Estimating COVID19 Border Arrival Risk in New Zealand

Richard Arnold[1], Rachelle N. Binny[2], Thomas Lumley[3], Audrey Lustig[2], Matthew Parry[4], Michael Plank[5]

1. School of Mathematics and Statistics, Victoria University of Wellington, New Zealand.
2. Manaaki Whenua, Lincoln, New Zealand.
3. Department of Statistics, University of Auckland, Auckland, New Zealand.
4. Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand.
5. School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand.

17 May 2022[1]

Executive Summary

The New Zealand Ministry of Health needs to assess the risk of COVID19 infection among international travellers arriving in New Zealand, principally by air.

The model described here assesses country level infection risk using daily infection rates among recent arrivals, and daily disease incidence in the country of origin.

1. Data used are:
    - **Arrivals**: Arrival date and source country of all passengers arriving at the New Zealand border (made available by Statistics New Zealand);
    - **Case Registrations**: Case status (confirmed, probable), vaccination status (unvaccinated, partly, fully) from the EpiSurv database, and matched to the arrivals data (provided by ESR);
    - **Source country incidence**: Daily new cases per head of population in all countries (From the Our World in Data website https://ourworldindata.org/)
2. The data are aggregated to weekly intervals.
3. A statistical model is implemented to estimate the future rate of infection among arriving passengers at the New Zealand border.
4. The model is calibrated using source country incidence, the observed numbers of arrivals, and cases in the previous 30 week period.

[1]This version, released on 15 September 2022, corrects a small number of typographical errors in the original version

5. Predictions are made using simple forecasts of the source country incidence, and the expected number of arrivals in New Zealand. Confidence intervals are calculated for the expected infection rates.

6. Each country is assigned to a 4-level risk category for each future week using the predicted rates and their uncertainties. The forecasts are reliable for the 3 weeks following the prediction.

7. The model incorporates a country-specific offset in risk which absorbs differences in rates among travellers compared to the general population in their source country, systematic differences between countries in the quality of their incidence reporting and vaccination coverage. In-flight/transit risk is not modelled explicitly, but is captured to some extent by the country level offset.

8. A further country specific and time dependent offset allows the model to adjust for surges in risk (due to sudden outbreaks, changes in policy or practice in a source country, and even new variants). Such changes can however take several weeks to be learned by the model.

9. The outputs of the model are available in a visualisation app. The outputs provide an initial assessment of the infection risk from every country suitable for setting border policies for each country. The assessments should not be used without reference to other information from source countries.

1 Introduction

The first confirmed case of COVID19 (SARS-Cov-2) in New Zealand was reported on 28 February 2020 [1].

One month later on 25 March 2020 a national state of emergency was declared and at 11:59pm that day New Zealand went into lockdown at Alert Level 4, the highest of a four stage system of restrictions designed to contain the spread of the virus. At that time there were 205 confirmed cases.

From then until October 2021 New Zealand pursued an elimination strategy, with occasional regional lockdowns (principally affecting the largest city, Auckland) to contain potential outbreaks. The measures were successful in containing spread, with New Zealand enjoying sustained periods at the lowest Alert level.

The progress of the pandemic is shown in Figure 1. The initial wave of cases was fully contained by the lockdown, with the country returning to Alert Level 1 at 11:59pm on 8 June 2020, with a total of 1504 cases (none active) and 22 deaths.

In the period June 2020-July 2021 there were no full national lockdowns, however regional outbreaks in Auckland raised the countrywide Alert Level to Level 2 with Auckland at Level 3 in August-September 2020, February-March 2021. The Wellington re-

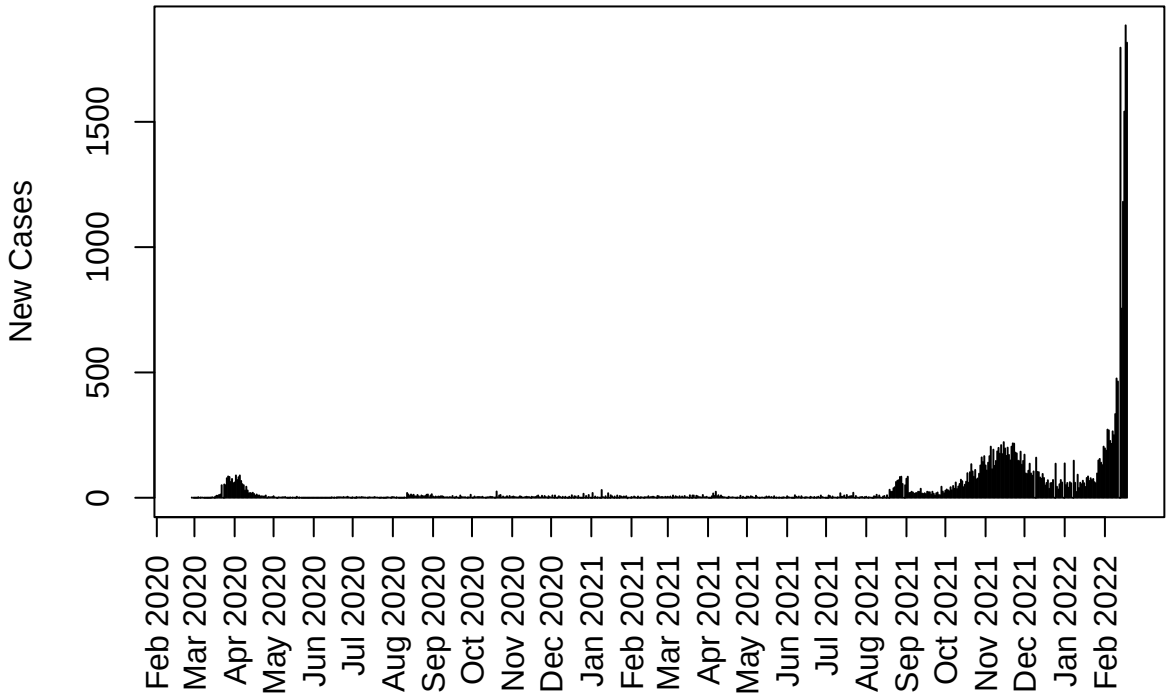


Figure 1: Daily COVID19 incidence in New Zealand (March 2022-February 2022)

gion went briefly into Alert Level 2 in June 2021 when following the diagnosis of a recent visitor from Australia with the Delta Variant.

The report of a new community case of the Delta Variant in Auckland sent the whole of New Zealand into Alert Level 4 lockdown on 17 August 2021. All regions except for Auckland moved to Alert Level 3 on 31 August, and then to Alert Level 2 on 7 September. Auckland followed with a move to Alert Level 3 on 21 September. After an extended lockdown failed to eliminate Delta cases in Auckland, and with increasing vaccination coverage, New Zealand moved to a new three tier traffic light system on 2 December 2021. The Omicron variant, which first appeared at the end of November 2021, has been found to be significantly more transmissible that other variants. As a result Omicron has dramatically changed patterns of incidence across the world, and affected border policies [2].

Restrictions at the border have been a part of New Zealand's pandemic control measures since the initial border closure on 19 March 2020. A timeline of key border policy changes is given in Table 1.

Throughout the pandemic the New Zealand Government has sought a robust means of assessing the risk of importing cases of COVID19.

