1	Improving estimates of epidemiological quantities by combining
2	reported cases with wastewater data: a statistical framework
3	with applications to COVID-19 in Aotearoa New Zealand

4	Leighton M. Watson ¹ , Michael J. Plank ¹ , Bridget A. Armstrong ² , Joanne R. Chapman ² ,					
5	Joanne Hewitt ² , Helen Morris ² , Alvaro Orsi ² , Michael Bunce ^{2,3} , Christl A. Donnelly ^{4,5} ,					
6	and Nicholas $Steyn^4$					
7	¹ School of Mathematics and Statistics, University of Canterbury, New Zealand					
8	² Institute of Environmental Science and Research Ltd, New Zealand					
9	³ Department of Conservation, New Zealand					
10	⁴ Department of Statistics, University of Oxford, United Kingdom					
11	⁵ Pandemic Sciences Institute, University of Oxford, United Kingdom					

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract

Background: Timely and informed public health responses to infectious diseases such as COVID-19 necessitate reliable information about infection dynamics. The case ascertainment rate (CAR), the proportion of infections that are reported as cases, is typically much less than one and varies with testing practices and behaviours, making reported cases 16 unreliable as the sole source of data. The concentration of viral RNA in wastewater samples provides an alternate measure of infection prevalence that is not affected by human behaviours. Here, we investigated how these two data sources can be combined to inform estimates of the instantaneous reproduction number, R, and track changes in the CAR over time.

Methods: We constructed a state-space model that we solved using sequential Monte 22 Carlo methods. The observed data are the levels of SARS-CoV-2 in wastewater and reported 23 case incidence. The hidden states that we estimate are R and CAR. Model parameters are 24 estimated using the particle marginal Metropolis Hastings algorithm. 25

Findings: We analysed data from 1 January 2022 to 31 March 2023 from Aotearoa 26 New Zealand. Our model estimates that R peaked at 2.76 (95% CrI 2.20, 3.83) around 18 27 February 2022 and the CAR peaked around 12 March 2022. Accounting for reduced CAR, 28 we estimate that New Zealand's second Omicron wave in July 2022 was similar in size to the 29 first, despite fewer reported cases. We estimate that the CAR in the BA.5 Omicron wave 30 in July 2022 was approximately 50% lower than in the BA.1/BA.2 Omicron wave in March 31 2022. The CAR in subsequent waves around November 2022 and April 2023 was estimated 32 to be comparable to that in the second Omicron wave. 33

Interpretation: This work on wastewater-based epidemiology (WBE) can be used 34 to give insight into key epidemiological quantities. Estimating R, CAR, and cumulative 35 number of infections provides useful information for planning public health responses and 36 understanding the state of immunity in the population. This model is a useful disease 37 surveillance tool, improving situational awareness of infectious disease dynamics in real-38 time, which may be increasingly useful as intensive pandemic surveillance programmes are 39 wound down. 40

Funding: New Zealand Ministry of Health, Department of Prime Minister and Cabinet, 41 the Royal Society Te Apārangi, Imperial College London, and University of Oxford. 42

12

13

14

15

17

18

19

20

21

⁴³ Research in Context

44 Evidence before this study

There has been a substantial increase in the number of publications focusing on wastewater-45 based epidemiology (WBE) in recent years, particularly during the COVID-19 pandemic. We 46 searched PubMed for "wastewater based epidemiology" and found fewer than 10 papers per 47 year prior to 2014 with a drastic increase to 463 in 2022. Approximately 52% of the WBE 48 publications are related to COVID-19 ("wastewater based epidemiology" AND ("SARS-CoV-49 2" OR "COVID-19")). Many studies have focused on detecting SARS-CoV-2 in wastewater 50 systems but only 10 have estimated the reproduction number ("wastewater based epidemiology" 51 AND ("SARS-CoV-2" OR "COVID-19") AND "reproduction number"). No previous work has 52 combined WBE with reported case data to estimate (relative) case ascertainment rate ("waste-53 water based epidemiology" AND ("SARS-CoV-2" OR "COVID-19") AND "case ascertainment 54 rate"). Previous work has estimated the reproduction number from reported cases assuming 55 constant under-ascertainment but the issue of time-varying case ascertainment has not yet been 56 addressed, except to demonstrate the effect of a pre-determined change in case ascertainment. 57

58 Added value of this study

We present a model that, for the first time, enables reported case information to be combined with wastewater data to estimate epidemiology quantities. This work further demonstrates the utility of WBE; the reproduction number can be estimated in the absence of reported case information (although results are more reliable when case data are included), and wastewater data include information that, when combined with case data, can be used to estimate the time-varying relative case ascertainment rate.

⁶⁵ Implications of all the available evidence

In order to make informed and timely public health decisions about infectious diseases, it is important to understand the number of infections in the community. WBE provides a useful source of data that is not impacted by time-varying testing practices. Wastewater data can be quantitatively combined with case information to better understand the state of an epidemic. In order to determine the absolute case ascertainment rate (rather than the relative rate calculated in this work), there is a need for infection prevalence surveys to calibrate model results against.

72 **1** Introduction

Understanding and predicting the trajectory of infectious diseases is important in planning an 73 effective public health response. Reported case data depend heavily on testing modalities and 74 practices which typically change over time, resulting in considerable uncertainty in the case 75 ascertainment rate (CAR; the fraction of infections that are officially reported). During the 76 COVID-19 pandemic, many countries relied primarily on symptom-based testing programmes 77 to inform situational awareness and public health responses. In Aotearoa New Zealand, the CAR 78 for COVID-19 has been influenced by factors such as access to testing, a shift from healthcare 79 worker-administered polymerase chain reaction (PCR) tests to self-administered rapid antigen 80 tests (RATs), reduction in rates of symptomatic and severe disease due to rising population 81 immunity, relaxation of testing requirements and recommendations, and/or lack of perceived 82 need to test or 'pandemic fatigue' [1, 2, 3]. As a result, over time, officially reported cases of 83 COVID-19 have become a less reliable measure of levels of SARS-CoV-2 infection. 84

