility that decisions about how to allocate tests across the population may have changed from late March to early May. Because policy decisions regarding testing of elderly populations may have differed across jurisdictions, we also report estimates based solely on cases among individuals aged younger than 70.

Finally, our analysis depends on the quality of diagnostic testing, and systematic false-negative test results may affect the population disease prevalence estimates (Ai et al. 2020; Liu et al. 2020; Yang et al. 2020). Because our analysis focuses on day-to-day variation, however, changes in the rates of misdiagnosis should not be systematically related to changes in the number of implemented tests. As a result, these errors should not bias the coefficient estimates, but they may reduce precision through classical measurement error (Wooldridge 2002).

Table 1: Coefficient Estimates from Equation (6)

| Coefficient | Period 1, 31 March–7 April | Period 2, 14–21 April | Period 3, 28 April–5 May |
|-------------|---------------------------|-----------------------|-------------------------|
| $\alpha_1$  | 11.704 (2.157) | 11.303 (2.186) | 10.702 (2.105) |
| $\alpha_2$  | 17.630 (2.645) | 17.505 (2.602) | 16.420 (2.551) |
| $\alpha_3$  | 8.310 (1.763) | 8.250 (1.738) | 7.760 (1.704) |
| $\beta$     | -1,391.209 (156.412) | -1,381.209 (156.412) | -1,608.010 (204.343) |
| $\sigma_\epsilon$ | 0.527 (0.018) | 0.579 (0.020) | 0.593 (0.021) |
| No. of observations | 443 | 443 | 399 |
| Notes: This table reports the estimation of the coefficients from Equation (6). We estimate the model separately for each time period and for all ages versus cases among individuals aged younger than 70. Panel A reports the coefficient estimates from the baseline model. Panel B reports the estimates from augmented models that include province and state fixed effects. Heteroskedasticity robust standard errors are reported in parentheses. Source: Authors’ calculations. |
that have no structural relationship with the dependent tests. A symptom of selection bias is that variables population disease prevalence except through a selection channel.

So, these patterns strongly suggest non-random testing, because daily changes in testing should be unrelated to the estimates for all-age population infection rates on 4 April, 18 April, and 2 May, along with heteroskedasticity robust 95 percent confidence intervals. Column (4) reports the average estimates for the three time periods 31 March–4 April, 14–18 April, and 28 April–2 May. These

Figure 2: Daily Changes in Testing and the Share of Positive Cases: (a) Period 1, 31 March–April 7; (b) Period 2, 14–21 April; and (c) Period 3, 28 April–May 5
Notes: This figure reports the relationship between daily changes in the exponential of per capita testing and daily changes in the log share of positive tests for the three time periods. The relationship in each period is obtained using the estimated coefficient of $\beta$ from the main estimates of Equation (6). See Table 1, Columns (1), (3), and (5).
Source: Authors’ calculations.

larger day-to-day increases in the number of conducted tests are associated with decreases in the share of positive tests. A symptom of selection bias is that variables that have no structural relationship with the dependent variable may appear to be significant (Heckman 1979). So, these patterns strongly suggest non-random testing, because daily changes in testing should be unrelated to population disease prevalence except through a selection channel.

Table 2 reports the results for Quebec, Ontario, Alberta, and British Columbia that adjust observed COVID-19 case rates for non-random testing on the basis of the procedure described in the Methodology section. Column (2) reports the estimates for all-age population infection rates on 4 April, 18 April, and 2 May, along with heteroskedasticity robust 95 percent confidence intervals. Column (4) reports the average estimates for the three time periods 31 March–4 April, 14–18 April, and 28 April–2 May. These
Table 2: Estimated Population Infection Rates for COVID-19

