



# On testing for infections during epidemics, with application to Covid-19 in Ontario, Canada



Jerald F. Lawless <sup>a,\*</sup>, Ping Yan <sup>a,b</sup>

<sup>a</sup> Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

<sup>b</sup> Public Health Agency of Canada, 130 Colonnade Rd., Ottawa, ON, K1A 0K9, Canada

## ARTICLE INFO

### Article history:

Received 29 April 2021

Received in revised form 16 July 2021

Accepted 18 July 2021

Available online 22 July 2021

Handling editor: Dr HE DAIHAI HE

### Keywords:

Count data

COVID-19

Modelling

Testing strategies

Ascertainment rate

## ABSTRACT

During an epidemic, accurate estimation of the numbers of viral infections in different regions and groups is important for understanding transmission and guiding public health actions. This depends on effective testing strategies that identify a high proportion of infections (that is, provide high ascertainment rates). For the novel coronavirus SARS-CoV-2, ascertainment rates do not appear to be high in most jurisdictions, but quantitative analysis of testing has been limited. We provide statistical models for studying testing and ascertainment rates, and illustrate them on public data on testing and case counts in Ontario, Canada.

© 2021 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Accurate estimation of the extent and location of viral infections during an epidemic is important for understanding transmission, estimating hospitalization and fatality rates, and informing public health actions. This requires effective testing strategies, especially if many infected persons remain asymptomatic. For the novel coronavirus SARS-CoV-2 and Covid-19 disease, official counts of tests, confirmed cases and test positivity rates give only a partial picture. Confirmed case counts and positivity rates depend on testing rates, testing criteria, individuals' willingness to be tested, and other factors, and a substantial proportion of cases remains unidentified. Our objective is to discuss testing models for SARS-CoV-2 and consider testing strategies, with reference to experience in the Canadian province of Ontario.

Since the arrival of SARS-CoV-2 in Canada in January 2020 (Marchand-Sénécal et al., 2020), provinces and territories have provided daily and weekly updates on the epidemic. All jurisdictions report newly confirmed cases (infections), deaths and hospitalizations along with breakdowns by sex, age, presumed exposure and other factors (e.g. see Government of Alberta, 2021; Government of Ontario, 2021a,b). Other public data include the daily numbers of tests processed and the corresponding positivity rates. Testing levels increased greatly between March and December 2020, and in Ontario, have been fairly stable since January 2021. We will examine test and case counts up to the end of January 2021 later in the paper.

\* Corresponding author.

E-mail address: [jlawless@uwaterloo.ca](mailto:jlawless@uwaterloo.ca) (J.F. Lawless).

Peer review under responsibility of KeAi Communications Co., Ltd.

Our specific aims are to consider the connection between testing, case counts and infection rates, and to examine variation in testing rates. We provide simple statistical models motivated by the SARS-COV-2 pandemic, discuss testing strategies, and how more detailed data and random testing would enable estimation of ascertainment rates and thus better assessment of strategies. Section 2 introduces the statistical framework and Section 3 discusses testing strategies. Section 4 considers analysis of testing rates and Section 5 illustrates the models and methods using data from the province of Ontario. Section 6 makes concluding remarks.

## 2. A statistical framework for testing and case counts

### 2.1. Tests in a specific time period

We consider test and case counts in some population for time periods indexed by the variable  $t$ . We first set notation for an arbitrary time period, and suppress the value of  $t$  for convenience: let  $N$  denote the population size,  $n$  the number of tests with results in the time period,  $C_{test}$  the number of positive tests, and  $C_{pop}$  the (unknown) number of active but unidentified cases (infections) in the population just before the current time period. We also define rates related to the data: the test positivity rate,  $PR = C_{test}/n$ ; the testing rate,  $TR = n/N$ ; the confirmed case rate  $CCR = C_{test}/N$ ; and the proportion of the population infected but not previously identified,  $p = C_{pop}/N$ . We will treat  $TR$  as the proportion of the population with test results in a specific time period; similarly we treat  $C_{test}$  as the number of persons testing positive. This is a slight idealization, since some individuals may be tested more than once. We typically consider days and weeks for time periods and assume the numbers of repeat tests in a period is negligible. Over longer periods a difficulty in allowing for repeated testing is that many jurisdictions publish only the numbers of tests and positive tests, and not the numbers of distinct persons tested. We note that  $C_{test}$  is distinct from the number of new cases reported during the same time period, which we denote by  $C$ . The count  $C$  includes some persons whose positive test results were in a previous time period, and depends on the time between a case being confirmed in a laboratory, reported to a local public health unit, and then to a higher authority. Conversely, some of the cases included in  $C_{test}$  are not reported in the current time period.

To address testing effectiveness we define an additional rate,  $TRI = C_{test}/C_{pop}$ , which is the proportion of infected but currently unconfirmed persons in the population who are tested (or “test rate for infecteds”). For now we assume that tests have perfect sensitivity and specificity, so that a tested person is infected if and only if the test is positive. Imperfect tests are discussed later in this section. With a perfect test the confirmed cases are all true cases, and the defined rates satisfy the relationship

$$CCR = p \times TRI = TR \times PR. \tag{1}$$

The rates  $CCR$ ,  $TR$  and  $PR$  are known, but  $p$  and  $TRI$  are not. The numbers of tests completed on a given day and the number of positive tests are random outcomes that depend on several factors, including personal behaviour in accessing testing and random variation in collecting and processing specimens. For a random individual in the population, excluding persons previously confirmed to be infected, let

$$Y = \begin{cases} 1 & \text{if infected} \\ 0 & \text{if not infected} \end{cases} \quad \Delta = \begin{cases} 1 & \text{if tested in the time period} \\ 0 & \text{if not tested} \end{cases}$$

The key parameters for the joint distribution of  $Y, \Delta$  are

$$p = Pr(Y = 1) \quad \pi_1 = Pr(\Delta = 1|Y = 1) \quad \pi_0 = Pr(\Delta = 1|Y = 0),$$

where  $\pi_1$  and  $\pi_0$  are the testing rates among infected and non-infected persons, and  $\theta = \pi_1 p + \pi_0(1 - p)$  is the overall testing rate. Table 1 shows the joint probabilities for  $Y$  and  $\Delta$ ; for a given person,  $Y$  is known only if they are tested ( $\Delta = 1$ ).

Letting  $r = E(CCR)$ , then analogous to (1), we have the relationship

$$r = p \pi_1 = \theta \gamma, \tag{2}$$

where  $\gamma = \pi_1 p / \theta = P(Y = 1|\Delta = 1) = E(PR|n) = E(PR)$ . The probability  $\pi_1$  applies to tests completed in a specific time period, and is different from the overall probability an infection is (eventually) identified by testing, which is discussed below. Data ( $n$ ,

**Table 1**  
Population infection and testing probabilities.

	$\Delta = 1$	$\Delta = 0$	Total
$Y = 1$	$\pi_1 p$	$(1 - \pi_1) p$	$p$
$Y = 0$	$\pi_0 (1 - p)$	$(1 - \pi_0) (1 - p)$	$1 - p$
Total	$\theta$	$1 - \theta$	1

$C_{test}$ ) for a given time period give estimates  $\hat{\theta} = TR = n/N$ ,  $\hat{\gamma} = PR = C_{test}/n$  and  $\hat{r} = CCR = C_{test}/N$ ; the rates  $p$  and  $\pi_1$  are not separately estimable. When repeat tests on individuals in a time period are negligible, the relationships (1) and (2) indicate that a trend in confirmed case counts or positivity rates across successive time periods does not imply a corresponding trend in actual cases or  $p$ ; trends in  $\theta$  and  $\pi_1$  also need to be considered. The value of  $\pi_1$  is unknown in most settings, but we note two extreme situations: (i) all unidentified infected persons get tested, so  $\pi_1 = 1$ , and (ii) those tested are a random sample of the population (excluding confirmed cases), so that  $\pi_1 = \theta$ . Neither situation is realistic across a large heterogeneous population, but sometimes apply to specific groups within the population. For example, in groups with severe outbreaks everyone might be tested, so  $\pi_1 = 1$ . Situation (ii) applies in the case of random surveillance testing, used to obtain an estimate of the true infection rate (Office of National Statistics U.K., 2020).