Data on hospital admissions and deaths are more consistent and are less affected by testing 85 practices and behavioural change than reported cases but are subject to additional delays [4] 86 that limit their usefulness for understanding disease dynamics. Infection prevalence surveys [5] 87 that aim to regularly test a representative sample of the population are the gold-standard for 88 tracking the spread of an infectious disease, but these surveys are resource intensive, making 89 them harder to justify as countries move out of the acute phase of the pandemic. For example, 90 regular SARS-CoV-2 infection prevalence surveys in the UK [6] have now been wound down 91 and there are no current plans for similar surveys in New Zealand. 92

Wastewater surveillance, where levels of SARS-CoV-2 RNA in wastewater samples are mea-93 sured, can provide additional data on the prevalence of the virus that are unaffected by individ-94 ual testing and self-reporting behaviours. Wastewater surveillance (also known as wastewater-95 based epidemiology or WBE) also has the potential to contribute to an integrated global network 96 for disease surveillance [7, 8, 9]. These data, however, can be highly variable and subject to other 97 biases, such as rainwater dilution, sampling methodologies, and changing locations of selected 98 sampling sites. To realise this potential, appropriate models and analytical tools are needed 99 to deliver epidemiological insights from raw data. Previous work by [10] and [11] estimated R_t 100 from wastewater data, compared the results with estimates derived from case data, and found 101 that WBE can provide independent and reliable estimates of R_t . 102

¹⁰³ Semi-mechanistic models based on the renewal equation are a popular method for epidemic ¹⁰⁴ forecasting and estimation of the instantaneous reproduction number [12, 13, 14]. Such methods ¹⁰⁵ are robust to constant under-ascertainment of cases, but may be biased by rapid changes in

CAR and cannot provide any information about the total number of infections. In this paper, 106 we extend the renewal equation framework [12, 13, 14] for reproduction number estimation 107 to incorporate wastewater time-series data. The model treats the instantaneous reproduction 108 number and CAR as hidden states and reported cases and quantity of viral RNA in wastewater 109 as observed states. We use a sequential Monte Carlo approach to infer the hidden states. We 110 apply the model to national data from Aotearoa New Zealand on reported COVID-19 cases and 111 the average number of SARS-CoV-2 genome copies per person per day measured in municipal 112 wastewater samples between January 2022 and March 2023. Because the relationship between 113 infections and wastewater concentration is only determined in the model up to an overall scaling 114 constant, it cannot be used to infer the absolute CAR but can be used to estimate relative 115 changes in case ascertainment over time. 116

From March 2020 until December 2021 New Zealand used strict border controls and intermittent 117 non-pharmaceutical interventions to suppress and eliminate transmission of SARS-CoV-2. By 118 the beginning of 2022, there had been a cumulative total of around 3 confirmed cases of COVID-119 19 per 1,000 people and around 90% of the population over 12 years old had received at least 120 two doses of the Pfizer-BioNTech vaccine. From October 2021, interventions were progressively 121 eased and in January 2022 the B.1.1.529 (Omicron) variant began to spread in the community, 122 causing the first large wave of infection. Since then community transmission has been sustained, 123 with multiple further waves of infection being driven by various Omicron subvariants. Between 124 1 January 2022 and 31 March 2023, there was a cumulative total of around 440 confirmed cases 125 per 1,000 people, most of which were from self-administered RATs. During this period, SARS-126 CoV-2 concentration was regularly measured at various wastewater treatment plants, providing 127 an additional data source on changes in community prevalence over time. 128

¹²⁹ 2 Materials and Methods

130 2.1 Data

National daily reported cases of COVID-19 were obtained from the New Zealand Ministry of Health [15]. Until February 2022, these cases were diagnosed solely by healthcare-administered PCR testing. From February 2022, in response to the rapid increase in reported cases, RATs were widely distributed. Since then, the vast majority of reported cases have been from selfadministered RATs, with results reported through an online portal. Reported cases are shown in Figure 1. As these data exhibit a clear day-of-the-week effect we remove the weekly trend before fitting the model (see Supplementary Material sec. 1.2 for details).

SARS-CoV-2 concentration data from wastewater samples collected by the Institute for Envi-138 ronmental Science and Research (ESR) were used for this study [16]. Wastewater samples were 139 collected every week at municipal wastewater treatment plants located throughout the coun-140 try, serving communities with populations ranging from 400 to over 500,000 people. Typically 141 70-90% of the national population connected to reticulated wastewater was covered by waste 142 water sampling in any given week (60-124 sites, usually sampled twice per week). We aggregate 143 the individual wastewater samples into daily estimates of genome copies per person by taking 144 volume-weighted averages of the samples on each day and dividing by the total population 145 connected to the sampled sites (see Figure 1). 146



Figure 1: Reported cases of COVID-19 (upper), daily SARS-CoV-2 genome copies per litre in sampled wastewater (middle), and proportion of the total population covered by sampled wastewater catchments (lower), between 1 January 2022 and 31 March 2023 in Aotearoa New Zealand. The black line in the upper plot shows the adjusted case series with the multiplicative day-of-the-week effect removed (see Supplementary Material section 1.2). The two outliers in wastewater data arise from estimates of a high wastewater flow-rate in Wellington following high rainfall. Since rainfall is a source of noise in wastewater sampling we retain these samples in our analysis. Reported case data were obtained from the New Zealand Ministry of Health [15] and wastewater data were obtained from ESR [16].

147 2.2 Hidden state model

We construct a state-space model (Figure 2) consisting of time-varying hidden states (the instantaneous reproduction number R_t , daily case ascertainment rate CAR_t , and daily infection incidence I_t) and time-varying observed states (daily reported cases of COVID-19 C_t and daily wastewater observations W_t). We use subscript s: t to refer to all values between day s and tinclusive.

¹⁵³ We assume the hidden states R_t and CAR_t follow independent Gaussian random walks, encoding ¹⁵⁴ the fact we expect them to vary continuously over time. We also assume that the hidden state ¹⁵⁵ I_t follows a Poisson renewal process, a simple epidemic model commonly used when estimating ¹⁵⁶ R_t [12]. Thus our state-space transitions are governed by:

$$(R_t | R_{t-1}) \sim N_{(0,\infty)}(R_{t-1}, \sigma_R R_{t-1})$$
$$(CAR_t | CAR_{t-1}) \sim N_{(0,1)}(CAR_{t-1}, \sigma_{CAR})$$
$$(I_t | R_t, I_{1:t-1}) \sim Poisson\left(R_t \sum_{u=1}^{t-1} g_u I_{t-u}\right)$$

Parameters σ_R and σ_{CAR} determine how quickly R_t and CAR_t vary. The standard deviation of the transition distribution for $R_t \to R_{t+1}$ is given by $\sigma_R R_t$, which means that R_t varies more rapidly at larger values. The distribution for R_t was truncated on $(0, \infty)$ and for CAR_t on (0, 1). Finally, g_u is the pre-determined generation time distribution, describing the proportion of transmission events that occur u days after infection (see Supplementary Material sec. 2.7).