The levels of risk depend on a number of factors including the incidence rate in the source country, levels of vaccination, disease control measures, and disease control along the route taken from the source country to New Zealand.

This paper describes a statistical model which provides an up to date assessment of the immediate risk among arriving passengers from other countries. Information on infections among historical arrivals to New Zealand as well as current information on incidence is used to create estimates of arrival risk for each country.

The proposed model accounts for possible variations in disease reporting between countries. Where a country is underreporting its disease incidence, the historical rate of disease among arrivals will appear to be too high. In that case future arrivals from that country will correspondingly be assessed as being at higher risk of infection than the in-country incidence might suggest. Conversely, in countries where the incidence of disease is low and mostly confined to managed isolation facilities, then arrivals from that country will have low rates of disease, and arrivals will be assessed as being at lower risk than the in-country incidence would suggest.

The model contains a country-specific time varying component which allows these higher or lower risks to change over time, as the pandemic progresses and as different disease control measures are implemented.

Section 2 describes the data sources used in the paper, and the model is explained in detail in Section 3. The practical implementation of the model is described in Section 4, and some example results are presented in Section 5. A short discussion concludes the paper in Section 6. Details of the model selection procedure are provided in Appendix A.

Table 1: New Zealand border restrictions during the COVID19 pandemic

Date	Measures taken	Reference
2 Feb 2020	Travellers leaving from or transitting through China will be refused entry; Foreign travellers subject to increased screening on arrival; NZ citizens and residents must self-isolate for 14 days	[3]
19 Mar 2020	NZ borders closed except to NZ citizens and permanent residents, their partners and children and some health and humanitarian workers	[4]
9 Apr 2020	Managed Isolation and Quarantine become mandatory for all arrivals	[5]
11 Aug 2020	Charges for some MIQ users are introduced	[5]
3 Nov 2020	Advance MIQ bookings become mandatory for all arrivals	[6]
15 Jan 2021	Travellers from the UK and US need negative COVID19 test results within 72 hours of departure	[7]
18 Jan 2021	Negative pre-departure tests required for all arrivals excepting those from Australia, Antarctica and some Pacific Island nations; Tests in MIQ at Day 0, 3 and 12;	[5]
21 Jan 2021	Quarantine-free travel from the Cook Islands starts	[8]
1 Apr 2021	Widening of criteria for emergency bookings in MIQ	[9]
11-18 Apr 2021	Arrivals from India banned for 2 weeks	[10]
19 Apr-23 Jul 2021	Quarantine-free travel from Australia	[5, 11, 12]
28 Apr 2021	India, Brazil, Papua New Guinea and Pakistan classified as Very High Risk: travel restricted to NZ Citizens and families	[13, 14]
17 May 2021	Two-way quarantine-free travel between NZ and the Cook Islands begins	[15]
15 Aug 2021	Fiji and Indonesia classified as Very High Risk: travel restricted to NZ Citizens and families	[16, 14]
1 Nov 2021	Full vaccination a requirement for entry for non-citizens	[17]
2 Dec 2021	Three tier traffic light system introduced	[18]
27 Feb 2022	No MIQ for vaccinated NZ residents/citizens & critical workers coming from Australia	[19]
4 Mar 2022	No MIQ for vaccinated NZ residents/citizens & critical workers coming from all origins	[19]
13 Mar 2022	No MIQ for critical workers coming from all origins	[19]
12 Apr 2022	No MIQ for vaccinated current visa holders	[19]
Jul 2022	No MIQ for visa holders and visitors from visa waiver countries	[19]
Oct 2022	No MIQ for vaccinated arrivals under normal visa processing arrangements	[19]

2 Data

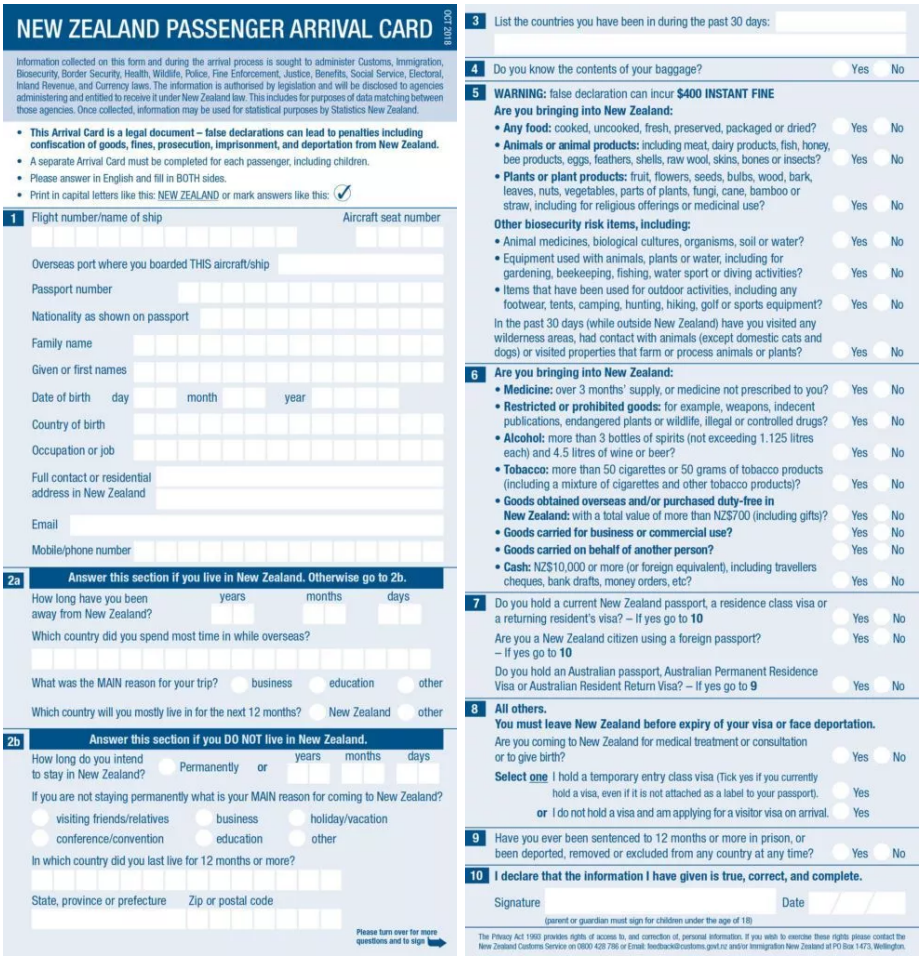
The model uses the following routine data sources.

- Our World in Data COVID19 data set [20]. The website provides daily data on incidence, effective reproduction number R_{eff}, new deaths, partial and full vaccination coverage, new tests, and the test positivity rate;
- Epiforecasts [21]. Effective reproduction number estimates R_{eff} from a research group and the London School of Hygiene and Tropical Medicine.
- Daily counts of new arrivals by country of origin (Statistics New Zealand);
- Daily counts of cases among arrivals by country (ESR, EpiSurv data set);

2.1 Data Quality

- We assume that the data on new arrivals and counts of cases are complete in New Zealand. New Zealand's geographic isolation means it has no land borders, and that unreported arrivals by sea are rare. Arrivals by air are controlled at New Zealand's international airports.
- The information on the country of origin of each arrival is collected in Question 2 of the passenger arrival card completed by all passengers and crew on aeroplanes and ships arriving in New Zealand.
  There are two versions of the question (see Figure 2):
  1. **Question 2a:** 'Answer this question if you live in New Zealand: ... Which country did you spend the most time in while overseas?'
  2. **Question 2b:** 'Answer this question if you DO NOT live in New Zealand: ... In which country did you last live for 12 months or more?'