| Province          | Positive Tests, % | Estimated Population Prevalence, % | Average Estimated Population Prevalence, % |
|-------------------|------------------|-----------------------------------|-------------------------------------------|
|                   | (1)              | All Ages < 70 y                   | All Ages < 70 y                           |
|                   | 2: Early April   |                                   |                                           |
| Quebec            | 14.22            | 2.22 (1.03, 4.82)                 | 1.95 (0.87, 4.35)                         |
| Ontario           | 7.31             | 0.86 (0.41, 1.79)                 | 0.61 (0.29, 1.32)                         |
| Alberta           | 4.94             | 0.69 (0.33, 1.43)                 | 0.63 (0.30, 1.34)                         |
| British Columbia  | 2.51             | 0.23 (0.11, 0.49)                 | 0.12 (0.05, 0.26)                         |
|                   | 3: Mid-April     |                                   |                                           |
| Quebec            | 13.95            | 2.70 (1.52, 4.81)                 | 2.56 (1.45, 4.53)                         |
| Ontario           | 5.93             | 1.21 (0.67, 2.18)                 | 0.80 (0.44, 1.48)                         |
| Alberta           | 4.03             | 1.11 (0.59, 2.09)                 | 1.05 (0.55, 2.00)                         |
| British Columbia  | 2.79             | 0.43 (0.24, 0.75)                 | 0.36 (0.20, 0.63)                         |
|                   | 4: Early May     |                                   |                                           |
| Quebec            | 10.87            | 2.91 (1.51, 5.63)                 | 1.98 (1.01, 3.90)                         |
| Ontario           | 2.69             | 0.76 (0.39, 1.47)                 | 0.75 (0.38, 1.48)                         |
| Alberta           | 2.56             | 0.60 (0.32, 1.13)                 | 0.57 (0.30, 1.10)                         |
| British Columbia  | 1.25             | 0.23 (0.13, 0.43)                 | 0.24 (0.13, 0.45)                         |

Notes: Column (1) reports the fraction of positive tests on the relevant day. Columns (2)–(3) report the coefficient estimates for population prevalence of COVID-19 based on the methodology described in the Methodology section. Heteroskedasticity robust 95% confidence intervals are reported in parentheses. We report the results for all age prevalence and prevalence among individuals aged younger than 70 y. Columns (4)–(5) report the average estimates for population prevalence of COVID-19 for the three time periods. COVID-19 = coronavirus disease 2019.

Source: Authors’ calculations.

The results reveal widespread disparities in COVID-19 prevalence across provinces. Population infection rates range from more than 2.2 percent in Quebec to less than 0.4 percent in British Columbia. Trends in infection rates differed significantly across provinces. Infection rates in British Columbia declined modestly over the sample period. In Ontario, infection rates rose from early to mid-April and subsequently declined. Meanwhile, Quebec and Alberta experienced steady increases in population infection rates over the sample period.

Columns (3) and (5) report the estimated population infection rates among individuals aged younger than 70 years. These estimates will not be influenced by specific policies regarding the testing of the elderly population and residents of senior residential facilities that may have differed across provinces. The results for Alberta and British Columbia are similar to the overall population prevalence. Meanwhile, the estimates are systematically lower in Ontario and Quebec, particularly in the latter
Estimates of COVID-19 Cases across Four Canadian Provinces

We multiply the average estimated prevalence from 14–18 April (Table 2, Panel B, Column 4) by the total province population and compare it with the cumulative diagnosed cases by 23 April. The gap in the two periods captures the typical five-day incubation period to account for the fact that individuals may not seek testing until symptom onset (Lauer et al. 2020; Li, Guan, et al. 2020). In principal, these numbers capture two distinct measures of COVID-19 spread: current infections versus cumulative infections. Nevertheless, given limited COVID-19 infection before mid-March and the fact that viral presence is detectable by PCR testing three weeks after initial symptom onset (Cai et al. forthcoming; Zhou et al. 2020), population infection rates in mid-April are likely to be similar to cumulative infections since onset.

Table 4 presents the results. We find widespread undetected population infection. By 23 April, 41,371 cases had been identified across the four provinces; however, our estimates suggest that there were more than a half million infected individuals. In Quebec and Alberta, we find widespread undetected population infection.