We assume that the expected number of reported cases μ_t^c at time t is equal to CAR_t multiplied by the convolution of past infections with the infection-to-reporting distribution L_u :

$$\mu_t^c = CAR_t \sum_{u=1}^t I_{t-u} L_u$$

Similarly, we assume that the expected number of genome copies per person μ_t^w at time t is equal to the convolution of past infections with the infection-to-shedding distribution ω_u , multiplied by a fixed parameter α , representing the average total genome copies produced by an infectious individual:

$$\mu_t^w = \alpha \sum_{u=1}^t I_{t-u} \omega_u$$

¹⁶⁸ We model reported cases using a negative binomial distribution:

$$(C_t|CAR_t, I_{1:t}) \sim NegBin\left(r = k_c, p = \frac{k_c}{k_c + \mu_t^c}\right)$$

which has mean μ_t^c and variance $\mu_t^c \left(1 + \frac{\mu_t^c}{k_c}\right)$. A negative binomial distribution is used to account for noise in the observations beyond that predicted by a binomial distribution. This is a common choice in other methods of reproduction number estimation [14, 17].

¹⁷² We model observed wastewater data using a shape-scale gamma distribution:

$$(W_t|I_{1:t}) \sim \Gamma\left(k_w \operatorname{pop}_t, \frac{\mu_t^w}{k_w \operatorname{pop}_t}\right)$$

which has mean μ_t^w and variance $\frac{(\mu_t^w)^2}{k_w \text{pop}_t}$. The variable pop_t refers to the daily population in the catchment areas of the sampled wastewater sites at time t. Scaling by this allows the model to account for additional noise when fewer or smaller sites were sampled. On days when no sites were sampled, we let $P(W_t = 0) = 1$, meaning that the model filters on case data alone. The gamma distribution is a reasonably flexible choice for a non-negative continuous random variable, however other distributions could be considered, such as a Weibull or log-normal.



Figure 2: Diagram of the state-space model showing the dependency between hidden-states (dashed circles) and the observed data (solid circles). R_t is the instantaneous reproduction number on day t, CAR_t is the case ascertainment rate on day t, I_t is the number of new infections on day t, C_t is the number of reported cases on day t, and W_t is the observed wastewater, measured as genome copies per person per day, on day t. $I_{1:t}$ denotes the set of states $\{I_1, I_2, \ldots, I_t\}$. In practice the current infections I_t , reported cases C_t and wastewater W_t depend only on recent values of I_t as specified by the generation interval distribution, the infection-to-reporting distribution, and infection-to-shedding distribution respectively (see Methods).

In the absence of additional information we are unable to estimate α , which represents the average total genome copies shed by an infected individual over the course of their infection.

Table 1: Parameter values used in the model. The infection-to-reporting and infection-toshedding distributions are calculated as convolutions of the incubation period distribution [18] and the onset-to-reporting and onset-to-shedding distribution [19] respectively (see Supplementary Material sec. 2.7).

Parameter	Symbol	Value
Coefficient of variation of R_t transitions	σ_R	Fitted
Std dev. of CAR_t transitions	σ_{CAR}	Fitted
Reported cases tuning parameter	k_c	Fitted
Wastewater tuning parameter	k_w	Fitted
Generation time distribution [20, 21]	g_u	Mean = 3.3 days, s.d. = 1.3 days
Infection-to-reporting distribution	L_u	Mean = 5.8 days, s.d. = 2.6 days
Infection-to-shedding distribution	ω_u	Mean = 5.2 days, s.d. = 2.9 days
Average total genome copies per infection	α	$3 \times 10^9 \ (2 \times 10^9, \ 4 \times 10^9)$
Fixed-lag resampling window	h	30 days

This means we are unable to estimate the absolute value of CAR_t . Instead, we run the model 181

with a range of different values for α , and estimate the change in CAR_t relative to its initial 182 value. This additionally requires the assumption that α is constant over time, which is unlikely

183

to be true in general and is a key limitation of our model (see Discussion). 184

The infection-to-reporting and infection-to-shedding distributions are calculated as the convo-185 lution of the incubation period distribution with the onset-to-reporting and onset-to-shedding 186 distribution respectively. The incubation period is modelled as a Weibull distribution with mean 187 2.9 days and standard deviation 2.0 days [18]. The onset-to-reporting distribution is estimated 188 empirically from New Zealand case data extracted on 16 September 2022, representing over 1.2 189 million cases, and has mean 1.8 days and standard deviation 1.8 days. The onset-to-shedding 190 distribution comes from [19] and has mean 0.7 days and standard deviation 2.6 days. The 191 resulting infection-to-reporting distribution has mean 5.8 days and standard deviation 2.6, and 192 the resulting infection-to-shedding distribution has mean 5.2 days and standard deviation 2.9 193 days (see Supplementary Figure S1). 194

The model is solved using a bootstrap filter [22] with fixed-lag resampling. This produces 195 estimates for the marginal posterior distribution of the hidden states at each time step. The 196 random walk step variance parameters (σ_R and σ_{CAR}) and observation variance parameters 197 $(k_c \text{ and } k_w)$ are estimated using a particle marginal Metropolis Hastings Markov chain Monte 198 Carlo method. We use uninformative uniform prior distributions for these parameters, with 199 the exception of σ_{CAR} , where we use an informative prior distribution to ensure an appropriate 200 level of smoothness in our estimates of CAR_t . Different parameter values are fitted in three-201