  There is also a free text response question **Question 3:** 'List the countries you have been in during the last 30 days:' This response is not routinely captured.
  The country of origin that is captured in Question 2 may not be the country at which the person was most at risk of infection.
- Arrivals that could not be matched to a standard country name, or where the country did not have corresponding data in the Our World in Data dataset were assigned to 'Unknown Origin'.
  Between 8 Jun 2020 and 20 Feb 2022 there were 4060 arrivals with unknown origin, out of a total of 207518 arrivals (2.0%).
- Cases are selected from the EpiSurv database if their Status flag has the value 'Confirmed' or 'Probable', and the Overseas flag is 'Yes'.
- When an arriving person is found to be infected with COVID19 more information is collected: the three most recently visited countries are captured in the EpiSurv database, along with the departure dates from each country.
  Of these three countries the least recently visited country is selected from the countries that were visited within the previous 14 days.



NEW ZEALAND PASSENGER ARRIVAL CARD

Information collected on this form and during the arrival process is sought to administer Customs, Immigration, Biosecurity, Border Security, Health, Wildlife, Police, Fine Enforcement, Justice, Benefits, Social Service, Electoral, Inland Revenue, and Currency laws. The information is authorised by legislation and will be disclosed to agencies administering and entitled to receive it under New Zealand law. This includes for purposes of data matching between those agencies. Once collected, information may be used for statistical purposes by Statistics New Zealand.

- **This Arrival Card is a legal document – false declarations can lead to penalties including confiscation of goods, fines, prosecution, imprisonment, and deportation from New Zealand.**
- A separate Arrival Card must be completed for each passenger, including children.
- Please answer in English and fill in BOTH sides.
- Print in capital letters like this: NEW ZEALAND or mark answers like this: ✓

1 Flight number/name of ship — Aircraft seat number

Overseas port where you boarded THIS aircraft/ship

Passport number

Nationality as shown on passport

Family name

Given or first names

Date of birth day month year

Country of birth

Occupation or job

Full contact or residential address in New Zealand

Email

Mobile/phone number

2a Answer this section if you live in New Zealand. Otherwise go to 2b.

How long have you been away from New Zealand? years months days

Which country did you spend most time in while overseas?

What was the MAIN reason for your trip? business education other

Which country will you mostly live in for the next 12 months? New Zealand other

2b Answer this section if you DO NOT live in New Zealand.

How long do you intend to stay in New Zealand? Permanently or years months days

If you are not staying permanently what is your MAIN reason for coming to New Zealand?

visiting friends/relatives business holiday/vacation

conference/convention education other

In which country did you last live for 12 months or more?

State, province or prefecture Zip or postal code

Please turn over for more questions and to sign

3 List the countries you have been in during the past 30 days:

4 Do you know the contents of your baggage? Yes No

5 **WARNING:** false declaration can incur **$400 INSTANT FINE**

Are you bringing into New Zealand:

- **Any food:** cooked, uncooked, fresh, preserved, packaged or dried? Yes No
- **Animals or animal products:** including meat, dairy products, fish, honey, bee products, eggs, feathers, shells, raw wool, skins, bones or insects? Yes No
- **Plants or plant products:** fruit, flowers, seeds, bulbs, wood, bark, leaves, nuts, vegetables, parts of plants, fungi, cane, bamboo or straw, including for religious offerings or medicinal use? Yes No

Other biosecurity risk items, including:

- Animal medicines, biological cultures, organisms, soil or water? Yes No
- Equipment used with animals, plants or water, including for gardening, beekeeping, fishing, water sport or diving activities? Yes No
- Items that have been used for outdoor activities, including any footwear, tents, camping, hunting, hiking, golf or sports equipment? Yes No

In the past 30 days (while outside New Zealand) have you visited any wilderness areas, had contact with animals (except domestic cats and dogs) or visited properties that farm or process animals or plants? Yes No

6 **Are you bringing into New Zealand:**

- **Medicine:** over 3 months' supply, or medicine not prescribed to you? Yes No
- **Restricted or prohibited goods:** for example, weapons, indecent publications, endangered plants or wildlife, illegal or controlled drugs? Yes No
- **Alcohol:** more than 3 bottles of spirits (not exceeding 1.125 litres each) and 4.5 litres of wine or beer? Yes No
- **Tobacco:** more than 50 cigarettes or 50 grams of tobacco products (including a mixture of cigarettes and other tobacco products)? Yes No
- **Goods obtained overseas and/or purchased duty-free in New Zealand:** with a total value of more than NZ$700 (including gifts)? Yes No
- **Goods carried for business or commercial use?** Yes No
- **Goods carried on behalf of another person?** Yes No
- **Cash:** NZ$10,000 or more (or foreign equivalent), including travellers cheques, bank drafts, money orders, etc? Yes No

7 Do you hold a current New Zealand passport, a residence class visa or a returning resident's visa? – If yes go to **10** Yes No

Are you a New Zealand citizen using a foreign passport? – If yes go to **10** Yes No

Do you hold an Australian passport, Australian Permanent Residence Visa or Australian Resident Return Visa? – If yes go to **9** Yes No

8 **All others.**
You must leave New Zealand before expiry of your visa or face deportation.

Are you coming to New Zealand for medical treatment or consultation or to give birth? Yes No

Select one I hold a temporary entry class visa (Tick yes if you currently hold a visa, even if it is not attached as a label to your passport). Yes

or I do not hold a visa and am applying for a visitor visa on arrival. Yes

9 Have you ever been sentenced to 12 months or more in prison, or been deported, removed or excluded from any country at any time? Yes No

10 **I declare that the information I have given is true, correct, and complete.**

Signature Date

(parent or guardian must sign for children under the age of 18)

The Privacy Act 1993 provides rights of access to, and correction of, personal information. If you wish to exercise these rights please contact the New Zealand Customs Service on 0800 428 786 or Email: feedback@customs.govt.nz and/or Immigration New Zealand at PO Box 1473, Wellington.

Figure 2: New Zealand Passenger Arrival Card

Inevitably there are inaccurate dates (e.g. departures taking place before arrivals) and missing data on recently visited countries. Between 8 Jun 2020 and 20 Feb 2022 there were 2 cases with unknown origin, out of a total of 3047 cases arriving at the border (0.1%).