Figure 3: Population COVID-19 Infection Rates by Province and Period

Notes: This figure reports average population infection rates across provinces for three different time periods: (1) 31 March–4 April, (2) 14 April–18 April, and (3) 28 April–2 May. These average infection rates are obtained using the estimation procedure described in the Methodology section. All age population prevalence estimates are based on all cases and tests. To derive population prevalence for those aged younger than 70 y, we subtract the number of cases among elderly persons from the total daily cases and total tests across provinces. COVID-19 = coronavirus disease 2019.

Source: Authors’ calculations.
Discussion

This article provides new evidence on the population prevalence of COVID-19 in Quebec, Ontario, Alberta, and British Columbia from late March to early May. Our analysis adapts a sample selection model approach developed in Benatia et al. (2020). We find widespread population

limited in British Columbia, allowing officials to better identify clusters of cases.

Table 3: Robustness Exercises: Fixed-Effects Models

| COVID-19 Prevalence in | Estimated Population Prevalence, % | Average Estimated Pop. Prevalence, % |
|------------------------|------------------------------------|--------------------------------------|
|                        | All Ages                           | < 70 y                               | All Ages                           | < 70 y                               |
|                        | Baseline (1)                        | Add Fixed Effects (2)                | Baseline (3)                        | Add Fixed Effects (4)                | Baseline (5)                        | Add Fixed Effects (6)                | Baseline (7)                        | Add Fixed Effects (8)                |
| A: Early April         |                                     |                                     |                                     |                                     |                                     |                                     |                                     |                                     |
| Quebec                 | 4 April                             |                                     |                                     |                                     |                                     |                                     |                                     |
| Baseline               | 2.22                                | 2.32                                | 1.95                                | 2.04                                | 2.08                                | 2.17                                | 1.85                                | 1.93                                |
|                        | (1.03, 4.82)                        | (1.33, 4.08)                        | (0.87, 4.35)                        | (1.14, 3.67)                        | (1.23)                              | (1.28)                              | (0.96)                              | (1.00)                              |
| Ontario                | 0.86                                | 0.89                                | 0.61                                | 0.64                                | 1.23                                | 1.28                                | 0.96                                | 1.00                                |
|                        | (0.41, 1.79)                        | (0.52, 1.52)                        | (0.29, 1.32)                        | (0.37, 1.11)                        | (0.51)                              | (0.54)                              | (0.46)                              | (0.48)                              |
| Alberta                | 0.69                                | 0.72                                | 0.63                                | 0.65                                | 0.36                                | 0.38                                | 0.28                                | 0.29                                |
|                        | (0.33, 1.43)                        | (0.42, 1.22)                        | (0.30, 1.34)                        | (0.38, 1.14)                        |                                    |                                    |                                    |                                    |
| British Columbia       | 0.23                                | 0.24                                | 0.12                                | 0.12                                |                                    |                                    |                                    |                                    |
|                        | (0.11, 0.49)                        | (0.14, 0.41)                        | (0.05, 0.26)                        | (0.07, 0.22)                        |                                    |                                    |                                    |                                    |
| B: Mid-April           | 18 April                            |                                     |                                     |                                     |                                     |                                     |                                     |                                     |
| Quebec                 | 1.21                                | 1.19                                | 1.05                                | 1.05                                | 1.00                                | 0.98                                | 0.89                                | 0.89                                |
|                        | (0.59, 2.09)                        | (0.70, 1.72)                        | (0.55, 2.00)                        | (0.66, 1.66)                        | (1.00)                              | (0.98)                              | (0.89)                              | (0.89)                              |
| Ontario                | 0.43                                | 0.42                                | 0.36                                | 0.35                                | 0.31                                | 0.31                                | 0.24                                | 0.23                                |
|                        | (0.24, 0.75)                        | (0.28, 0.63)                        | (0.20, 0.63)                        | (0.23, 0.53)                        |                                    |                                    |                                    |                                    |
| C: Early May           | 2 May                               |                                     |                                     |                                     |                                     |                                     |                                     |                                     |
| Quebec                 | 2.91                                | 2.97                                | 1.98                                | 2.02                                | 3.60                                | 3.66                                | 2.35                                | 2.39                                |
|                        | (1.51, 5.63)                        | (1.87, 4.73)                        | (1.01, 3.90)                        | (1.25, 3.26)                        |                                    |                                    |                                    |                                    |
| Ontario                | 0.76                                | 0.77                                | 0.75                                | 0.78                                | 0.86                                | 0.88                                | 0.76                                | 0.78                                |
|                        | (0.39, 1.47)                        | (0.48, 1.24)                        | (0.38, 1.48)                        | (0.47, 1.24)                        |                                    |                                    |                                    |                                    |
| Alberta                | 0.60                                | 0.61                                | 0.57                                | 0.59                                | 1.16                                | 1.18                                | 1.13                                | 1.16                                |
|                        | (0.32, 1.13)                        | (0.39, 0.96)                        | (0.30, 1.10)                        | (0.37, 0.93)                        |                                    |                                    |                                    |                                    |
| British Columbia       | 0.23                                | 0.24                                | 0.24                                | 0.24                                | 0.27                                | 0.27                                | 0.26                                | 0.26                                |
|                        | (0.13, 0.43)                        | (0.15, 0.37)                        | (0.13, 0.45)                        | (0.16, 0.37)                        |                                    |                                    |                                    |                                    |