Figure 3: Results for New Zealand data from 1 January 2022 to 31 March 2023. (a) instantaneous reproduction number R_t , (b) relative case ascertainment rate CAR_t (compared to the central estimate on 1 April 2022), (c) wastewater data W_t measured in genome copies per person per day and (d) reported cases C_t . Results assume the average total shedding per infection does not vary over time ($\alpha = 3 \times 10^9$). Solid lines present central estimates. Shaded regions show 95% credible intervals on the value of the hidden states (subplots a and b), and 95%credible intervals on the expected reported cases and wastewater data (darker shaded regions in subplots c and d) and 95% credible intervals on the prediction distribution for wastewater data and reported cases (lighter shaded regions in subplots c and d). Black dots show the observed data.

month blocks to allow for some variation over time. See Supplementary Material sec. 2 for 202 further details of the numerical method. Code and data to reproduce the results are provided 203 at https://github.com/nicsteyn2/NZWastewaterModelling. 204

205 **3** Results

²⁰⁶ Reproduction number, relative case ascertainment, and infection incidence

The estimated value of the reproduction number R_t (Figure 3a) increased from around 1 at the beginning of 2022 to a peak of 2.46 (95% CrI 2.04, 3.20) on 18 February 2022 (95% CrI 10 Feb, 23 Feb), corresponding to the sharp increase in cases seen during the first Omicron wave, which was a mixture of the BA.1 and BA.2 variants [23]. The estimated value of R_t dropped below 1 on 1 March 2022 (95% CrI 25 Feb, 5 Mar) and infection incidence peaked on 28 February 2022

²¹² (95% CrI 23 Feb, 7 Mar), suggesting this is when the wave peaked.

The estimated CAR (Figure 3b) increased rapidly between mid-February and mid-March 2022. 213 RATs became widely available for the first time in the last week of February 2022. This likely 214 led to a significant increase in case ascertainment as the testing system, which had previously 215 relied solely on laboratory-processed PCR tests, had become overwhelmed [3]. The estimated 216 CAR approximately halved between April and July 2022, when a second wave of infection 217 caused by the BA.5 Omicron subvariant [23, 24] occurred. This second wave was visible in both 218 reported cases and wastewater sampling, with estimated peak infections occurring on 7 July 219 2022 (95% CrI 3 Jul, 12 Jul). The estimated CAR increased somewhat between mid 2022 and 220 early 2023, with a noticeable dip in December 2022, possibly reflecting reduced testing during 221 the Christmas and summer school holiday period (from mid-December to late-January/early-222 February). Alternatively, the estimated increase in CAR from mid-2022 could be explained by 223 a decrease in the average genome copies shed by an infected individual α , although without 224 further information we are unable to discern changes in α . Overall, the model provided a 225 reasonably good fit to the observed data on cases and wastewater (Figure 3c-d). 226

Figure 4a-b shows the estimated daily incidence and cumulative infections for three values of 227 α , corresponding to estimated CAR values on 1 April 2022 of 0.42 (95% CrI 0.35, 0.50), 0.61 228 (95% CrI. 0.51, 0.71), and 0.80 (95% CrI. 0.67, 0.93), for $\alpha = 2 \times 10^9$, 3×10^9 , and 4×10^9 229 respectively. For comparison, the graphs also show the number of cases per capita in a cohort of 230 approximately 20,000 border workers who were tested weekly between January and July 2022 231 [24], scaled according to population size. This cohort is not representative of the population 232 and may not have perfect case ascertainment, so we do not expect the results to match exactly. 233 However, they provide a limited validation that the model is producing plausible estimates for 234 total infections. 235

Whilst peak reported cases (adjusted for the day-of-the-week effect) in the second wave were only 49% of the peak in the first wave (10,879 vs 22,038 respectively), under the assumption



Figure 4: Estimated (a) daily infections I_t , (b) cumulative infections $\sum_{s=0}^{t} I_s$, (c) case ascertainment rate CAR_t , (d) relative case ascertainment rate (compared to the central estimate on 1 April 2022), and (e) instantaneous reproduction number, R_t . Results are presented for three values of α : 2×10^9 , 3×10^9 , and 4×10^9 . Solid lines show central estimates and coloured regions are the 95% CrIs. Estimates and credible intervals on cumulative infections are calculated by taking cumulative sums of the estimates and credible intervals in panel (a). Black dots in panels (a) and (b) show the number of per capita cases in a cohort of regularly-tested border workers, scaled according to population size. The horizontal dashed black lie in panel (b) shows the New Zealand population at the end of 2022 (5.15 million people) [25]. While changing α results in different estimates of infections and absolute CAR, the relative CAR and reproduction number estimates are robust to different values, provided α remains relatively constant.

Table 2: Central estimates and 95% CrIs for estimated model parameters in each time period. Dates in the 'Period' column are the start date for the three-month period. All outputs presented to 2 s.f. Higher values of σ_R and σ_{CAR} suggest R_t and CAR_t vary faster. Higher values of k_c and k_w indicate a lower variance in the corresponding observation distribution. Note a different prior distribution was used for σ_{CAR} in the first period (see Supplementary Material, sec. 2.4), which may also impact estimates of other parameters in this period.

Period starting	σ_R	σ_{CAR}	k_c	$k_w(\times 10^{-6})$
1 Jan 2022	$0.12 \ (0.069, \ 0.21)$	$0.03\ (0.017,\ 0.043)$	31 (20, 49)	1.5(1.1, 2)
$1~{\rm Apr}~2022$	$0.069 \ (0.041, \ 0.12)$	$0.0099 \ (0.0053, \ 0.014)$	$170\ (100,\ 250)$	4.8 (3.2, 6.8)
1 Jul 2022	$0.037 \ (0.02, \ 0.066)$	$0.0063 \ (0.0018, \ 0.01)$	$330\ (220,\ 400)$	4.8 (3.3, 6.5)
$1 \ {\rm Oct} \ 2022$	$0.038\ (0.02,\ 0.068)$	$0.011 \ (0.0073, \ 0.014)$	$170\ (110,\ 270)$	7.2 (4.7, 10)
1 Jan 2023	$0.038 \ (0.018, \ 0.073)$	$0.0093 \ (0.0041, \ 0.015)$	150 (84, 330)	$6.8 \ (4.4, \ 10)$

of constant α , the central estimate from the model suggests that true infections peaked at approximately 78% of the peak of the initial wave (Figure 4a). Figure 4c-e shows the estimated absolute and relative CAR and R. These panels show that, while we are uncertain about the absolute level of infections and CAR, the relative CAR and reproduction number estimates are robust to reasonable choices for (constant) α .