- There is a potential mismatch between the reported origin of cases (from EpiSurv) and the counts of arrivals by country (from the arrival card). For example, there were 10 cases recorded as being from Ukraine reported on 26 January 2021, but only 2 people classified as having arrived from the Ukraine that day.
  Prior to 1 April 2021 the arrival card country is used in the denominator, and the EpiSurv country is used in the numerator.
- All analyses from 1 April 2021 onwards use the Arrival Card country if available. If that is not available then the last known port along the passenger's route is used. If that is not available **and** if the arrival is a case, then the country of origin reported in the EpiSurv database is used (or the 'Travelled from' country if that is not available). This consistency of approach minimises any numerator/denominator bias, although it likely does introduce some misclassification, since the EpiSurv country is more likely to be correct.
  During the period 1 April 2021-21 November 2021 there were 60779 arrivals and 486 cases.
  Among the 60293 non-cases, there were 695 (1.2%) with no Arrival Card country, similar to the rate (7/486=1.4%) among the cases.
  Among the 486 cases:
  - 3 (0.6%) have neither an Arrival Card nor an EpiSurv country;
  - 4 (0.8%) have an EpiSurv country only;
  - 39 (2.5%) have an Arrival Card country only;
  - 68 (14%) have both countries, but they don't match
  - 372 (77%) have both countries and they do match

  The 68 cases with non-matching EpiSurv and Arrival Card countries record 26 different EpiSurv countries and 30 different Arrival Card countries. These are all mostly small numbers (1-3 cases per country), but the largest EpiSurv counts are from Russia (7), the USA (7) and the UK (15). These 29 cases are distributed among 10 different Arrival Card countries, with small numbers from each with the notable exception of 11 cases with the UK as their EpiSurv country, but Singapore as their Arrival Card country.
  We can expect over-attribution of cases to the United Arab Emirates, the United States, Qatar and Singapore, which account for the most frequent routes by which travellers come to New Zealand. The consequent under-attribution of cases to the genuine source countries is a small effect for most source countries, given that we have seen very small counts (1-3) missing from a large number of countries (26).
- In situations where the number of cases from a country exceeds the number of arrivals from that country on a particular day, we exclude **ALL** of those cases from the model.
- Inflight infection is not accounted for in the model. However, with the highly transmissible Omicron variant this is more likely to be an issue than with other variants.

Where inflight transmission is known to have occurred, then in future implementations of the model these cases could either be (a) deleted from the case counts (in order not to underestimate this risk of border cases), or (b) reclassified, either at random or proportionately, across a set of countries.

- Data are aggregated to weekly totals (Monday-Sunday) before modelling. This reduces the proportion of counts that are very small, as well as eliminating any effect of irregular reporting during the various days of the week.
  Weeks are numbered in some output: the Monday of Week 1 is 13 January 2020.
  If the final week in the available data has 5 or 6 days we scale the number of cases and arrivals so that the counts are equivalent to a weekly total. If the final week has fewer than 5 days of data all the data for that week are excluded.
  We thus have one record of data for each country for each week.
- The Our World in Data dataset contains occasional negative counts of cases, deaths or tests. These are corrections to earlier overcounting. These negative counts are all set to zero, with no other adjustments made.

3 Methods

For country c in time period t we use the following notation:

- I_{ct} = in-country reported incidence (cases per head of population)
- $\mathbf{x}_{ct}$ = vector of other characteristics (proportion of the population tested, proportion of tests that are positive, proportion of the population at least partially vaccinated, proportion fully vaccinated, effective reproduction number);
- N_{ct} = total number of passengers arriving at the New Zealand border
- Y_{ct} = number of cases arriving at the New Zealand border

All data are aggregated to weekly periods.

We then model the number of arriving cases Y_{ct} among N_{ct} arrivals from country c in time period t using a Binomial mixed effects model with lagged in country incidence as the only predictors. Appendix A gives details of the model selection procedure. Briefly, we found that the available predictors (death rates, testing rates, test positivity rates and vaccination rates) were either too incomplete or too weakly informative to be useful as predictors. In particular, testing rates and vaccination rates were too strongly varying over time, due to changing in-country policies, that they could not be used with confidence in a predictive model. Ultimately the vector of predictors $\mathbf{x}_{ct}$ was reduced to the set of lagged in-country incidence rates I_{ct}.

The model is a Binomial count model,

$$Y_{ct} \sim \text{Binomial}(N_{ct}, \mu_{ct}) \tag{1}$$

and the mean μ_{ct} is modelled by

$$\begin{aligned} \eta_{ct} = \text{logit}\mu_{ct} &= \alpha + \sum_{k=k_1}^{k_2} \beta_k \text{logit}\left(\delta + I_{c,t-k}\right) + u_c + v_{ct} \\ u_c &\overset{\text{iid}}{\sim} N(0, \sigma_u^2) \\ v_{ct} &= \rho v_{c,t-1} + \varepsilon_{ct} \\ \varepsilon_{ct} &\overset{\text{iid}}{\sim} N(0, \sigma_e^2) \end{aligned} \tag{2}$$

We use logits to transform proportions and logs to transform non-negative quantities onto the unrestricted linear predictor scale. Since the incidence rates are in general low ($< 1\%$), this specification encodes the approximate proportionality:

$$\mu_{ct} \propto \prod_{k=k_1}^{k_2} \left(I_{c,t-k}\right)^{\beta_k} \tag{3}$$

Note also that:

- The logistic (logit) transformation is defined $\eta = \text{logit}(p) = \log \frac{p}{1-p}$, with inverse transformation $p = \text{expit}(\eta) = \frac{e^\eta}{1+e^\eta}$.
- $\delta = 10^{-7}$ is a small offset added to incidence observations which allows for situations where the reported incidence rate I_{ct} is 0.
- Lagged incidence is used in the predictors, with minimum lag $k_1 = 0$ and maximum lag $k_2 = 2$.

The random effects structure assigns a country level random effect u_c to country c to account for arrivals from that country differing in risk from the risk level suggested by the reported incidence rate I_{ct}.

The autoregressive AR(1) error structure v_{ct} allows this country level effect to change over time, but with temporal correlation.

The full set of parameters of the model is thus

- α: intercept
- $\beta_{k_1}, \ldots, \beta_{k_2}$: parameters associated with the logit of lagged incidence rates $I_{t-k_1}, \ldots, I_{t-k_2}$
- ρ: AR(1) correlation parameter
- σ_u^2: variance of country level random effects
- σ_e^2: one step variance of time varying AR(1) random effects

The vector $\boldsymbol{\beta} = (\alpha, \beta_{k_1}, \ldots, \beta_{k_2})^T$ contains the $p = 2 + k_2 - k_1$ parameters of the linear predictor in (2), and X is the associated design matrix (which has p columns, and one row $\mathbf{x}_{ct}$ for each country$\times$week entry in the dataset).

3.1 Estimation

The model is fit using the R package `glmmTMB` [22] which implements in R the TMB package [23] for use with generalised linear mixed effects models. We restrict the

data to be modelled to data in a fixed window, the most recent 30 weeks, to allow for long term changes in the pandemic.