Notes: This table explores the sensitivity of the findings to controls for province and state fixed effects. Columns (1)–(4) report the estimated population infection rates on the relevant date. Columns (5)–(8) report the average estimates for population prevalence of COVID-19 for the three time periods. Columns (1), (3), (5), and (7) report the baseline estimates, and Columns (2), (4), (6), and (8) report the estimates based on augmented models that include province and state fixed effects. Heteroskedasticity robust 95% confidence intervals are reported in parentheses.

Source: Authors’ calculations.
infection that exceeds official reported cases by factors of 9 to 15 across provinces.

Our findings are comparable to recent prevalence estimates from the United States and countries in Western Europe. The estimated infection rates in Quebec are similar to those of the United Kingdom (2.7%) and several US states (Pennsylvania, 2.4%; Rhode Island, 2.4%; and Massachusetts, 3.4%). Meanwhile, the rates in Ontario are similar to those in Austria (1.1%), Denmark (1.1%), Vermont (1.4%), Virginia (1.4%), and Idaho (1.5%) in early April (see Benatia et al. 2020; Ferguson et al. 2020; Johndrow et al. 2020; Javan et al. 2020). Our results are also consistent with recent evidence from serological testing across several US jurisdictions that shows widespread undetected infection by mid-April (Bendavid et al. 2020; Conarck and Chang 2020; Goodman and Rothfeld 2020).

Our analysis provides a complement to existing methods used to estimate population infection rates. These approaches require either strong assumptions about unknown disease parameters or accurate measurement of COVID-related deaths, which may be undercounted over the course of the pandemic. Our estimation approach builds on standard econometric techniques. As high-frequency test data become more widely available at finer geographic units, this approach could be applied to estimate population infection rates at the city or district level. The current estimates depend on the accuracy of an extrapolation of the functional form relating the number of tests and the positive case rate to the large untested population. In ongoing work, we hope to refine the estimation procedure and explore the sensitivity of the results to various functional form assumptions.

As physical distancing policies continue to be relaxed, it will be essential that policy-makers have access to timely data on infection rates. Given the potential for widespread undiagnosed infection, the expansion of randomized population-based PCR testing may play a key role in identifying localized outbreaks. Meanwhile, widespread implementation of serological testing will help identify the large numbers of individuals with some level of immunity to the virus.

Acknowledgements
We thank seminar participants at McGill University, Université de Montréal, and École Polytechnique for valuable suggestions. We are especially grateful to Xavier d’Hautefoeuill, Tim Guinnane, Angelo Melino, Peter Tankov, and anonymous referees for detailed comments. This study was supported by funding from the Social Sciences and Humanities Research Council.