243 Parameter estimates

The estimated standard deviation σ_R of the random walk on R_t was greatest in the first time 244 period (1 Jan - 31 Mar 2022) - see Table 2. This is unsurprising as it coincided with the rapid 245 increase and then decrease in incidence associated with the first Omicron wave. σ_R decreased 246 in the second period (1 Apr – 30 Jun 2022) and then remained relatively constant throughout 247 the remaining periods (1 Jul 2022 – 31 Mar 2023). The estimated standard deviation σ_{CAR} of 248 the random walk on CAR_t was also estimated to be greatest in the first time period, although 249 this is primarily because we applied a prior distribution with a higher mean in this period (see 250 Supplementary Material sec. 2.4). 251

The estimated variance parameters, k_c and k_w , for cases and wastewater observations, were lowest in the first time period (1 Jan 2022 – 31 Mar 2022). This implies there is more variability in the data that is not explained by the model in this time period, possibly as a consequence of the sharper variations in incidence compared to the later time periods. A less consistent weekly pattern in reported cases during the first time period, and higher levels of noise in wastewater observations at the low concentrations seen at the beginning of 2022, could also be contributing factors.

Discussion 4 259

WBE has been used globally for COVID-19 surveillance and has been shown to be a useful 260 public health tool for policy and public health responses [26]. We have presented a semi-261 mechanistic model that combines reported cases with wastewater data to estimate the time-262 varying reproduction number and CAR. This work demonstrates the value of WBE and how 263 the additional data that it provides can be combined with traditional monitoring (e.g., reported 264 cases) to learn more about the state of an epidemic, disease dynamics, and the true number 265 of infections in the community. This provides useful information to inform the public health 266 response. 267

To make reliable estimates of the state of the epidemic from reported cases, it is essential to 268 understand how case ascertainment changes with time. For example, are there fewer cases 269 because there are fewer infections or because fewer people are reporting? We applied our model 270 to national data from Aotearoa New Zealand and derived important insights into changes in 271 case ascertainment over the 15-month period considered. Reported cases during the second 272 wave in July 2022 were significantly lower than in the first wave in February and March 2022. 273 However, the model inferred that there was a substantial drop in case ascertainment between 274 these waves, and the true number of infections was likely more similar in each wave. The 275 reduced CAR during the second and subsequent waves may have been due to a higher number 276 of reinfections with individuals displaying fewer symptoms or due to "pandemic fatigue" and 277 reduced compliance with public health measures, including testing. This type of insight would 278 not be possible without regular wastewater surveillance data and without a robust analytical 279 framework in which to integrate these data with traditional epidemiological data streams. 280

Strengths of our model include the fact that it has relatively minimal data requirements, re-281 quiring only time series for reported cases and wastewater concentrations. This means that it 282 could be readily applied in other jurisdictions with wastewater surveillance programs, either 283 for SARS-CoV-2 or other pathogens such as influenza viruses [26, 27]. It is a relatively simple 284 model with minimal mechanistic assumptions and parsimonious parameterisation. This means 285 that results are less sensitive to model misspecification or parameter uncertainty than more 286 complex mechanistic models. The model presented here was operationalised by ESR in late 287 2022 and results for R_t and relative CAR are regularly provided to the Ministry of Health to 288 inform situational awareness and decision-making. 289

There are several limitations to this model and the results. We assume that the average number 290 of genome copies shed by an infected individual (the α parameter) was constant through 2022 291 and 2023 and did not depend on the infecting variant or history of prior infection or vaccination. 292

It is possible that some of the inferred changes in CAR may be partly explained by these 293 factors. For example, some of the inferred increase in case ascertainment between October and 294 December 2022 may have been due to decreasing α , caused by a combination of new immune 295 evasive subvariants displacing the previously dominant BA.5 variant [28] and/or an increase in 296 the proportion of reinfections or asymptomatic infections [15]. Furthermore, as we are unable to 297 estimate the true value of α , we are unable to estimate the absolute CAR. Nonetheless, relative 298 CAR is a useful metric and, given an estimated range of values for α , we are able to provide 299 plausible bounds on the total number of infections (Figure 4). 300

Wastewater surveillance does not provide any information on how infections are distributed among population groups (e.g. age groups, ethnicity) and biases in self-administered testing mean that case counts are not representative either. This information is important for assessing the clinical burden of disease and addressing health inequities [29]. Thus, other approaches are needed to determine the distribution of disease burden, such as representative sampling [6, 30], cohort studies [31] or sentinel surveillance [32].

As our model is flexible, future work could integrate hospitalisations (such as in [33]) and deaths data. In principle, this could allow the effects of varying CAR and varying rate of shedding per infection to be separated. However, this would additionally require the effects of age, immunity, ethnicity, and other variables on clinical severity to be accounted for.

The model could also be implemented at a regional level so that local epidemic dynamics can be compared. This paper has focused on modelling for inference: understanding epidemic dynamics that have already occurred. However, the state-space transition model coupled with the estimated parameters provides a natural method for forecasting [34, 14]. Forecasts generated using this state-space transition model naturally incorporate increasing uncertainty about the future reproduction number and CAR.

While this model has focused on COVID-19, there is a wealth of genetic information within 317 municipal wastewater that could also benefit from modelling. The detection and concentration 318 of viral, bacterial and anti-microbial resistance genes within wastewater have the ability to 319 inform public health decision-making in a number of ways, especially as methodology is refined 320 allowing more rapid turn-around times. As many jurisdictions seek to retain the wastewater 321 capabilities they built during the pandemic phase of COVID-19 (and to diversify microbial 322 targets), there is an 'opportunity springboard' to build tools that can predict the trajectories 323 and spread of pathogens - modelling has a key role to play in this journey. 324

325 Data availability

³²⁶ Daily reported case data for Aotearoa New Zealand are available from the Ministry of Health at

327 https://github.com/minhealthnz/nz-covid-data and seven-day average wastewater data

are available from ESR at https://github.com/ESR-NZ/covid_in_wastewater.

³²⁹ Code to run the model and reproduce the results in this paper are available at https://github.