A primary aim of testing is to identify a high proportion of cases. Success depends on how effectively testing targets those who are infected, as measured by the rate  $\pi_1$ . The ascertainment rate, or overall probability a case is identified through testing, is denoted as  $\pi^*$  for infections occurring over some specific time period, whereas  $\pi_1$  applies to tests reported for a given time period. The value of  $\pi^*$  depends on the accuracy of tests and on the values of  $\pi_1$  for about two weeks following the time of infection; we discuss this in the next section.

### 2.2. Time series of test results, ascertainment and infection rates

To consider multiple time periods we denote infection, testing and expected positivity rates for time period  $t$  as  $p(t)$ ,  $\theta(t)$ ,  $\gamma(t)$  respectively; similarly we have  $\pi_1(t)$ ,  $\pi_0(t)$ ,  $Y(t)$ ,  $\Delta(t)$ ,  $n(t)$ ,  $C_{test}(t)$ ,  $PR(t)$  and so on. Days are the usual time unit for reporting results; however, there is systematic day-of-the-week variation in testing and reporting in most jurisdictions so weekly counts and rates will be used in later analyses. Population size  $N(t)$  is roughly constant across time periods and treated as a fixed value  $N$ .

The unconfirmed infection rate  $p(t)$  is related to a population's rate of new infections as follows. Let  $I(t)$  denote the number of new infections that become detectable on day  $t$ ; this is a function of the infection rate for recent days and the time delay before infections become detectable by testing. The number of unidentified infections in the population at the start of day  $t$  satisfies the relationship

$$C_{pop}(t) = d(t)\{C_{pop}(t - 1) - C_{test}(t - 1)\} + I(t),$$

where  $d(t)$  is the proportion of unidentified cases at the end of day  $t - 1$  that remain detectable on day  $t$ . In terms of testing and infection rates this relationship becomes

$$p(t) = d(t)p(t - 1)\{1 - \pi_1(t - 1)\} + i(t) = d(t)\{p(t - 1) - r(t - 1)\} + i(t), \tag{3}$$

where  $i(t) = I(t)/N$ . Thus  $p(t)$  is determined by the rate  $i(t)$  of newly detectable infections and the rates at which existing infections are either identified through testing or progress to the point where they are no longer detectable.

If the values of  $\pi_1(t)$  were close to one, most new infections would be identified soon after they become detectable. This is rarely the case because of limited tracing and testing resources, and in most settings the daily values of  $\pi_1(t)$  are small. They can be related to the overall probability an infection is detected, or the ascertainment rate,  $\pi^*$ ; to consider variation over time we let  $\pi^*(t)$  denote the probability an infection that first becomes detectable on day  $t$  is eventually confirmed. If the maximum time an infection remains detectable is  $s_{max}$  days and if  $\pi^*(s; t)$  denotes the probability an infection that becomes detectable on day  $t$  is confirmed on day  $s$  ( $s = t, t + 1, \dots, t + s_{max}$ ), then

$$\pi^*(t) = \sum_{s=t}^{t+s_{max}} \pi^*(s; t). \tag{4}$$

The  $\pi^*(s; t)$  may vary with  $s - t$ ; they also vary with  $s$  if testing levels and strategies change over time. In the case where dependence on  $s - t$  is negligible,  $\pi^*(s; t) = \pi_1(s)$  and  $\pi^*(t) = \pi_1(t, t + s_{max})$ , where  $\pi_1(a, b)$  denotes the sum of rates  $\pi_1(s)$  over days  $a$  to  $b$ .

Ascertainment rates have been estimated in various countries and regions using a number of methods (e.g. Pullano et al., 2021; Shaman, 2021). Repeated random testing is the only reliable way to estimate  $p(t)$  and thus  $\pi_1(t)$ , and some countries have regular random surveillance surveys; we discuss this in Section 3. Random testing has not been widespread in Canada, but antibody seroprevalence surveys have provided some estimates of overall ascertainment. In the province of Ontario a July 2020 study (Ontario Agency for Health Protection and Promotion (Public Health Ontario), 2020) estimated that about 1.1% of the population had been infected by July 26. This was not a random survey since it was based on persons who had blood tests for other reasons, but for illustration suppose the 1.1% figure is accurate. The number of cases confirmed and reported by July 26 was about 38,800; with the Ontario population of about 14,700,000 this implies an estimate of about 0.24 for the average ascertainment rate  $\pi^*$  up to that time. If  $\pi^*(t)$  were about 0.24 and  $s_{max}$  is, say, 14 days, then the average daily rates  $\pi_1(s)$  would be about 0.017. A small but better designed national seroprevalence survey covering March to August 2020 in Canada (Ab-C Study Investigators, 2021) estimated the seroprevalence rate for Ontario adults to be 2.35% or higher. There were

approximately 42,400 confirmed cases (including children) by the end of August, so this implies an overall ascertainment rate substantially lower than 0.24.

### 2.3. Population stratification

Confirmed case and positivity rates can vary widely across parts of a population, according to age, residence, place of work and other factors. To consider variation in testing and case rates we suppose the population or some part of it has been partitioned into strata. In a population (of size  $N$ ) with strata  $S_1, S_2, \dots, S_K$  let  $z$  denote which stratum an individual is in. Let  $N_k$  be the size of stratum  $k$ , so for a randomly chosen person,  $\lambda_k = Pr(z = k) = N_k/N$ . Suppressing notation for the time period, we now have  $\theta_k = Pr(\Delta = 1|z = k)$ ,  $p_k = Pr(Y = 1|z = k)$ , and expected confirmed case rates  $r_k$  satisfy  $r_k = p_k\pi_{1k} = \theta_k\gamma_k$ , for  $k = 1, \dots, K$ . The overall expected positivity rate is then

$$\gamma = Pr(Y = 1|\Delta = 1) = \frac{r}{\theta} = \frac{\sum_k \gamma_k \theta_k \lambda_k}{\sum_k \theta_k \lambda_k} = \frac{\sum_k p_k \pi_{1k} \lambda_k}{\sum_k \theta_k \lambda_k}, \tag{5}$$

and the infection rate is  $p = Pr(Y = 1) = \sum_k p_k \lambda_k$ . If a higher proportion of testing is allocated to strata with higher positivity rates, the effect is to increase  $\gamma$  and  $r$ . However, information about the  $\pi_{1k}$  is needed to determine the effect on

$$\pi_1 = Pr(\Delta = 1|Y = 1) = \frac{r}{p} = \frac{\sum_k \gamma_k \theta_k \lambda_k}{\sum_k p_k \lambda_k} = \frac{\sum_k p_k \pi_{1k} \lambda_k}{\sum_k p_k \lambda_k}, \tag{6}$$

and the ascertainment rate. Section 3 considers how testing might be allocated across strata.