Notes
1 A large body of research in economics is devoted to the issue of non-random sampling. See Blundell and Costa Dias (2002); Das, Newey, and Vella (2003); Heckman (1976, 1979); Heckman, Lalonde, and Smith (1999); and Newey (2009). Epidemiologists have also devoted considerable attention to the issue of sample representativeness (for a detailed discussion, see Davey Smith and Ebrahim 2013).
2 To improve the precision of the estimates, we include all Canadian provinces and US states to estimate the selection bias gradient. Once this gradient has been estimated, however, our estimates of underlying prevalence rely solely on the shares of positive cases across the four Canadian provinces we study.
3 There is an extended period over which individuals may test positive for COVID-19. Polymerase chain reaction (PCR) testing has identified cases days before symptom onset and detected continued viral RNA presence more than three weeks after symptom onset (Cai et al. 2020; Huang et al. 2020; Zhou et al. 2020). These positive cases often occur among individuals who are no longer symptomatic, and it is believed that they reflect lingering viral material that no longer poses a risk of transmission.
4 British Columbia has a large number of missing observations because daily testing results were not always released.
5 The median ratio of negative to positive tests, $q$, across $p_A$, across Canadian provinces was 21 during the sample period.
6 The first difference derivation is based on the following equation: $\log \frac{s}{n} = \log (1 + e^{\rho_{\text{dn}}}) + \log \frac{A}{B} = \sum_{i=1}^{M} (-1)^{i+1} e^\frac{\rho_{\text{dn}}}{k} e^\rho_{\text{dn}} + \log \frac{A}{B}$.
The results are not sensitive to the inclusion of higher-order terms. Our baseline estimating equation does not include a constant, although we explore the robustness of the results to jurisdiction-specific intercepts.

Similar standard errors are found in models that allow for within-state/within-province serial correlation.

There is little guidance from theory about the functional form of relationship, and there are no individual-level survey data to shed light on who is tested.

To see why this is the case, let $\pi < 1$ denote the fixed probability that a test is positive if an individual is sick, so that some fraction of sick individuals may not be detected by testing. In this case, the researcher observes $s/n$, but the actual share of positive cases among the tested sample is $\frac{s}{n}$.

Provided that the rate of false negatives is constant over time, the term $\pi$ drops out of the first difference Equation (4), so it will not affect the main estimates.

Figure 3 presents both all-age and aged-younger-than-70-years population infection rates.

To derive these estimates, we subtract the number of cases among elderly individuals from the total daily cases and to-total daily tests across provinces. Because we lack data on total tests by age group, negative tests among elderly individuals are included in the denominator, so these estimates should be interpreted as a lower-bound estimate for disease prevalence.

References

Ai, T., Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia. 2020. “Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report on 1014 Cases.” Radiology 296(2):E32–E40. https://doi.org/10.1148/radiol.2020200642.

Aspelund, K., M. Drost, J. Stock, and C. Walker. 2020. “Identification and Estimation of Undetected COVID-19 Cases Using Testing Data from Iceland.” Working Paper 27528. Cambridge, MA: National Bureau of Economic Research.

Bai, Y., L. Yao, T. Wei, F. Tian, D. Jin, L. Chen, and M. Wang. 2020. “Presumed Asymptomatic Carrier Transmission of COVID-19.” JAMA 323(14):1406. https://doi.org/10.1001/jama.2020.2565.

Benatia, D., R. Godefroy, and J. Lewis. 2020. “Estimating COVID-19 Prevalence in the United States: A Sample Selection Model Approach.” medRxiv Working Paper. https://doi.org/10.1101/2020.04.20.20072942.

Bendavid, E., J. Bhattacharya, B. Mulaney, N. Sood, S. Shah, E. Ling, R. Bromley-Duffano, C. Lai, Z. Weissberg, R. Saavedra-Walker, et al. 2020. “COVID-19 Antibody Seroprevalence in Santa Clara County, California.” medRxiv Working Paper. https://doi.org/10.1101/2020.04.14.20062463.