330 com/nicsteyn2/NZWastewaterModelling.

331 Acknowledgements

The authors acknowledge the role of the New Zealand Ministry of Health in supplying data 332 in support of this work. The authors thank the many city and district council staff members 333 who collected the wastewater samples and the ESR laboratory staff who processed and tested 334 the samples used in this study. This work was funded by the New Zealand Ministry of Health 335 and the Department of Prime Minister and Cabinet (DPMC). This work was supported by 336 the NIHR HPRU in Emerging and Zoonotic Infections, a partnership between PHE, University 337 of Oxford, University of Liverpool, and Liverpool School of Tropical Medicine (grant number 338 NIHR200907 supporting C.A.D.). L. M. W. was supported by a Rutherford Foundation Post-339 doctoral Fellowship from New Zealand government funding, administered by the Royal Society 340 Te Apārangi. N.S. acknowledges support from the Oxford-Radcliffe Scholarship from University 341 College, Oxford, and the Engineering and Physical Sciences Research Council (EPSRC) Centre 342 for Doctoral Training (CDT) in Modern Statistics and Statistical Machine Learning (Imperial 343 College London and University of Oxford). We thank A. Maslov for supporting this research 344 through studentship support for N.S. 345

346 **References**

- [1] Ewan Colman, Gavrila A. Puspitarani, Jessica Enright, and Rowland R. Kao. Ascertain ment rate of SARS-CoV-2 infections from healthcare and community testing in the UK.
 Journal of Theoretical Biology, 558, 2 2023.
- [2] Oliver Eales, David Haw, Haowei Wang, Christina Atchison, Deborah Ashby, Graham S.
 Cooke, Wendy Barclay, Helen Ward, Ara Darzi, Christl A. Donnelly, Marc Chadeau-Hyam,
 Paul Elliott, and Steven Riley. Dynamics of SARS-CoV-2 infection hospitalisation and
 infection fatality ratios over 23 months in England. *PLOS Biology*, 21(5):e3002118, 5 2023.

[3] Giorgia Vattiatio, Audrey Lustig, Oliver J. Maclaren, and Michael J. Plank. Modelling the
 dynamics of infection, waning of immunity and re-infection with the Omicron variant of
 SARS-CoV-2 in Aotearoa New Zealand. *Epidemics*, 41:100657, 12 2022.

[4] Kris V. Parag, Christl A. Donnelly, and Alexander E. Zarebski. Quantifying the information
 in noisy epidemic curves. *Nature Computational Science*, 2(9):584–594, 9 2022.

[5] Fatimah S. Dawood, Christina A. Porucznik, Vic Veguilla, Joseph B. Stanford, Jazmin Duque, Melissa A. Rolfes, Ashton Dixon, Priyam Thind, Emily Hacker, Maria Julia E.
Castro, Zuha Jeddy, Michael Daugherty, Kim Altunkaynak, Danielle Rentz Hunt, Utsav Kattel, Jennifer Meece, and Melissa S. Stockwell. Incidence Rates, Household Infection Risk, and Clinical Characteristics of SARS-CoV-2 Infection Among Children and Adults in Utah and New York City, New York. JAMA Pediatrics, 176(1):59–67, 1 2022.

[6] Koen B. Pouwels, Thomas House, Emma Pritchard, Julie V. Robotham, Paul J. Birrell, 365 Andrew Gelman, Karina Doris Vihta, Nikola Bowers, Ian Boreham, Heledd Thomas, James 366 Lewis, Iain Bell, John I. Bell, John N. Newton, Jeremy Farrar, Ian Diamond, Pete Ben-367 ton, Ann Sarah Walker, Koen B. Pouwels, A. Sarah Walker, Derrick Crook, Philippa C. 368 Matthews, Tim Peto, Nicole Stoesser, Alison Howarth, George Doherty, James Kavanagh, 369 Kevin K. Chau, Stephanie B. Hatch, Daniel Ebner, Lucas Martins Ferreira, Thomas Chris-370 tott, Brian D. Marsden, Wanwisa Dejnirattisai, Juthathip Mongkolsapaya, Sarah Hoos-371 dally, Richard Cornall, David I. Stuart, Gavin Screaton, David Eyre, John Bell, Stuart 372 Cox, Kevin Paddon, Tim James, John N. Newton, Julie V. Robotham, Paul Birrell, Helena 373 Jordan, Tim Sheppard, Graham Athey, Dan Moody, Leigh Curry, Pamela Brereton, Jodie 374 Hay, Harper Vansteenhouse, Alex Lambert, Emma Rourke, Stacey Hawkes, Sarah Henry, 375 James Scruton, Peter Stokes, Tina Thomas, John Allen, Russell Black, Heather Bovill, 376 David Braunholtz, Dominic Brown, Sarah Collyer, Megan Crees, Colin Daglish, Byron 377 Davies, Hannah Donnarumma, Julia Douglas-Mann, Antonio Felton, Hannah Finselbach, 378 Eleanor Fordham, Alberta Ipser, Joe Jenkins, Joel Jones, Katherine Kent, Geeta Kerai, 379 Lina Lloyd, Victoria Masding, Ellie Osborn, Alpi Patel, Elizabeth Pereira, Tristan Pett, 380 Melissa Randall, Donna Reeve, Palvi Shah, Ruth Snook, Ruth Studley, Esther Sutherland, 381 Eliza Swinn, Anna Tudor, Joshua Weston, Shayla Leib, James Tierney, Gabor Farkas, Raf 382 Cobb, Folkert Van Galen, Lewis Compton, James Irving, John Clarke, Rachel Mullis, Lor-383 raine Ireland, Diana Airimitoaie, Charlotte Nash, Danielle Cox, Sarah Fisher, Zoe Moore, 384 James McLean, and Matt Kerby. Community prevalence of SARS-CoV-2 in England from 385 April to November, 2020: results from the ONS Coronavirus Infection Survey. The Lancet 386 Public Health, 6(1):e30–e38, 1 2021. 387

[7] Christian G. Daughton. Wastewater surveillance for population-wide Covid-19: The
 present and future, 9 2020.

[8] Harsh Dutta, Geetanjali Kaushik, and Venkatesh Dutta. Wastewater-based epidemiol ogy: a new frontier for tracking environmental persistence and community transmission of
 COVID-19. Environmental Science and Pollution Research, 29(57):85688-85699, 12 2022.