### 2.4. Imperfect tests

Tests are almost never perfect, and may give false positive and false negative results. In this case we observe  $\tilde{Y} = I(\text{test is positive})$  for an individual who is tested, and  $\tilde{Y}$  replaces  $Y$  in Table 1; in addition we denote the number of positive tests in some time period as  $\tilde{C}_{test}$  and the positivity rate as  $\tilde{PR}$ . If  $\alpha = Pr(\tilde{Y} = 1|Y = 1)$  is the sensitivity of the test and  $\beta = Pr(\tilde{Y} = 1|Y = 0)$  is one minus the specificity, or false positive rate, then Table 1 implies that  $\tilde{PR}$ ,  $\tilde{CCR} = \tilde{C}_{test}/N$  and the estimate  $\tilde{p} = \tilde{PR}(TR)/\pi_1$  have expectations

$$E(\tilde{PR}|n) = (\alpha - \beta)\gamma + \beta, \quad E(\tilde{CCR}) = \theta E(\tilde{PR}|n), \quad E(\tilde{p}) = (\alpha - \beta)p + \theta\beta/\pi_1,$$

where  $\gamma = E(PR)$  is the expected positivity rate if the test is perfect. In addition, for time period  $t$  we have

$$E(\tilde{C}_{test}(t)|n(t)) = (\alpha - \beta)E(C_{test}(t)|n(t)) + \beta n(t), \tag{7}$$

where  $C_{test}(t)$  is the true number of infected persons among those tested. This suggests the estimate

$$\hat{C}_{test}(t) = (\tilde{C}_{test}(t) - \beta n(t))/(\alpha - \beta) \tag{8}$$

for the true number of infected persons among those tested in the time period.

PCR tests (also known as viral RNA or nucleic acid tests) for SARS-CoV-2 have very low false positive rates  $\beta$  (high specificity) but false negative rates  $\alpha$  can be substantial (e.g. see [Burstyn et al., 2020](#); [Mina et al., 2020](#)). In this case the infection and positivity rates are underestimated by the factor  $\alpha$ . More rapid diagnostic tests such as antigen tests ([Mina et al., 2020](#)) also have high specificity but lower sensitivity than a PCR test. The sensitivity of most tests depends on the time since infection ([Mina et al., 2020](#)) and to integrate this with variation in infection and testing rates over time, we can utilize the framework of Section 2.2. In the context of (4), the probability  $\pi^*(s, t)$  an infection that first becomes detectable on day  $t$  is confirmed on day  $s$  will depend on the test sensitivity  $\alpha(s - t)$  on day  $s$  as well as the testing rate  $\pi(s)$  on day  $s$ . It might also depend further on  $s - t$  because as symptoms become more severe, an individual is more likely to seek a test. Detailed modelling and estimation of  $\pi^*(s, t)$  requires more comprehensive data than are usually available. In the absence of such information we suggest an approximation  $\pi^*(s, t) = \alpha\pi_1(s)$ , where  $\alpha$  is the average sensitivity. The probability  $\pi^*(t)$  that an infection which becomes detectable on day  $t$  is eventually confirmed is then given by (4) as  $\alpha\pi_1(t, t + s_{max})$ . We discuss ascertainment rates further in Section 3.3.

### 3. Testing strategies

Jurisdictions usually target certain groups such as health care workers for regular testing, and provide criteria for who should be tested in the general population. A report from the European Centre for Disease Prevention and Control (ECDC) (2020) discussed five objectives of testing: A - controlling transmission, B - monitoring incidence and trends and assessing severity (of outcomes) over time, C - mitigating the impact of COVID-19 in healthcare and social care settings, D - rapid identification of clusters or outbreaks in specific settings, and E - preventing re-introduction of the virus into regions or countries with sustained control of the virus. Up to early 2021, most Canadian provinces focussed on testing individuals at the sites of outbreaks, health care workers, residents in long term care, returning travellers, plus persons in the general population with covid-related symptoms and possible close contact with a known case. However, there is scant evidence concerning ascertainment rates; random surveillance testing has been very limited (Waldner et al., 2021).

#### 3.1. Strategies

The testing objectives above require that infections be identified rapidly; this requires high values for daily rates  $\pi_1(t)$ , especially in places where transmission risk and potential severity of disease are high. We consider testing strategies by assuming a population, group or region is divided into strata that reflect testing criteria and infection levels. Our discussion will focus on accurate PCR tests; rapid tests are discussed at the end of Section 3.2 and in Section 6.

Suppose first the objective is to identify the largest possible number of cases across a population or equivalently, to maximize the population's ascertainment rate. For a given time period, if  $K$  strata can be ordered such that  $p_1 \geq p_2 \geq \dots \geq p_K$ , and if no further subdivision of strata according to likelihood of infection is possible, the best strategy for a total allotment of  $n$  tests would be to allot  $n_k$  tests to  $S_k (k = 1, \dots, K)$  as follows: test as many persons as possible from  $S_1$ , then if  $n_1 < n$ , test as many as possible from  $S_2$ , and so on. That is, we take  $n_1 = \min(n, N_1)$ ,  $n_2 = \max(0, \min(n - n_1, N_2))$ , ...,  $n_k = \max(0, \min(n - n_1 - \dots - n_{k-1}, N_k))$ . Since the true rates  $p_k$  are unknown, we could instead order strata according to their recent confirmed case rates.

More refined objectives would prioritize the identification of infections in critical strata with high transmission risk or potential disease severity. This would call for testing even when recent case rates were low. The ability to target testing so as to maximize  $\pi_{1k}(t)$  in the face of limited resources would remain crucial. Asymptomatic infections are a particular problem. To identify them effectively we must have other predictors of infection that can be used to define strata; main ones are close contact with a known case or exposure to populations with high case rates. If such predictors are not identified and if a substantial proportion of infections is asymptomatic, a high ascertainment rate will be impossible to obtain without very high testing rates.

Another potential objective is to estimate the proportion  $p_k$  of a stratum that is currently infected but not yet confirmed. As discussed earlier, a random sample from stratum  $k$  would enable this; in that case  $PR_k$  is an estimate of  $p_k$ . An estimate of  $p_k$  enables estimation of the testing rate  $\pi_{1k}$  under regular non-random diagnostic testing; this is discussed in section 3.3.

#### 3.2. Conditionally independent testing

An alternative approach to estimation of ascertainment rates when the population is stratified is through an assumption of conditionally independent testing. This requires that those tested within each stratum in a given period are a random sample from the stratum or, in the notation of Section 2, that  $\Delta$  is independent of  $Y$ , given  $z$ . We assumed this in our discussion of testing to maximize the number of infections identified in the preceding section. This assumption is restrictive, but may be a reasonable approximation when the strata partition the population into relatively homogeneous groups according to testing criteria, access to testing and likelihood of infection; persons with Covid-like symptoms would be in different strata than

**Table 2**

Weekly ascertainment, confirmed case and expected positivity rates for infection scenarios A, B, C and alternative testing schemes. The overall infection rate is  $p = 0.005$  and the weekly test rate is  $\theta = 0.02$ .