Berry, L., J. Soucy, A. Tuite, and D. Fisman. 2020. “Open Access Epidemiological Data and an Interactive Dashboard to Monitor the COVID-19 Outbreak in Canada.” CMAJ 192(15):E420. https://doi.org/10.1503/cmaj.75262.

Blundell, R., and M. Costa Dias. 2002. “Evaluation Methods for Non-Experimental Data.” Fiscal Studies 21(4):427–68. https://doi.org/10.1111/j.1475-5890.2000.tb00031.x.

Cai, J., J. Xu, D. Lin, Z. Yang, L. Xu, Z. Qu, Y. Zhang, H. Zhang, R. Jia, P. Liu, et al. forthcoming. “A Case Series of Children with 2019 Novel Coronavirus Infection: Clinical and Epidemiological Features.” Clinical Infectious Disease. https://doi.org/10.1093/cid/ciaa198.

Center for Systems Science and Engineering (CSSE), Johns Hopkins University. 2020. “COVID-19 in the USA.” At https://coronavirus.jhu.edu/.

Centers for Disease Control and Prevention (CDC). 2020a. “Commercial Laboratory Seroprevalence Survey Data.” At https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/commercial-lab-surveys.html.

Centers for Disease Control and Prevention (CDC). 2020b. “Coronavirus (COVID-19).” At https://www.cdc.gov/coronavirus/2019-ncov/index.html.

Clay, K., J. Lewis, and E. Severini. 2018. “Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic.” Journal of Economic History 78(4):1179–209. https://doi.org/10.3386/w21635.

Clay, K., J. Lewis, and E. Severini. 2019. “What Explains Cross-city Variation in Mortality during the 1918 Influenza Pandemic? Evidence from 440 U.S. Cities.” Economics and Human Biology 35:42–50. https://doi.org/10.1016/j.ehbi.2019.03.010.

Clay, K., J. Lewis, E. Severini, and X. Wang. 2020. “The Value of Health Insurance during a Crisis: Effects of Medicaid Implementation on Pandemic Influenza Mortality.” Working Paper 27120. Cambridge, MA: National Bureau of Economic Research.

Conarck, B., and D. Chang. 2020. “Miami-Dade Has Tens of Thousands of Missed Coronavirus Infections, UM Survey Finds.” Miami Herald, 24 April. At https://www.miamiherald.com/news/coronavirus/article242260406.html.

Das, M., W. Newey, and F. Vella. 2003. “Nonparametric Estimation of Sample Selection Models.” Review of Economic Studies 70(1):33–58. https://doi.org/10.1111/1467-937x.00236.

Davey Smith, G., and S. Ebrahim, ed. 2013. International Journal of Epidemiology 42(4).

Dong, Y., X. Mo, Y. Hu, X. Qi, F. Jiang, Z. Jiang, and S. Tong. 2020. “Epidemiological Characteristics of 2143 Pediatric Patients with 2019 Coronavirus Disease in China.” Pediatrics 145(6):e20200702. https://doi.org/10.1542/peds.2020-0702.

Ferguson, N., D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba, G. Cuomo-Dannenburg, et al. 2020. Impacts of Non-Pharmaceutical Interventions to Reduce COVID-19 Mortality and Healthcare Demand. London: Imperial College COVID-19 Response Team.

Goodman, D., and M. Rothfeld. 2020. “1 in 5 New Yorkers May Have Had COVID-19, Antibody Tests Suggest.” New York Times, 23 April. At https://www.nytimes.com/2020/04/23/nyregion/coronavirus-antibodies-testy.html.

Han, Y., J. Lam, V. Li, P. Guo, Q. Zhang, A. Wang, J. Crowcroft, S. Wang, J. Fu, Z. Gilani, et al. 2020. “The Effects of Outdoor Air Pollution Concentrations and Lockdowns on COVID-19 Infections in Wuhan and Other Provincial Capital in China.” Working Paper. https://doi.org/10.20944/preprints202003.0364.v1.
Heckman, J. 1976. “The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models.” *Annals of Economics and Social Measurement* 5(4):475–92.