Given data on cases, arrivals, in-country incidence rates (Y_{ct}, N_{ct}, I_{ct}) at the n weekly time points $(t_1, \ldots, t_n)$ and countries $c = 1, \ldots, C$ the TMB estimation function returns:

- Parameter estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{\rho}$, $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_e^2$;
- Variance-covariance of $\widehat{\boldsymbol{\beta}}$ estimates: $\widehat{V}_\beta$;
- Estimated country level random effects $\widehat{u}_c$;
- Estimated time dependent random effects $\widehat{v}_{ct}$ for each country at the observed data time points $(t_1, \ldots, t_n)$;
- Fitted values and variances on the logit scale

$$\widehat{\eta}_{ct} = \mathbf{x}_{ct}^T \widehat{\boldsymbol{\beta}} + \widehat{u}_c + \widehat{v}_{ct} \tag{4}$$

$$\text{Var}[\widehat{\eta}_{ct}] = \widehat{S}_{ct}^2 \tag{5}$$

  i.e. $\widehat{S}_{ct}$ is the standard error of the estimate of η_{ct};
- Fitted values and variance on the observation scale

$$\widehat{Y}_{ct} = N_{ct}\widehat{\mu}_{ct} = N_{ct}\text{expit}(\mathbf{x}_{ct}^T \widehat{\boldsymbol{\beta}} + \widehat{u}_c + \widehat{v}_{ct}) \tag{6}$$

$$\text{Var}[\widehat{Y}_{ct}] = N_{ct}^2 \widehat{S}_{ct}^2 \, \widehat{\mu}_{ct}^2 (1 - \widehat{\mu}_{ct})^2 \tag{7}$$

Confidence intervals for η_{ct}, μ_{ct} and $\widehat{Y}_{ct}$ can be constructed as follows:

$$\begin{aligned} (\widehat{\eta}_{ct}^-, \widehat{\eta}_{ct}^+) &= \widehat{\eta}_{ct} \pm Z_\alpha \widehat{S}_{ct} \\ (\widehat{\mu}_{ct}^-, \widehat{\mu}_{ct}^+) &= \text{expit}(\widehat{\eta}_{ct} \pm Z_\alpha \widehat{S}_{ct}) \end{aligned} \tag{8}$$

$$(\widehat{Y}_{ct}^-, \widehat{Y}_{ct}^+) = N_{ct}\text{expit}(\widehat{\eta}_{ct} \pm Z_\alpha \widehat{S}_{ct}) \tag{9}$$

where Z_α is the appropriate quantile of a standard Normal distribution for the desired level of confidence $1 - \alpha$.

In-sample prediction intervals can be constructed as:

$$(Y_{ct}^-, Y_{ct}^+) = N_{ct}\text{expit}\left(\widehat{\eta}_{ct} \pm Z_\alpha \sqrt{\frac{1}{N_{ct}\widehat{\mu}_{ct}(1 - \widehat{\mu}_{ct})} + \widehat{S}_{ct}^2}\right) . \tag{10}$$

3.2 Aggregating over fitted values

We need to aggregate

- over time when computing observed and expected values over periods of multiple weeks,
- over countries when computing the expected total numbers of cases arriving at the border from all countries combined.

Given a set of independent Binomial random variables

$$Y_k|N_k, \mu_k \sim \text{Binomial}(N_k, \mu_k) \tag{11}$$

where we have estimated $\eta_k = \text{logit}(\mu_k)$ with fitted values $\widehat{\eta}_k = \text{logit}(\widehat{\mu}_k)$ and variances $\text{Var}[\widehat{\eta}_k] = \widehat{S}_k^2$, we can write

$$\sigma_k^2 = N_k\mu_k(1-\mu_k) \qquad \text{and} \qquad \widehat{\sigma}_k^2 = N_k\widehat{\mu}_k(1-\widehat{\mu}_k)$$

Aggregation then proceeds as follows:

$$\begin{aligned}
Y &= \sum_k Y_k \\
N &= \sum_k N_k \\
E[Y|\{\widehat{\mu}_k\}] &= \sum_k N_k\widehat{\mu}_k \\
\text{Var}[Y|\{\widehat{\mu}_k\}] &= \sum_k N_k\widehat{\mu}_k(1-\widehat{\mu}_k) \\
E[\text{Var}[Y|\{\widehat{\mu}_k\}]] &= \sum_k \widehat{\sigma}_k^2 - \sum_k \frac{1}{N_k}\widehat{\sigma}_k^4\widehat{S}_k^2 \\
\text{Var}[E[Y|\{\widehat{\mu}_k\}]] &= \sum_k \widehat{\sigma}_k^4\widehat{S}_k^2 \\
\text{Var}[Y] &\simeq \sum_k \widehat{\sigma}_k^4\widehat{S}_k^2 + \sum_k \widehat{\sigma}_k^2
\end{aligned} \tag{12}$$

So that if we set $\widehat{\mu} = Y/N$ then the variances needed for confidence and prediction intervals for $\text{logit}(Y/N)$ are, respectively:

$$\text{Var}\left[E\left[\text{logit}\,\frac{Y}{N}\middle|\{\widehat{\mu}_k\}\right]\right] = \frac{\sum_k \widehat{\sigma}_k^4\widehat{S}_k^2}{N^2\widehat{\mu}^2(1-\widehat{\mu})^2} \tag{13}$$

$$\text{Var}\left[\text{logit}\,\frac{Y}{N}\middle|\{\widehat{\mu}_k\}\right] = \frac{\sum_k \left(\widehat{\sigma}_k^4\widehat{S}_k^2 + \widehat{\sigma}_k^2\right)}{N^2\widehat{\mu}^2(1-\widehat{\mu})^2}\,. \tag{14}$$

3.3 Forecasts

To forecast the model a further k time steps beyond the last observation t_n, we require first a method for forecasting the number of arrivals N_{ct} and the in country incidence rate I_{ct}.

- We forecast **arrivals** N_{ct} using known Managed Isolation and Quarantine (MIQ) bookings, or if these are not available, we set $N_{ct} = \bar{N}_c$ for all $t > t_n$ where $\bar{N}_c$ is the mean number of arrivals for country c in a fixed window prior to the last observation at t_n.

- We forecast **in-country incidence** using a weighted linear fit to logit$(\delta + I_{ct})$ using the last $m = 3$ observations with weights $1/m, 2/m, \ldots, m/m$. δ is again a small offset (10^{-7}) added to all incidence observations to avoid zeros.

We then need to forecast the other components of the model:

- The fixed effect parameters $\widehat{\boldsymbol{\beta}}$ and covariances $\widehat{V}_\beta$ are time invariant;
- The country level random effect errors $\widehat{\sigma}_u$ are time invariant;
- The AR(1) correlation and error $\widehat{\rho}$, $\widehat{\sigma}_e$ are time invariant;
- The country level random effects $\widehat{u}_c$ are time invariant;
- The country-period level random effects, predicting k steps forward:

$$\widehat{v}_{ct_{n+k}} = \widehat{\rho}^k v_{ct_n} + \zeta_{ct_{n+k}}$$

where

$$\zeta_{ct_{n+k}} \overset{\text{iid}}{\sim} N\left(0, \widehat{\sigma}_e^2 \frac{1-\widehat{\rho}^{2k}}{1-\widehat{\rho}^2}\right)$$

Forecasts use the expected value $\widehat{v}_{ct_{n+k}} = \widehat{\rho}^k v_{ct_n}$.

Conveniently, forecasts of the infection rates of arrivals on the logit scale ($\widehat{\eta}_{ct}$) and their associated standard errors ($\widehat{S}_{ct}$) can be computed when the model is being fitted at the estimation stage. In this procedure we set both the observed numbers of arrivals (N_{ct}) and numbers of observed cases (Y_{ct}) to zero for the time periods to be forecasted, and include them in the data set during the fitting process. The likelihood contributions from these observations are all independent of the parameters, and do not affect the estimation process. These time periods are thus treated in the same way as time periods that were truly observed within the data set, but at which no arrivals occurred.

When estimating confidence and prediction intervals for the future observations, we replace the zero values of the arrivals N_{ct} with the forecast arrival numbers, and then use the methods of Section 3.1.