$k$ (stratum)	A		B		C	
	$\lambda_k$	$p_k$	$\lambda_k$	$p_k$	$\lambda_k$	$p_k$
1	0.01	0.30	0.01	0.10	0.02	0.10
2	0.10	0.01	0.10	0.02	0.08	0.02
3	0.89	0.00112	0.89	0.00225	0.90	0.00156
(i)	$\theta = (1.0, 0.1, 0)$ $\pi_1 = 0.62, r = 0.00310$ $\gamma = 0.155$		$\theta = (0.8, 0.12, 0)$ $\pi_1 = 0.208, r = 0.00104$ $\gamma = 0.052$		$\theta = (0.8, 0.05, 0)$ $\pi_1 = 0.336, r = 0.00168$ $\gamma = 0.084$	
(ii)	$\theta = (0.8, 0.08, 0.0045)$ $\pi_1 = 0.496, r = 0.00248$ $\gamma = 0.124$		$\theta = (0.5, 0.1, 0.0056)$ $\pi_1 = 0.142, r = 0.00071$ $\gamma = 0.0356$		$\theta = (0.5, 0.1, 0.0022)$ $\pi_1 = 0.232, r = 0.00116$ $\gamma = 0.058$	

those without symptoms. The positivity rate  $PR_k$  in stratum  $k$  then estimates the infection rate  $p_k$ , and  $\pi_{1k} = \theta_k$ , so from (6) the overall ascertainment rate is estimated by

$$\hat{\pi}_1 = \frac{\hat{r}}{\hat{p}} = \frac{\sum_k PR_k \theta_k \lambda_k}{\sum_k PR_k \lambda_k}. \tag{9}$$

Exact stratum sizes  $N_k$  are typically unknown so the  $\lambda_k$  and  $\theta_k$  are in that case estimates.

To achieve a high overall ascertainment rate it is necessary to have high testing rates in strata with high positivity or case rates, particularly when the strata are large. This was the basis for the optimal testing allocation discussed in Section 3.1. More generally, the relationship (6) with  $\pi_{1k} = \theta_k$  shows the effects of stratum-specific random testing rates on overall ascertainment. For illustration we consider three simplified scenarios, each involving three strata and giving an overall unconfirmed infection rate of  $p = 0.005$  (5 cases per thousand persons); this rate has been plausible in many countries at various points in the pandemic. Table 2 shows the values of  $\lambda_k$  and  $p_k$  for each scenario and gives weekly ascertainment, confirmed case and expected positivity rates for two sets of testing rates. In each scenario both sets of testing rates give an average weekly testing rate of 0.02, or 2 tests per hundred persons. This rate is similar to the rates in Ontario data discussed in Section 5. For each scenario the rates (i) focus more heavily on the stratum with a high infection rate, and do not test anyone in the stratum with a small infection rate, whereas (ii) is more balanced. The notable implications from these scenarios are that for a high ascertainment rate  $\pi_1$  we need the strata to have widely varying infection rates and for testing to focus on the strata with high rates, as suggested earlier. The decrease in  $\pi_1$  when this does not occur indicates the inevitable tradeoff between identifying the largest number of infections and conducting precautionary surveillance testing in areas with low rates.

Rapid testing is being increasingly deployed, with persons who test positive subsequently given a PCR test. This strategy increases testing rates  $\theta_k$  at the cost of reduced sensitivity  $\alpha$ . Assuming both the rapid test and PCR test have false positive rates close to zero, the confirmed case rate  $r$  is reduced by the factor  $\alpha$  and from (6) the effective ascertainment rate is reduced by the same factor. If testing is random within strata then  $\gamma_k = p_k$  but in any case, the tradeoff between decreased  $\alpha$  and increased testing rates  $\theta_k$  can be determined. If  $\alpha = 0.5$ , for example, testing rates would need to be doubled to achieve the same ascertainment rate as when  $\alpha = 1$ .

### 3.3. Combining information

To estimate ascertainment rates for regular diagnostic testing we need information on the true numbers of infections in a population, which can then be compared with cases identified by regular testing. One approach has been to use unverifiable modelling assumptions involving transmission or fatality rates (e.g. Dougherty et al., 2021; Shaman, 2021), but the only reliable method is to estimate true infection rates and total cases through random testing. The United Kingdom is exemplary in this regard, with two sets of repeated surveys based on PCR tests: one is conducted by the Office of National Statistics with Oxford University (Office of National Statistics U.K., 2020) and a second by Imperial College London and Ipsos MORI (Riley et al., 2021). These surveys provide estimates of the infection rate in the general population over time, with breakdowns by region, age and other factors. In Canada, random testing has been done only occasionally in certain regions (Waldner et al., 2021).

An estimate  $\hat{p}(t)$  of the infection rate  $p(t)$  for some population or group in time period  $t$  can be combined with data from regular diagnostic testing so as to estimate the positive testing rate  $\pi_1(t)$  for regular testing in period  $t$ . By (2) we have

$$\pi_1(t) = \hat{\gamma}(t) \hat{\theta}(t) / \hat{p}(t) = CCR(t) / \hat{p}(t) = C_{test}(t) / \hat{C}_{pop}(t), \tag{10}$$

where  $CCR(t) = C_{test}(t)/N(t)$  is from regular testing and  $\hat{C}_{pop}(t) = N(t)\hat{p}(t)$ . We assume here that tests are perfectly accurate. When this is not the case we can replace  $C_{test}(t)$  with the estimate (8). Since testing patterns usually vary by day of the week, it is best to consider weekly or bi-weekly rates.

Uncertainty concerning  $\pi_1(t)$  arises from random variation in  $\hat{p}(t)$  and in regular testing. This can be addressed through either frequentist or Bayesian procedures. The following frequentist approach which uses normal approximations for the estimators is simple to apply. Bayesian approaches can in principle deal more comprehensively with uncertainty in all its forms, including uncertainty around test sensitivity and specificity. However, it has more difficulty dealing with features such as complex survey designs, whose analysis utilizes estimating functions rather than likelihood. We condition on the number of regular tests  $n(t)$ , in which case  $C_{test}(t)$  depends on  $\gamma(t)$ . The estimators  $\hat{p}(t)$  and  $\hat{\gamma}(t) = PR(t)$  are asymptotically independent and so the asymptotic variance of  $\log \hat{\pi}_1(t)$  is (Boos & Stefanski, 2013)

$$Var(\log \hat{\pi}_1(t)) = \gamma(t)^{-2} Var(\hat{\gamma}(t)) + p(t)^{-2} Var(\hat{p}(t)).$$

The variance of  $\hat{p}(t)$  will depend on the sample design for random testing, which may be complex (Wu & Thompson, 2020). If  $C_{test}(t)$ , given  $n(t)$ , is a binomial random variable then  $Var(\hat{\gamma}(t)) = \gamma(t)(1 - \gamma(t))/n(t)$ . The binomial assumptions here might not be completely satisfactory due to population heterogeneity and clustering affecting testing in a given period; stratification as described in Section 2.3 can be used to mitigate such effects, provided stratum - specific estimates  $\hat{p}_k$  are available.