Heckman, J. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica* 47(1):153–62. https://doi.org/10.2307/1912352.

Heckman, J., R. Lalonde, and J. Smith. 1999. “The Economics of Active Labor Market Programs.” In *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card, 1866–2097. Amsterdam: North-Holland.

Hinkson, K., and V. Stevenson. 2020. “Quebec Premier Says All Patients and Staff to Be Tested at Long-Term Care Homes.” *CBC News*, 8 April. At https://www.cbc.ca/news/canada/quebec/quebec-covid-19-april-8-1-15525861.

Hoehl, S., H. Rabenau, A. Berger, M. Kortenbusch, J. Cinatlı, D. Bojkova, P. Behrens, B. Böddinghaus, U. Götsch, F. Naujoks, et al. 2020. “Evidence of SARS-CoV-2 Infection in Returning Travelers from Wuhan, China.” *New England Journal of Medicine* 382(13):1278–80. https://doi.org/10.1056/NEJMoa2001899.

Huang, R., J. Xia, Y. Chen, C. Shan, and C. Wu. 2020. “A Family Cluster of SARS-CoV-2 Infection Involving 11 Patients in Nanjing, China.” *Lancet Infectious Disease* 20(5):534–35. https://doi.org/10.1016/S1473-3099(20)30147-X.

Javan, E., S. Fox, and L. Meyers. 2020. “Probability of Current COVID-19 Outbreaks in All US Counties.” Working Paper. Austin: University of Texas at Austin, John Ring LaMontagne Center for Infectious Disease. At https://cid.utexas.edu/sites/default/files/cid/files/covid-risk-maps_counties_4.3.2020.pdf?m=1585958755.

Johndrow, J., K. Lum, and P. Ball. 2020. “Estimating SARS-CoV-2 Positive Americans Using Deaths-Only Data.” At https://www.researchgate.net/publication/340474998_Estimating_SARS-CoV-2-positive_Americans_using_deaths-only_data.

Jones, A. 2020. “Coronavirus: Ontario to Test All Long Term Care Residents, Staff for COVID-19.” *Global News*, 22 April. At https://globalnews.ca/news/6852825/ontario-test-all-long-term-care-residents-staff-coronavirus.

Katz, J., and M. Sanger-Katz. 2020. “Death in New York City Are More than Double the Usual Total.” *New York Times*, 10 April. At https://www.nytimes.com/interactive/2020/04/10/upa/sub/2020-coronavirus-death-new-york-city.html.

Lauer, S., K. Grantz, Q. Bi, F. Jones, Q. Zheng, H. Meredith, A.S. Azman, N.G. Reich, and J. Lessler. 2020. “The Incubation Period of Coronavirus Disease 2019 (COVID-19) from Publicly Reported Confirmed Cases: Estimation and Application.” *New England Journal of Medicine* 172(9):577–82. https://doi.org/10.1056/NEJMoa2001316.

Li, Q., X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. M. Leung, E. H. Y. Lau, J. Y. Wong, et al. 2020. “Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus—Infected Pneumonia.” *New England Journal of Medicine* 26;582(13):1199–207. https://doi.org/10.1056/NEJMoa2001316.

Li, R., S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman. 2020. “Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV-2).” *Science* 368(6490):489–93. https://doi.org/10.1126/science.abb3221.

Liu, J., X. Xie, Z. Zhong, W. Zhao, C. Zheng, and F. Wang. 2020. “Chest CT for Typical 2019-nCoV Pneumonia: Relationship to Negative RT-PCR Testing.” *Radiology* 296(2):E41–E45. https://doi.org/10.1148/radiol.2020200330.

Lu, X., L. Zhang, H. Du, J. Zhang, Y. Li, J. Qu, W. Zhang, Y. Wang, S. Bao, Y. Li, et al. 2020. “SARS-CoV-2 Infection in Children.” *New England Journal of Medicine* 382(17):1663–65. https://doi.org/10.1056/NEJMc2005073.