Note that these confidence and prediction intervals both neglect any uncertainty introduced in the forecasting of the numbers of arrivals and of the in-country incidence. The number of arrivals may be estimated from Managed Isolation and Quarantine (MIQ) bookings, and for instances where passengers are not required to use MIQ estimates can be made from flight schedules.

3.4 One step ahead forecasts

One step ahead forecasts are useful in-sample measures of goodness of fit, and were used in the model selection procedure described in Appendix A.

On the logit scale the one step ahead forecast is

$$\widehat{\eta}_{ct|t-1} = \mathbf{x}_{ct}^T \widehat{\boldsymbol{\beta}} + \widehat{u}_c + \widehat{\rho}\widehat{v}_{c,t-1} \tag{15}$$

with variance

$$\text{Var}[\widehat{\eta}_{ct|t-1}] = \text{Var}[\mathbf{x}_{ct}^T \widehat{\boldsymbol{\beta}} + \widehat{u}_c + \widehat{\rho}\widehat{v}_{c,t-1}] \tag{16}$$

which we estimate by its approximate upper bound

$$\text{Var}[\widehat{\eta}_{ct|t-1}] \leq \widehat{S}^2_{c,t-1} + \max\left(0, \mathbf{x}_{ct}^T \widehat{V}_\beta \mathbf{x}_{ct} - \mathbf{x}_{c,t-1}^T \widehat{V}_\beta \mathbf{x}_{c,t-1}\right) . \quad (17)$$

3.5 Risk Classification

In each week each country can be classified by a multi-level risk categorisation using the fitted and forecasted estimates of cases $\widehat{Y}_{ct}$ and rates $\widehat{\mu}_{ct}$. Thresholds can be set for point estimates, or the upper bounds of either the confidence or prediction intervals, at some chosen significance level.

Thus a risk classification for country c at time t can be made according to:

1. The expected number of cases arriving at the border $\widehat{Y}_{ct}$;
2. The upper bound of a **confidence interval** for $\widehat{Y}_{ct}$ (this includes the uncertainty in the estimate of $\widehat{Y}_{ct}$)
3. The upper bound of a **prediction interval** for Y_{ct} (this includes **both** the uncertainty in the estimate of $\widehat{Y}_{ct}$, as well as the uncertainty in the observed value of the random variable Y_{ct})

Risk classification based on the expected number of **cases** at the border requires some knowledge of the expected number of **arrivals** N_{ct}. An alternative risk classification can be based only on the infection **rates** instead:

1. The expected infection rate amongst people arriving at the border $\widehat{\mu}_{ct}$;
2. The upper bound of a **confidence interval** for $\widehat{\mu}_{ct}$
3. The upper bound of a **prediction interval** for the observed rate Y_{ct}/N_{ct} (however this does require an estimate of N_{ct})

A suitable risk classification might have four classes, for example:

- Class 1. The expected number of cases $E[Y_{ct}]$ is less than 3 in a given week;
- Class 2. $3 \leq E[Y_{ct}] < 8$;
- Class 3. $8 \leq E[Y_{ct}] < 20$;
- Class 4. $E[Y_{ct}] \geq 20$;

If confidence or prediction intervals are used, it is suitable to use confidence intervals with confidence levels of the order of 50% (rather than, say 95%), so that the upper bound of the interval is the upper quartile. This avoids overly conservative risk classifications.

3.6 Low information countries

We cannot make good estimates using the above model for countries where there is low information, either due to low numbers of arrivals, low numbers of cases, or both.

The model described above is fitted only for countries with more than 50 arrivals and more than 5 cases in the most recent 30 weeks.

The fitted model has estimates of the parameters ($\boldsymbol{\beta}$, σ_u, σ_e, ρ) and the random effects u_c, v_{ct} for the modelled countries only. The fitted model also estimates the variance-covariance of the fixed effect parameters $\widehat{V}_\beta$.

In Sections 3.6.1 and 3.6.2 below we refer to parameter estimates from this data set as the main model estimates.

3.6.1 Low case numbers

For the remaining countries where there has been at least one arrival ($\sum_t N_{ct} > 0$) and at least one case ($\sum_t Y_{ct} > 0$), we remove the autoregressive AR(1) component from (2), and fit the simpler model to the data (N_{ct}, y_{ct}):

$$Y_{ct}|N_{ct}, \mu_{ct} \sim \text{Binomial}(N_{ct}, \mu_{ct}) \quad (18)$$

$$\eta_{ct} = \text{logit}\mu_{ct} = \mathbf{x}_{ct}^T\widehat{\boldsymbol{\beta}} + u_c \quad (19)$$

where u_c is a (fixed effect) parameter to be estimated, and the term $\mathbf{x}_{ct}^T\widehat{\beta}$ is treated as a known offset (i.e. $\boldsymbol{\beta}$ is not re-estimated).

We set u_c to the fitted value $\widehat{u}_c$ from the fitted model, and set all $v_{ct} = 0$. Thus

$$\widehat{\eta}_{ct} = \mathbf{x}_{ct}^T\widehat{\boldsymbol{\beta}} + \widehat{u}_c \quad (20)$$

$$\text{Var}[\widehat{\eta}_{ct}] = \widehat{S}_{ct}^2 \simeq \mathbf{x}_{ct}^T\widehat{V}_\beta\mathbf{x}_{ct} + \text{Var}[\widehat{u}_c] \quad (21)$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{V}_\beta$ come from the original fit and $\text{Var}[\widehat{u}_c]$ comes from the fit to the low case data.

Forecasting k time steps beyond the final observation at t_n uses the main model estimates of the AR(1) correlation $\widehat{\rho}$ and variance $\widehat{\sigma}_e^2$:

$$\widehat{\eta}_{ct_{n+k}} = \mathbf{x}_{ct_{n+k}}^T\widehat{\boldsymbol{\beta}} + \widehat{u}_c \quad (22)$$

$$\text{Var}[\widehat{\eta}_{ct_{n+k}}] = \widehat{S}_{ct_{n+k}}^2 \simeq \mathbf{x}_{ct_{n+k}}^T\widehat{V}_\beta\mathbf{x}_{ct_{n+k}} + \text{Var}[\widehat{u}_c] + \widehat{\sigma}_e^2\frac{1-\widehat{\rho}^{2k}}{1-\widehat{\rho}^2} \quad (23)$$

These formulae apply to all forecasts (including the one step ahead forecast $\widehat{\eta}_{ct|t-1}$).

3.6.2 Zero cases

For all other countries (where there have been zero cases) we set $u_c = 0$ and $v_{ct} = 0$. With no further model fitting we simply set

$$\begin{aligned} \widehat{\eta}_{ct} &= \mathbf{x}_{ct}^T \widehat{\beta} \\ \text{Var}[\widehat{\eta}_{ct}] = \widehat{S}_{ct}^2 &\simeq \mathbf{x}_{ct_n}^T \widehat{V}_\beta \mathbf{x}_{ct_n} \end{aligned}$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{V}_\beta$ come from the original main model fit.