Regular random testing is rare, and a more feasible objective than estimating time-dependent ascertainment rates is to estimate the average ascertainment rate over a given period of time by estimating the true number of infections and comparing that with confirmed cases. The true number of infections could be estimated from seroprevalence surveys earlier in the pandemic, and this was illustrated for the province of Ontario in Section 2.2, although the accuracy of such estimates depends on antibody persistence and the sensitivity of antibody tests (Accorsi et al., 2021). Now that vaccination is widespread, this approach is no longer widely feasible. An alternative approach that uses information about symptomatic infection rates could be applied, if appropriate information is collected. In particular, if the proportion  $q_S$  of infections that are symptomatic were known, and if the number of confirmed cases  $C_{test}$  in a time period were split into symptomatic cases  $C_{test-symp}$  and asymptomatic cases, then the overall ascertainment rate for the period could be estimated as  $\pi^* = C_{test-symp} \pi_S^* / C_{test}$ , where  $\pi_S^*$  is the ascertainment rate for symptomatic infections. The rate  $\pi_S^*$  is unknown, but might often be close to 1.

The proportion  $q_S$  of infections that are symptomatic varies with age and other factors, and there have been wide ranges in estimates from different studies (Yanes-Lane et al., 2020). For the 2020 Canadian seroprevalence survey (Ab-C Study Investigators, 2021), the proportion of “definitely” seroprevalent adults who were estimated to be symptomatic was 0.32. A similar alternative approach to estimating ascertainment would be to replace “symptomatic” with “hospitalized” in the preceding discussion. In this case  $q_S$  represents the proportion of infections that lead to hospitalization, and in this case  $\pi_S^* = 1$  is reasonable. However, to estimate the hospitalization rate  $q_S$  we once again need estimates of the true numbers of infections, and not just confirmed cases. We are thus lead back to the need for some level of random testing, or exhaustive testing. For Canada, Dougherty et al., 2021 used another version of this approach with death replacing hospitalization, but failed to address the difference between infection and case fatality rates.

#### 4. Analysis of testing data

Studies have linked factors such as such as race, population density, population mobility, income, and work environments to confirmed case rates, but usually they do not account for testing and ascertainment rates. Analysis of testing rates  $\theta(t)$  can provide additional insight. Section 3 demonstrated the importance of high diagnostic testing rates in groups or regions with high infection rates, and we consider two types of analysis: comparisons of testing rates across specified regions or groups, and analysis of variation in testing over time. In each case the number of tests  $n(t)$  for a specific group and time period is a response variable, and  $z(t)$  denotes a vector of observable covariates associated with  $n(t)$ . Unobserved factors such as the underlying symptomatic infection rate and individual behaviour also affect testing rates, and to accommodate heterogeneity in counts from such factors, we consider negative binomial mixed Poisson models (Lawless, 1987), which are widely used for modelling count data (Venables & Ripley, 2002, Section 7.4). We use the log-linear form, with the mean and variance of  $n(t)$  given  $z(t)$  and the population size  $N(t)$  as

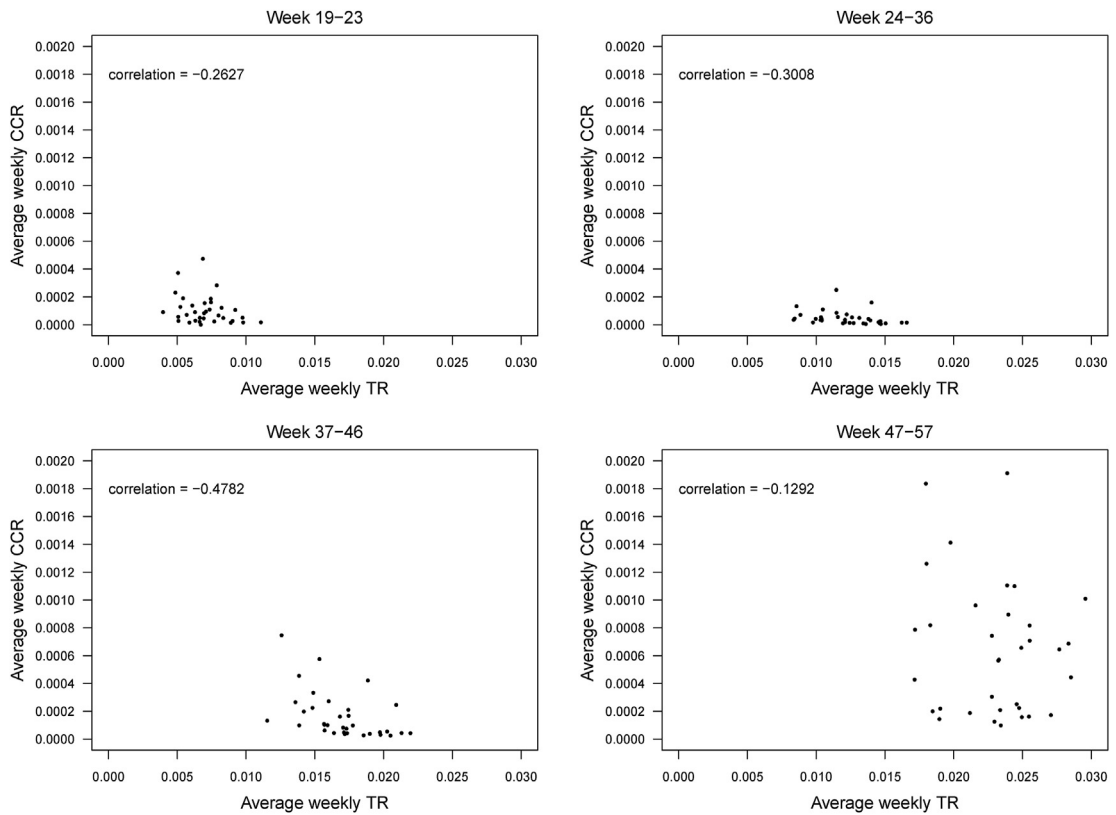
$$\begin{aligned} E\{n(t)|z(t)\} &= \mu(t; z(t)) = N(t)\exp(\beta_0 + \beta'z(t)), \\ \text{Var}\{n(t)|z(t)\} &= \mu(t; z(t)) + \tau\mu(t; z(t))^2, \end{aligned}$$

where  $\beta$  is a vector of regression coefficients and  $\tau > 0$  is the dispersion parameter; vectors are written in column form. In Section 5 we use the R function `glm.nb` in the MASS library to fit models (Venables & Ripley, 2002) for different public health regions in Ontario. Ideally, it would be best to examine separately testing in targeted groups (e.g. health care workers, long term care residents, industrial locations) and in the general population, but published data do not provide breakdowns that allow this. Testing for different age groups is also of interest; in Ontario, tests by age group are available for the province as a whole but are not provided for separate regions.

#### 5. Analysis of ontario data

We analyze data for Ontario, Canada, focusing on variation in testing rates over time and across regional public health units (PHUs). Daily data on new reported cases of Covid-19, numbers of tests completed and positivity rates are available for download at <https://data.ontario.ca/dataset>. Data for the 34 separate PHUs in the province have been available since late April 2020. We note a few points: (i) The number of tests reported on a given day is the number completed during the previous 24 h, but specimens could have been taken from some of the individuals up to a few days earlier. (ii) The number of tests completed on a given day may be slightly higher than the number of persons tested, as some tests may need to be repeated. (iii) Positivity rates are calculated as the number of positive tests divided by the total number of tests. (iv) Daily counts show strong day of the week effects: the numbers of tests are smallest on Tuesdays as their specimens were mostly collected over the previous weekend and processed on Monday; the numbers are largest on Thursdays, Fridays and Saturdays. The daily positivity rates vary inversely to the numbers of tests. The reasons for this are not fully clear, but would be expected from (2) if  $\pi_1(t)$  and  $p(t)$  were fairly stable from day to day.