Manski, C., and F. Molinari. forthcoming. “Estimating the COVID-19 Infection Rate: Anatomy of an Inference Problem.” *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.04.041.

Meyer, R., E. Kissane, and A. Madrigal. 2020. “The COVID Tracking Project.” At https://covidtracking.com/.

Newey, W. 2009. “Two-Step Series Estimation of Sample Selected Models.” *Econometrica* 12(Supplement 1):S217–29. https://doi.org/10.1111/j.1468-423X.2008.00263.x.

Pan, X., D. Chen, Y. Xia, X. Wu, T. Li, X. Ou, L. Zhou, and J. Liu. 2020. “Asymptomatic Cases in a Family Cluster with SARS-CoV-2 Infection.” *Lancet Infection Disease* 20(4):410–11. https://doi.org/10.1016/S1473-3099(20)30114-6.

Perkins, A., S. Cavany, S. Moore, R. Oldtman, A. Lerch, and M. Poterkek. 2020. “Estimating Unobserved SARS-CoV-2 Infections in the United States.” medRxiv Working Paper. https://doi.org/10.1101/2020.03.15.20036582.

Prakash, N., and E. Hall. 2020. “Doctors and Nurses Say More People Are Dying of COVID-19 in the US than We Know.” *Buzzfeed News*, 25 March. At https://www.buzzfeednews.com/article/nidhiprakash/coronavirus-update-dead-covid19-doctors-hospitals.

Reid, A. 2020. “The Incidence of COVID-19 Infection in Canada? New Survey Points to Over 100,000 Household.” Angus Reid Institute, 8 April. At http://angusreid.org/covid-epidemiology-study/.

Riou, J., A. Hauser, M. Cournou, and C. Althaus. 2020. “Adjusted Age-Specific Case Fatality Rates during the COVID-19 Epidemic in Hubei, China, January and February.” medRxiv Working Paper. At https://www.medrxiv.org/content/10.1101/2020.03.04.20031104v1.full.pdf.

Riou, J., A. Hauser, M. Cournou, M. Margossian, G. Constantinoudis, N. Low, and C.L. Althaus. 2020. “Estimation of SARS-CoV-2 Mortality during the Early Stages of an Epidemic: A Modelling Study in Hubei, China and Northern Italy.” *PLOS Medicine* 17(7):e1003189. https://doi.org/10.1371/journal.pmed.1003189.

Statistics Canada. 2020. “Population Estimates on July 1, by Age and Sex (Table 17-10-0005-01).” At https://www15.statcan.gc.ca/t1/tbl1/en/tv.action?pid=171000501.

Verity, R., L. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P.G.T. Walker, H. Fu, et al. 2020. “Estimates of the Severity of Coronavirus Disease 2019: A Model-Based Analysis.” *Lancet Infectious Disease* 20(6):669–77. https://doi.org/10.1016/S1473-3099(20)30243-7.

Wooldridge, J. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
Wu, X., R. Nethery, B. Sabath, D. Braun, and F. Dominici. 2020. “Exposure to Air Pollution and COVID-19 Mortality in the United States.” medRxiv Working Paper. https://doi.org/10.1101/2020.04.05.20054502.

Yang, Y., M. Yang, C. Shen, F. Wang, J. Yuan, J. Li, M. Zhang, Z. Wang, and L. Xing. 2020. “Evaluating the Accuracy of Different Respiratory Specimens in the Laboratory Diagnosis and Monitoring the Viral Shedding of 2019-nCoV Infections.” medRxiv Working Paper. https://doi.org/10.1101/2020.02.11.20021493.

Zhou, F., T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, et al. 2020. “Clinical Course and Risk Factors for Mortality of Adult Inpatients with COVID-19 in Wuhan, China.” Lancet 395(10229):1054–62. https://doi.org/10.1016/S0140-6736(20)30566-3.