Forecasting k time steps beyond the final observation at t_n uses the main model estimates of the AR(1) correlation $\widehat{\rho}$ and variance $\widehat{\sigma}_e^2$:

$$\widehat{\eta}_{ct_{n+k}} = \mathbf{x}_{ct_{n+k}}^T \widehat{\boldsymbol{\beta}} \tag{24}$$

$$\text{Var}[\widehat{\eta}_{ct_{n+k}}] = \widehat{S}_{ct_{n+k}}^2 \simeq \mathbf{x}_{ct_{n+k}}^T \widehat{V}_\beta \mathbf{x}_{ct_{n+k}} + \widehat{\sigma}_e^2 \frac{1 - \widehat{\rho}^{2k}}{1 - \widehat{\rho}^2} \tag{25}$$

These formulae apply to all forecasts (including the one step ahead forecast $\widehat{\eta}_{ct|t-1}$).

4 Implementation

The methods described in this documented have been implemented in R [24] in the `crater` R package which uses the `glmmTMB` R package [25] for model fitting.

An R Shiny app is available to explore the output object from the model fitting. The app allows the user to:

1. Choose a method of risk classification, and suitable thresholds (Section 3.5);
2. Adjust scenarios for numbers of future arrivals (Section 3.3).

The app then displays:

1. Fitted models by country, including estimates of cases and rates of infection. The app also displays the other available information (death rates, testing rates, vaccination coverage, reproduction number etc.);
2. Aggregate estimates of total number of cases arriving at the border, split by geographic region and risk classification;
3. A world map of risk classification for any selected week;
4. A set of risk classifications over time.

Table 2: Parameter estimates. The estimates on the left are returned by the fitting routine, and those on the right are useful transformations of those estimates. (Standard errors are calculated using the delta method.)

Parameter	Estimate	Std. Error	Parameter	Estimate	Std. Error
α	-1.1648	0.9706	α	-1.1647	0.9706
β_0	0.2373	0.3080	β_{ave}	0.4441	0.1108
β_1	0.0858	0.4038	β_{1-0}	0.0892	0.3013
β_2	0.1210	0.2948	β_{2-1}	0.0270	0.2905
$\log(\sigma_u)$	-0.2927	0.3591	σ_u	0.7462	0.2680
$\log(\sigma_e)$	0.3314	0.1275	σ_e	1.3929	0.1776
$\theta = \rho/\sqrt{(1-\rho^2)}$	0.7845	0.2266	ρ	0.6172	0.1104

5 Results

In this section we display the results for the model fitted to data in the time interval 25 January 2021 to 22 August 2021. Although data are available back to 8 Jun 2020 we only use the most recent 30 weeks of data when fitting the model. This is to allow for large scale changes in the situation in a country which cannot easily be allowed for in the model specification in Section 3: in particular this allows the country level random effect u_c to have a different value at different epochs.

Out of a total of 224 countries, 19 countries had sufficient numbers of arrivals and cases to be included in the full model, 44 otherwise had 1 case or more, and 161 had zero cases.

5.1 Fitted model

The parameter estimates and their standard errors are shown in Table 2.

The parameter estimates $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2$ show the dependence of the disease risk on disease incidence in the source country.

These are transformed in parameters $\widehat{\beta}_{\text{ave}}$, the effect of the mean logit(incidence) of the last three weeks, and parameters $\widehat{\beta}_{1-0}$ and $\widehat{\beta}_{2-1}$ which are the coefficients of the changes in logit(incidence) from lag 1 to lag 0, and lag 2 to lag 1 respectively.

We now consider the various aspects of the fitted model, and use the United Kingdom as an example.

5.2 Country level estimates

Firstly, Figure 3 displays the rate of new cases per week in the United Kingdom. Strong surges in cases are visible in October 2020, January 2021 and in July 2021. The Omi-

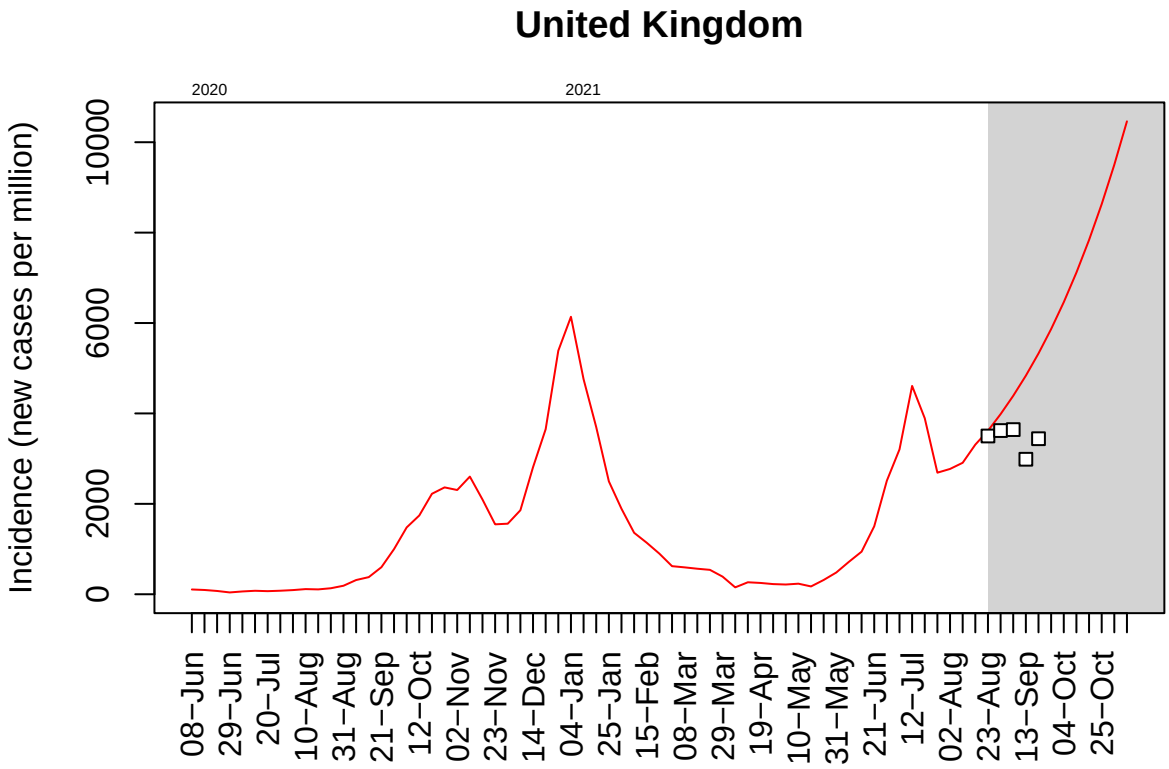


Figure 3: Incidence rates in the United Kingdom. Actual data are shown by the red line up to 22 August 2021, and a prediction continues into the shaded area at right. Actual observations in the shaded area are shown as white squares.

cron outbreak had not yet begun. At the right in this diagram, and in subsequent diagrams, is a shaded area which is a 12 week period immediately after the end of the observed data. We make forecasts into this period, and as a check on our findings compare with true observations in the first 5 weeks of the forecast period. We expect our model to provide reliable forecasts for only the first three weeks in the forecast period. The true incidence is shown by white squares. Incidence has been forecast into the period by a linear extrapolation of the logit of incidence: in this case we overestimate the incidence by the linear extrapolation.

Next, Figure 4 shows the numbers of arrivals in New Zealand coming from the United Kingdom per week. We assume that that the arrivals over the next 12 weeks will be an average of the number of arrivals during the last 3 weeks of the data set. The white squares are actual observations of the next 5 weeks of arrival volumes.