[9] Aparna Keshaviah, Megan B Diamond, Matthew J Wade, Samuel V Scarpino, Warish 393 Ahmed, Fabian Amman, Olusola Aruna, Andrei Badilla-Aguilar, Itay Bar-Or, Andreas 394 Bergthaler, Julie E Bines, Aaron W Bivins, Alexandria B Boehm, Jean-Martin Brault, 395 Jean-Baptiste Burnet, Joanne R Chapman, Angela Chaudhuri, Ana Maria de Roda Hus-396 man, Robert Delatolla, John J Dennehy, Megan Beth Diamond, Celeste Donato, Erwin 397 Duizer, Abiodun Egwuenu, Oran Erster, Despo Fatta-Kassinos, Aldo Gaggero, Deirdre F 398 Gilpin, Brent J Gilpin, Tyson E Graber, Christopher A Green, Amanda Handley, Joanne 399 Hewitt, Rochelle H Holm, Heribert Insam, Marc C Johnson, Rabia Johnson, Davey L 400 Jones, Timothy R Julian, Asha Jyothi, Aparna Keshaviah, Tamar Kohn, Katrin G Kuhn, 401 Giuseppina La Rosa, Marie Lesenfants, Douglas G Manuel, Patrick M D'Aoust, Rudolf 402 Markt, John W McGrath, Gertjan Medema, Christine L Moe, Indah Kartika Murni, Hu-403 mood Naser, Colleen C Naughton, Leslie Ogorzały, Vicka Oktaria, Christoph Ort, Popi 404 Karaolia, Ekta H Patel, Steve Paterson, Mahbubur Rahman, Pablo Rivera-Navarro, Alex 405 Robinson, Monica C Santa-Maria, Samuel V Scarpino, Heike Schmitt, Theodore Smith, 406 Lauren B Stadler, Jorgen Stassijns, Alberta Stenico, Renee A Street, Elisabetta Suffredini, 407 Zachary Susswein, Monica Trujillo, Matthew J Wade, Marlene K Wolfe, Habib Yakubu, 408 and Maria Ines Zanoli Sato. Wastewater monitoring can anchor global disease surveillance 409 systems. The Lancet Global Health, 11(6):e976–e981, 6 2023. 410

[10] Jana S. Huisman, Jérémie Scire, Lea Caduff, Xavier Fernandez-Cassi, Pravin Ganesanandamoorthy, Anina Kull, Andreas Scheidegger, Elyse Stachler, Alexandria B. Boehm, Bridgette Hughes, Alisha Knudson, Aaron Topol, Krista R. Wigginton, Marlene K. Wolfe,
Tamar Kohn, Christoph Ort, Tanja Stadler, and Timothy R. Julian. Wastewater-Based
Estimation of the Effective Reproductive Number of SARS-CoV-2. Environmental Health *Perspectives*, 130(5), 5 2022.

[11] Guangming Jiang, Jiangping Wu, Jennifer Weidhaas, Xuan Li, Yan Chen, Jochen Mueller,
Jiaying Li, Manish Kumar, Xu Zhou, Sudipti Arora, Eiji Haramoto, Samendra Sherchan,
Gorka Orive, Unax Lertxundi, Ryo Honda, Masaaki Kitajima, and Greg Jackson. Artificial
neural network-based estimation of COVID-19 case numbers and effective reproduction rate
using wastewater-based epidemiology. *Water Research*, 218, 6 2022.

[12] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework
and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology*, 178(9):1505–1512, 11 2013.

425 [13] R. N. Thompson, J. E. Stockwin, R. D. van Gaalen, J. A. Polonsky, Z. N. Kamvar, P. A.

Demarsh, E. Dahlqwist, S. Li, E. Miguel, T. Jombart, J. Lessler, S. Cauchemez, and
A. Cori. Improved inference of time-varying reproduction numbers during infectious disease
outbreaks. *Epidemics*, 29:100356, 12 2019.

[14] Sam Abbott, Joel Hellewell, Robin N. Thompson, Katharine Sherratt, Hamish P. Gibbs,
Nikos I. Bosse, James D. Munday, Sophie Meakin, Emma L. Doughty, June Young Chun,
Yung-Wai Desmond Chan, Flavio Finger, Paul Campbell, Akira Endo, Carl A. B. Pearson,
Amy Gimma, Tim Russell, Stefan Flasche, Adam J. Kucharski, Rosalind M. Eggo, and
Sebastian Funk. Estimating the time-varying reproduction number of SARS-CoV-2 using

⁴³⁴ national and subnational case counts. *Wellcome Open Research*, 5:112, 12 2020.

- ⁴³⁵ [15] Ministry of Health. COVID-19 data for New Zealand, 2023.
- [16] ESR. COVID-19 Data Repository by the Institute of Environmental Science and Research,
 2023.

[17] Nick Golding, David J. Price, Gerard E. Ryan, Jodie McVernon, James M. McCaw, and
Freya M. Shearer. A modelling approach to estimate the transmissibility of SARS-CoV 2
during periods of high, low, and zero case incidence. *eLife*, 12, 1 2023.

- [18] Jantien A. Backer, Dirk Eggink, Stijn P. Andeweg, Irene K. Veldhuijzen, Noortje van Maarseveen, Klaas Vermaas, Boris Vlaemynck, Raf Schepers, Susan van den Hof, Chantal B.E.M. Reusken, and Jacco Wallinga. Shorter serial intervals in SARS-CoV-2 cases with Omicron BA.1 variant compared with Delta variant, the Netherlands, 13 to 26 December 2021. Euro Surveillance, 27(6), 2 2022.
- Ioanne Hewitt, Sam Trowsdale, Bridget A. Armstrong, Joanne R. Chapman, Kirsten M.
 Carter, Dawn M. Croucher, Cassandra R. Trent, Rosemary E. Sim, and Brent J. Gilpin.
 Sensitivity of wastewater-based epidemiology for detection of SARS-CoV-2 RNA in a low
 prevalence setting. *Water Research*, 211:118032, 3 2022.
- [20] Sam Abbott, Katharine Sherratt, Moritz Gerstung, and Sebastian Funk. Estimation of
 the test to test distribution as a proxy for generation interval distribution for the Omicron
 variant in England. medRxiv, 1 2022.