We focus on weekly counts and rates for analysis. Weeks run from Monday to Sunday and we consider test counts  $n(t)$  from 2020 week 22 (May 25–31) to 2021 week 4 (January 25–31) as responses in regression analyses; in addition, data from

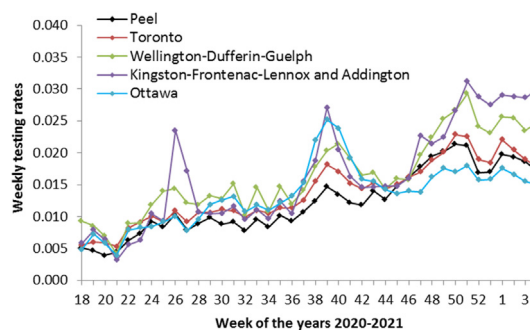


**Fig. 1.** Plots of average weekly testing rate (TR) versus average weekly confirmed case rate (CCR) for 34 Ontario public health units and four time periods in 2020–2021.

weeks 20 and 21 are used as covariates. Rapid tests have affected testing processes since February 2021, but rapid test results are not reported. We do not include test counts since then in our analysis, but discuss them briefly at the end of this section.

Weekly testing rates have increased three-to four-fold in most PHUs since May 2020. We compare testing rates across PHUs and association with confirmed case rates by splitting time into four periods during which testing rates were reasonably stable: weeks 19–23, 24–36, 37–46, and week 47–2021 week 4. For each period, Fig. 1 shows the average weekly testing rate against the average confirmed case rate for each PHU. The median testing rate increased from about 0.07 (7 tests per thousand persons) in weeks 19–23 to about 0.024 in weeks 47–57. Testing rates within each time period vary on average about two-fold across the 34 PHUs. Confirmed case rates vary much more widely across PHUs and across time, particularly from week 37 onward. Within each time period there is a moderate negative correlation between a PHU’s average TR and average CCR; that is, regions with higher case rates tend to have lower testing rates.

Further analysis of temporal variation in testing rates is based on negative binomial models. For illustration we consider five PHUs, three with large and two with smaller populations: Toronto, Peel, Wellington-Dufferin-Guelph (WDG), Kingston-



**Fig. 2.** Weekly testing rates for five Ontario public health units.



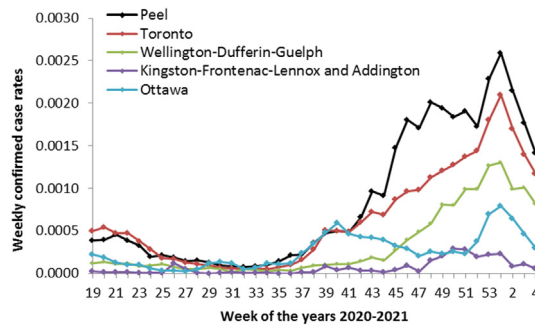


Fig. 3. Weekly confirmed case rates for five Ontario public health units.

Frontenac-Lennox-Addington (KFLA), and Ottawa. Estimated 2020 population sizes for these PHUs are respectively  $N = 1,519,200; 3,109,676; 304,360; 209,023; \text{ and } 1,111,773$ . Figs. 2 and 3 show weekly variation in testing rates and confirmed case rates for each PHU. Peel and Toronto have had by far the highest case rates but along with Ottawa, the lowest testing rates. Testing rates for all PHUs increased approximately three-fold since May 2020 but have levelled off since December 2020. Plots of weekly positivity rates follow a similar pattern to case rates but with more pronounced high rates in the early weeks, when testing rates were low. Fig. 2 shows considerable week to week variation in test counts for PHUs and to examine the effects of time-varying factors, we fitted negative binomial regression models for weekly test counts described in Section 4. The expected value of  $n(t)$  given a covariate vector  $z(t)$  and population size  $N(t)$  takes the form

$$\mu(t; z(t)) = N(t)\exp(\alpha_0 + \alpha_1 t + \beta' z(t)). \tag{11}$$

To consider the effect of recent confirmed case rates we defined covariates  $z_1(t) = CCR(t - 1)$  and  $z_2(t) = CCR(t - 1) - CCR(t - 2)$ ; the time origin  $t = 0$  corresponds to week 20. We also defined binary covariates to reflect holiday periods and changes in testing criteria, as follows: a covariate  $Holidays(t)$  that equals 1 for weeks 21, 27, 32, 36, 42, 52–53 and 0 otherwise, reflecting Victoria Day, July 1, August, Labour Day, Thanksgiving and Christmas holiday periods; a covariate  $x_1(t) = 1$  for weeks 22–24 and 0 otherwise; a covariate  $x_2(t) = 1$  for weeks 37–40 and 0 otherwise; a covariate  $x_3(t) = 1$  for weeks 52–53 and weeks 1–4 of 2021 and 0 otherwise. Covariates  $x_1$  and  $x_2$  correspond to two periods where testing criteria were temporarily relaxed by the province; the rationale for  $x_3$  is to reflect the effect of the Christmas - New Year period and the transition into 2021. These covariates are included as covariates  $z_3(t)$  to  $z_6(t)$  in (11). Other factors such as lockdowns were also considered, but not found significant. We considered alternative functions of  $t$  to represent the temporal increase in testing rates, but found the form in (11) to give the smallest value for AIC among parametric models considered.

Table 3 shows estimates of model parameters for the five PHUs. A high case rate in the preceding week ( $z_1$ ) has a significant effect on the current testing rate in all PHUs but aside from Ottawa, the change over the previous two weeks ( $z_2$ ) is not significant. Holidays is not significant but the effects of eased testing criteria ( $x_1, x_2$ ) and the transition into 2021 ( $x_3$ ) are. We refitted models with  $z_2$  and Holidays dropped and calculated the likelihood ratio statistics (twice the difference in log likelihoods) for a test of no combined effect. Based on chisquare with two degrees of freedom, p-values for the five regions were 0.193 (Peel), 0.438 (Toronto), 0.593 (WDG), 0.452 (KFLA) and 0.032 (Ottawa); as expected, only Ottawa gave a small p-value. The regression coefficients for the other covariates were similar in the full and reduced models. Paradoxically, eased criteria in weeks 22–24 was mildly associated with a decrease in testing. The temporary easing of criteria in weeks 37–40 (covariate  $x_2$ )

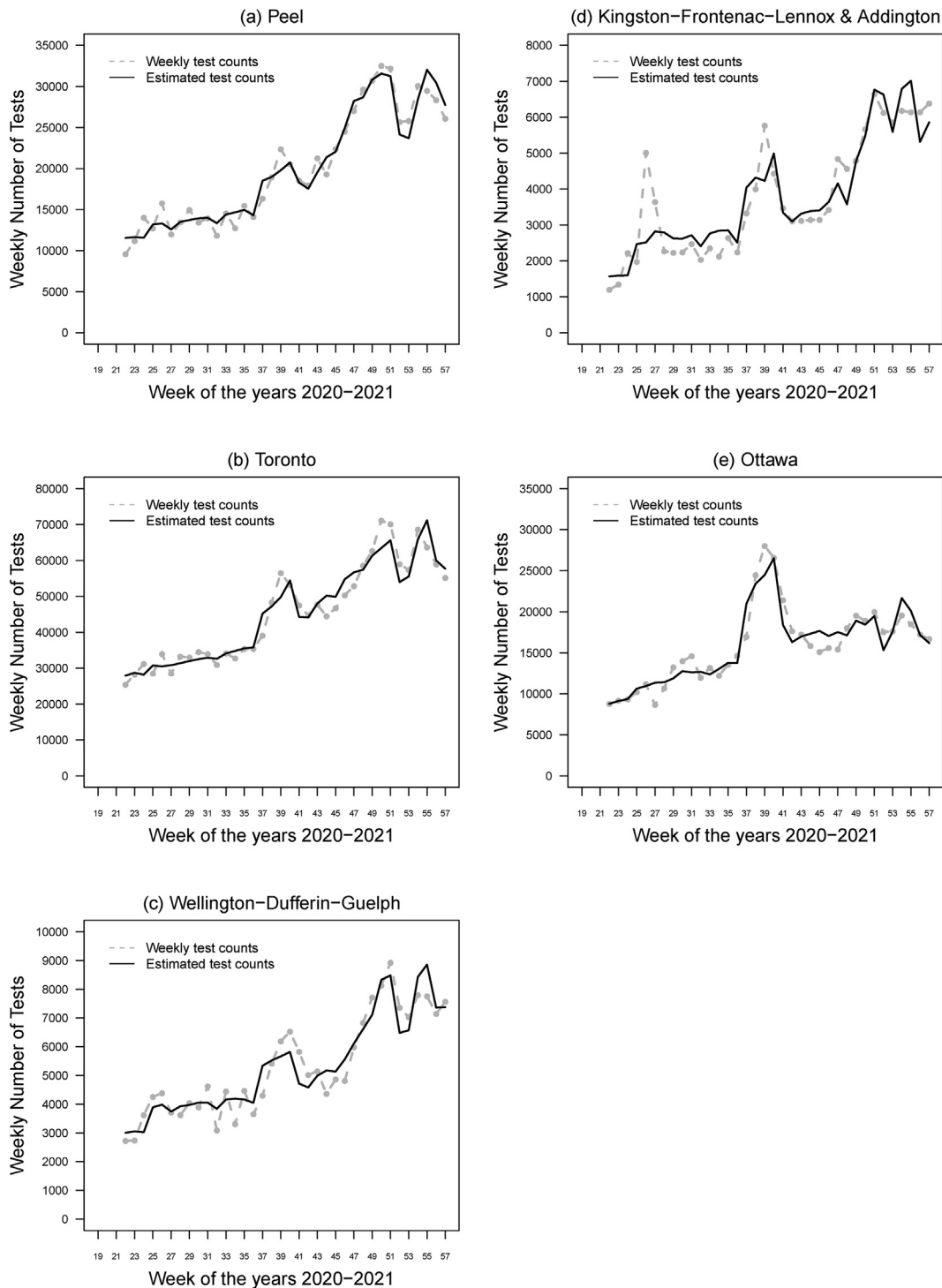
Table 3

Estimated regression coefficients in negative binomial regression models for weekly numbers of tests in five Ontario public health units over May 2020 to January 2021.

Covariate	Estimate (SE <sup>a</sup> )				
	Peel	Toronto	WDG	KFLA	Ottawa
Intercept	-4.95 (0.05)	-4.78 (0.04)	-4.50 (0.06)	-4.55 (0.11)	-4.69 (0.05)
Time <sup>b</sup>	0.02 (0.004)	0.02 (0.003)	0.01 (0.004)	0.01 (0.006)	0.02 (0.003)
$z_1(t)$	0.29 (0.05)	0.25 (0.07)	0.68 (0.16)	2.01 (0.76)	0.28 (0.13)
$z_2(t)$	-0.08 (0.08)	0.13 (0.10)	-0.04 (0.24)	0.20 (0.98)	0.53 (0.19)
$x_1(t)$	-0.14 (0.07)	-0.09 (0.07)	-0.24 (0.09)	-0.43 (0.15)	-0.14 (0.07)
$x_2(t)$	0.16 (0.05)	0.19 (0.05)	0.22 (0.07)	0.33 (0.12)	0.39 (0.06)
$x_3(t)$	-0.22 (0.05)	-0.22 (0.06)	-0.34 (0.12)	0.15 (0.17)	-0.25 (0.06)
Holidays	-0.06 (0.04)	-0.02 (0.04)	-0.05 (0.05)	-0.13 (0.10)	0.02 (0.05)
Dispersion ( $\tau^{-1}$ )	155.5 (37.0)	206.8 (48.9)	82.5 (19.8)	24.0 (5.7)	112.7 (26.7)

<sup>a</sup> Standard error.

<sup>b</sup> Time in weeks.



**Fig. 4.** Testing rates and fitted regression curve for (a) Peel Region Public Health Unit, (b) City of Toronto Public Health Unit, (c) Wellington-Dufferin-Guelph Public Health Unit, (d) Kingston-Frontenac-Lennox-Addington Public Health Unit, and (e) Ottawa Public Health Unit, week 19, 2020 to week 4, 2021.

was however associated with increased testing. We note that schools began their fall re-opening in week 37 and province-wide data on testing by age group (Fisman et al., 2020; Government of Ontario, 2021b) show a large temporary increase in testing for children aged 9 and younger during this period, and a smaller increase for those aged 10–19. The Christmas and New Year period saw a significant decrease in testing for all except the KFLA region.

Fig. 4 (a) - (e) show the weekly test counts  $n(t)$  and the estimated expected values  $\hat{\mu}(t; z(t))$  under the fitted models for the five PHUs. The regression models track the actual test counts well. There is evidence of residual negative autocorrelation in counts for certain time periods; this may be due to backlogs related to laboratory capacity. Some regions show temporary aberrations, which may be due to local initiatives. In weeks 26 and 27, KFLA's testing rate was well above the predicted value; this was associated with an outbreak in a Kingston nail salon in weeks 25–26. In week 47 Ontario began to allow testing outside of laboratories (Hicks Morley, 2020); this may be associated with the slightly more rapid than predicted increase in testing over weeks 47–50 in some PHUs.

Instead of higher testing rates in regions with high case rates, we see a negative correlation with confirmed case rates. The highest case rates were mainly in regions with large populations, and further analysis could consider factors such as population density, socioeconomic conditions, and ease of access to testing. At present, public PHU-level data to support such analysis is unavailable. A comparison of ascertainment rates is crucial to truly understand differences in testing rates across PHUs, but since little random testing has been done, true infection and ascertainment rates in PHUs are very uncertain. As discussed earlier, Ontario-wide estimates based on serological surveys (Ontario Agency for Health Protection and Promotion (Public Health Ontario), 2020; Ab-C Study Investigators, 2021) suggest average ascertainment rates of 0.25 or less up to September 2020.

Since January 2021 the province-wide weekly testing rate has been stable at approximately 0.02, but the negative association between testing and confirmed case rates has eased. Testing rates still vary two-fold across regions, with Peel, Toronto and Ottawa among those with the five lowest rates. Confirmed case rates during the pandemic's third wave in April 2021 have once again varied widely across regions.