[21] Dasom Kim, Sheikh Taslim Ali, Sungchan Kim, Jisoo Jo, Jun Sik Lim, Sunmi Lee, and
Sukhyun Ryu. Estimation of Serial Interval and Reproduction Number to Quantify the
Transmissibility of SARS-CoV-2 Omicron Variant in South Korea. Viruses 2022, Vol. 14,
Page 533, 14(3):533, 3 2022.

⁴⁵⁷ [22] N. J. Gordon, D. J. Salmond, and A. F.M. Smith. Novel approach to nonlinear/non⁴⁵⁸ gaussian Bayesian state estimation. *IEE Proceedings, Part F: Radar and Signal Processing*,
⁴⁵⁹ 140(2):107–113, 1993.

[23] Jordan Douglas, David Winter, Andrea McNeill, Sam Carr, Michael Bunce, Nigel French,
James Hadfield, Joep de Ligt, David Welch, and Jemma L. Geoghegan. Tracing the international arrivals of SARS-CoV-2 Omicron variants after Aotearoa New Zealand reopened
its border. *Nature Communications*, 13(1), 12 2022.

- ⁴⁶⁴ [24] Audrey Lustig, Giorgia Vattiato, Oliver Maclaren, Leighton M. Watson, Samik Datta, and
 ⁴⁶⁵ Michael J. Plank. Modelling the impact of the Omicron BA.5 subvariant in New Zealand.
 ⁴⁶⁶ Journal of the Royal Society Interface, 20(199), 2 2023.
- ⁴⁶⁷ [25] Stats NZ. National population estimates: at 31 December 2022, 2023.

⁴⁶⁸ [26] Pruthvi Kilaru, Dustin Hill, Kathryn Anderson, Mary B. Collins, Hyatt Green, Brittany L.
⁴⁶⁹ Kmush, and David A. Larsen. Wastewater Surveillance for Infectious Disease: A Systematic
⁴⁷⁰ Review. American journal of epidemiology, 192(2):305–322, 2 2023.

[27] Daniel Toribio-Avedillo, Clara Gómez-Gómez, Laura Sala-Comorera, Lorena RodríguezRubio, Albert Carcereny, David García-Pedemonte, Rosa Maria Pintó, Susana Guix, Belén
Galofré, Albert Bosch, Susana Merino, and Maite Muniesa. Monitoring influenza and respiratory syncytial virus in wastewater. Beyond COVID-19. Science of The Total Environment, page 164495, 5 2023.

[28] Sarah M. Prasek, Ian L. Pepper, Gabriel K. Innes, Stephanie Slinski, Walter Q. Betancourt, Aidan R. Foster, Hayley D. Yaglom, W. Tanner Porter, David M. Engelthaler, and
Bradley W. Schmitz. Variant-specific sars-cov-2 shedding rates in wastewater. Science of
the Total Environment, 857, 1 2023.

⁴⁸⁰ [29] Nicholas Steyn, Rachelle N Binny, Kate Hannah, Shaun C Hendy, Alex James, Audrey
⁴⁸¹ Lustig, Kannan Ridings, Michael J Plank, and Andrew Sporle. Māori and Pacific people
⁴⁸² in New Zealand have a higher risk of hospitalisation for COVID-19. 134:1538, 2021.

[30] Steven Riley, Kylie E.C. Ainslie, Oliver Eales, Caroline E. Walters, Haowei Wang, Christina
Atchison, Claudio Fronterre, Peter J. Diggle, Deborah Ashby, Christl A. Donnelly, Graham
Cooke, Wendy Barclay, Helen Ward, Ara Darzi, and Paul Elliott. Resurgence of SARSCoV-2: Detection by community viral surveillance. *Science*, 372(6545):990–995, 5 2021.

[31] Q. Sue Huang, Tim Wood, Lauren Jelley, Tineke Jennings, Sarah Jefferies, Karen Daniells,
Annette Nesdale, Tony Dowell, Nikki Turner, Priscilla Campbell-Stokes, Michelle Balm,
Hazel C. Dobinson, Cameron C. Grant, Shelley James, Nayyereh Aminisani, Jacqui Ralston, Wendy Gunn, Judy Bocacao, Jessica Danielewicz, Tessa Moncrieff, Andrea McNeill,
Liza Lopez, Ben Waite, Tomasz Kiedrzynski, Hannah Schrader, Rebekah Gray, Kayla
Cook, Danielle Currin, Chaune Engelbrecht, Whitney Tapurau, Leigh Emmerton, Maxine Martin, Michael G. Baker, Susan Taylor, Adrian Trenholme, Conroy Wong, Shirley

Lawrence, Colin McArthur, Alicia Stanley, Sally Roberts, Fahimeh Rahnama, Jenny Bennett, Chris Mansell, Meik Dilcher, Anja Werno, Jennifer Grant, Antje van der Linden, Ben
Youngblood, Paul G. Thomas, and Richard J. Webby. Impact of the COVID-19 nonpharmaceutical interventions on influenza and other respiratory viral infections in New Zealand. *Nature Communications*, 12(1), 12 2021.

- ⁴⁹⁹ [32] M. C. Zambon, J. D. Stockton, J. P. Clewley, and D. M. Fleming. Contribution of in⁵⁰⁰ fluenza and respiratory syncytial virus to community cases of influenza-like illness: An
 ⁵⁰¹ observational study. *Lancet*, 358(9291):1410–1416, 10 2001.
- [33] Hannes Schenk, Petra Heidinger, Heribert Insam, Norbert Kreuzinger, Rudolf Markt, Fabiana Nägele, Herbert Oberacher, Christoph Scheffknecht, Martin Steinlechner, Gunther
 Vogl, Andreas Otto Wagner, and Wolfgang Rauch. Prediction of hospitalisations based
 on wastewater-based SARS-CoV-2 epidemiology. *Science of the Total Environment*, 873, 5
 2023.
- [34] R. Moss, A. Zarebski, P. Dawson, and J. M. McCaw. Retrospective forecasting of the
 2010-2014 Melbourne influenza seasons using multiple surveillance systems. *Epidemiology* and Infection, 145(1):156–169, 1 2017.