## 6. Conclusion

In Ontario, per capita testing rates have consistently varied approximately two-fold across regions at any given time, whereas confirmed case rates varied by factors of 20 or more. If testing were targeting higher risk areas, we would expect higher testing rates in regions with high case rates; instead there is a moderate negative association between testing and confirmed case rates across regions. Since January 2021 this association has eased, but regions with the highest confirmed case rates still tend to have lower testing rates. Exceptions are often for PHUs with smaller populations, which are perhaps more able to increase testing proportionally when there is an upsurge in confirmed cases. Investigation of testing rate variation within PHUs would also be of interest, but requires more detailed data than are available. We note that for the Toronto PHU, dashboard maps show wide variation in case rates across neighbourhoods, and much less variation in testing rates. Neighbourhoods with high case rates tend to be ones with features such as a high proportion of essential workers, lower income levels and more crowded living conditions. A more detailed analysis that relates testing rates to demographic and socioeconomic conditions, ease of access to testing, and behavioural factors would be valuable.

Rapid antigen tests (Larremore et al., 2021; Mina et al., 2020) are increasingly used to monitor high risk settings such as workplaces and schools but to be effective they must be repeated frequently. Testing strategies for schools have been an area of debate in some countries: some argue for frequent rapid testing but others point out pitfalls when tests have variable and sometimes low sensitivity (Deeks et al., 2021). These discussions reflect the inevitable tradeoffs between sensitivity and frequency of testing in any setting; once again, however, lack of information on ascertainment rates hampers comparisons.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

Research was supported in part by Discovery Grant RGPIN-2017-04055 to JFL from the Natural Sciences and Engineering Research Council of Canada. The data analyzed in the paper were provided by the Province of Ontario under the Open Government Licence - Ontario (url <https://www.ontario.ca/page/open-government-licence-ontario>). The authors thank Ker-Ai Lee of the University of Waterloo for computing assistance.

## References

- Accorsi, E., Qui, X., Rumpler, E., Kennedy-Shaffer, L., Kahn, R., Joshi, K., Goldstein, E., Stensrud, M. J., Niehus, R., Cevik, M., & Lipsitch, M. (2021). How to detect and reduce potential sources of bias in studies of SARS-CoV-2 and COVID-19. *European Journal of Epidemiology*, 36, 179–196.
- Boos, D., & Stefanski, L. (2013). *Essential statistical inference: Theory and methods*. New York, NY: Springer.
- Burstyn, I., Goldstein, N., & Gustafson, P. (2020). It can be dangerous to take epidemic curves of COVID-19 at face value. *Canadian Journal of Public Health*, 111, 397–400.
- Deeks, J., Gill, M., Bird, S., Richardson, S., & Ashby, D. (2021). Covid-19 INNOVA testing in schools: don't just test, evaluate Accessed <https://blogs.bmj.com/bmj/2021/01/12/covid-19-innova-testing-in-schools-dont-just-test-evaluate/>. (Accessed 12 January 2021).
- Dougherty, B., Smith, B., Carson, C., & Ogden, N. (2021). Exploring the percentage of COVID-19 cases reported in the community in Canada and associated case fatality ratios. *Infectious Disease Modelling*, 6, 123–132.
- European Centre for Disease Prevention and Control (ECDC). (2020). *COVID-19 testing strategies and objectives*. Stockholm: Technical Report ECDC.

- Fisman, D., Greer, A., Hillmer, M., O'Brien, S., Drews, S., & Tuite, A. (2020). COVID-19 case-age distribution: Correction on differential testing by age. *medRxiv*, p. 2020.09.15.21252540.
- Government of Alberta. (2021). COVID-19 info for Albertans accessed 2021-02-19 <https://www.alberta.ca/coronavirus-info-for-albertans.aspx>.
- Government of Ontario. (2021a). All Ontario: Case numbers and spread accessed <https://covid-19.ontario.ca/data>. (Accessed 19 February 2021).
- Government of Ontario. (2021b). COVID-19: Epidemiologic summaries from public health Ontario accessed <https://covid-19.ontario.ca/covid-19-epidemiologic-summaries-public-health-ontario>. (Accessed 19 February 2021).
- Investigators, A.b-C. S. (2021). COVID seroprevalence, symptoms and mortality during the first wave of SARS-CoV-2 in Canada. *medRxiv*, 2021.03.04.20193862.
- Larremore, D., Wilder, B., Lester, E., Shehata, S., Burke, J., Hay, J., Tambe, M., Mina, M., & Parker, R. (2021). Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Science Advances*, 7(1), Article eabd5393.
- Lawless, J. (1987). Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*, 15, 209–225.
- Marchand-Sénécal, X., Kozak, R., Mubareka, S., Salt, N., Gubbay, J., Eshaghi, A., Aalen, V., Li, Y., Bastien, N., Gilmour, M., Ozaldin, O., & Leis, J. (2020). Diagnosis and management of first case of COVID-19 in Canada: Lessons applied from SARS-CoV-1. *Clinical Infectious Diseases*, 71, 2207–2210.
- Mina, M., Parker, R., & Larremore, D. (2020). Rethinking Covid-19 test sensitivity - a strategy for containment. *New England Journal of Medicine*, 383, Article e120.
- Morley, H. (2020). Insights from Hicks Morley. <https://canlii.ca/t/t0dw>. (Accessed 8 April 2021).
- Office of National Statistics U.K. (2020). COVID-19 infection survey (Pilot): Methods and further information Accessed <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurvey-pilot-methods-and-further-information>. (Accessed 21 September 2020).
- Ontario Agency for Health Protection and Promotion (Public Health Ontario). (2020). *Covid-19 seroprevalence in Ontario: July 4, 2020 to July 31, 2020. Technical report*. Toronto, ON: Queens Printer for Ontario.
- Pullano, G., Di Domenico, L., Sabbatini, C., Valdano, E., Turbelin, C., Debin, M., Guerrisi, C., Kengne-Kuetché, C., Souty, C., Hanslik, T., Blanchon, T., Boëlle, P.-Y., Figoni, J., Vaux, S., Campese, S., Bernard-Stoecklin, C., & Colizza, V. (2021). Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature*, 590, 134–139.
- Riley, S., W, H., Eales, O., Walters, C., A, K. E. C., Atchison, C., Fronterre, C., Diggle, P., Ashby, D., Donnelly, C., Cooke, G., Barclay, W., Ward, H., Darzi, A., & Elliott, P. (2021). REACT-1 round 8 interim report: SARS-CoV-2 prevalence during the initial stages of the third national lockdown in England. *medRxiv*, (p. 2021.01.20.21250158).
- Shaman, J. (2021). An estimation of undetected COVID cases in France. *Nature*, 590, 38–39.
- Venables, W., & Ripley, B. (2002). *Modern applied statistics with S* (4th ed.). New York, NY: Springer.
- Waldner, D., Harrison, R., Johnstone, J., Saxinger, L., Webster, D., & Sligl, W. (2021). The epidemiology of COVID-19 in Canada 2020: The pre-vaccine era accessed <https://www.rsc-src.ca/en/research-and-reports/covid-19-policy-briefing/epidemiology-covid-19-in-canada-in-2020-pre-vaccine>. (Accessed 10 March 2021).
- Wu, C., & Thompson, M. (2020). *Sampling theory and practice*. New York, NY: Springer.
- Yanes-Lane, M., Winters, N., Fregonese, F., Bastos, M., Perlman-Arrow, S., Campbell, J., & Menzies, D. (2020). Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: A systematic review and meta-analysis. *PLoS One*, 15, Article e0241536.