Figures 5 and 6 show the data on numbers of cases and infection rates among arrivals from the United Kingdom. In each graph the fitted model is shown as the bold red line with thin red lines bounding a 50% confidence interval. A 50% prediction interval is shown by the dashed purple lines.

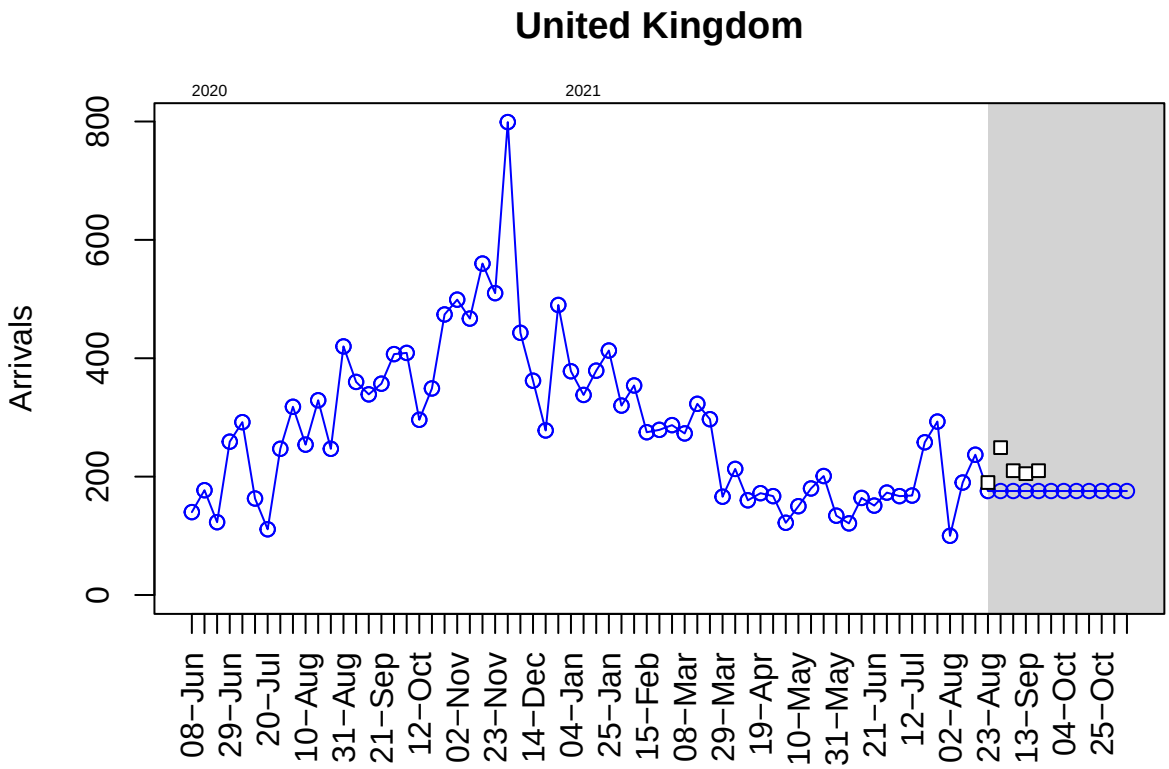


Figure 4: Arrivals from the United Kingdom. Actual data are shown by the blue line up to 22 August 2021, and a prediction continues into the shaded area at right. Actual observations in the shaded area are shown as white squares.

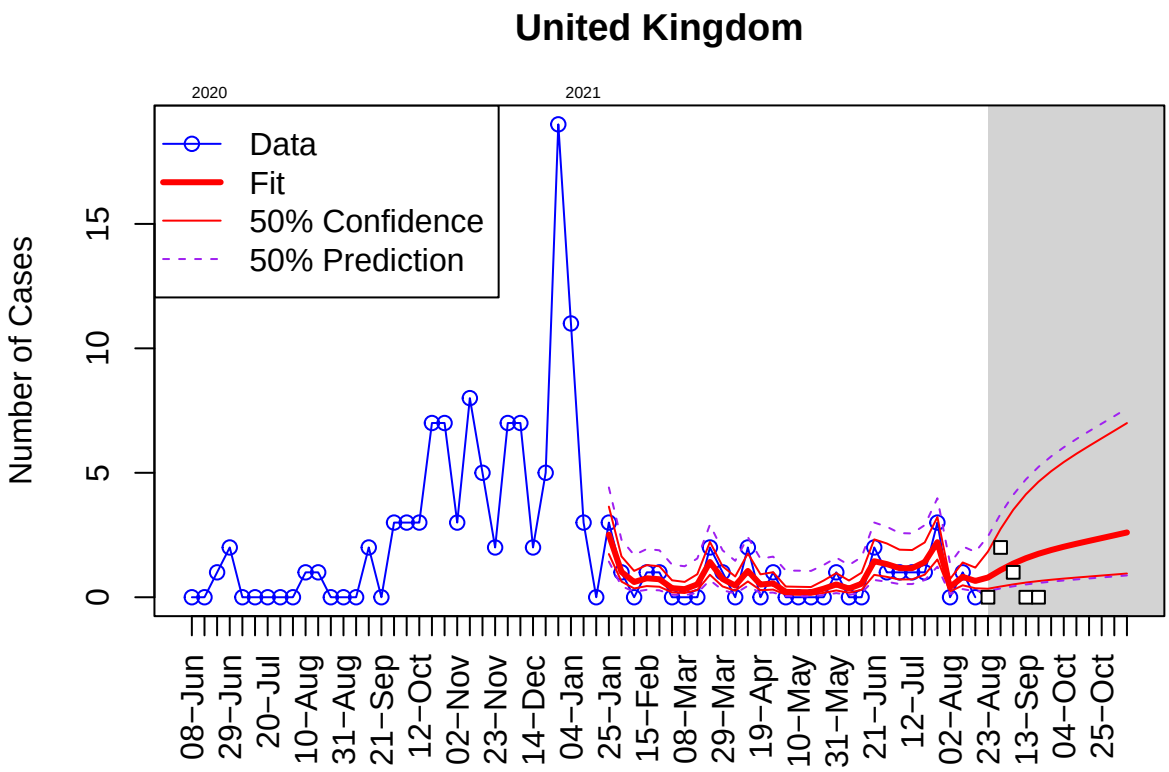


Figure 5: Cases arriving from the United Kingdom. The fitted number of cases is shown by the bold red line, with thinner red lines bounding a 50% confidence interval. A 50% prediction interval is shown by the dashed purple lines. Predictions are in the shaded region at right, with actual observations shown as white squares.

The model matches the data well throughout the range of the observations (up to 22 August 2021). The confidence and prediction intervals expand widely beyond the end of the observed data.

Figure 7 shows the random effects for the United Kingdom. In this graph the (time independent) country level random effect u_c is shown as the blue horizontal line. It is slightly negative (below the dashed line at zero), indicating that the infection rate among arrivals is in general somewhat less than the in country infection rate would predict on its own.

The time dependent random effect v_{ct} is an **autocorrelated** process: each value being in general closer to its immediate predecessor than it is to more distant observations. The value of v_{ct} is heavily influenced by the excursions in the observed infection rate among new arrivals, but the autoregressive structure resists sudden large movements. However in the prediction region after 22 August 2021 there is no data, and the random effect is pulled strongly and smoothly towards the horizontal blue line.

The random effects are the mechanism by which the infection rate between countries

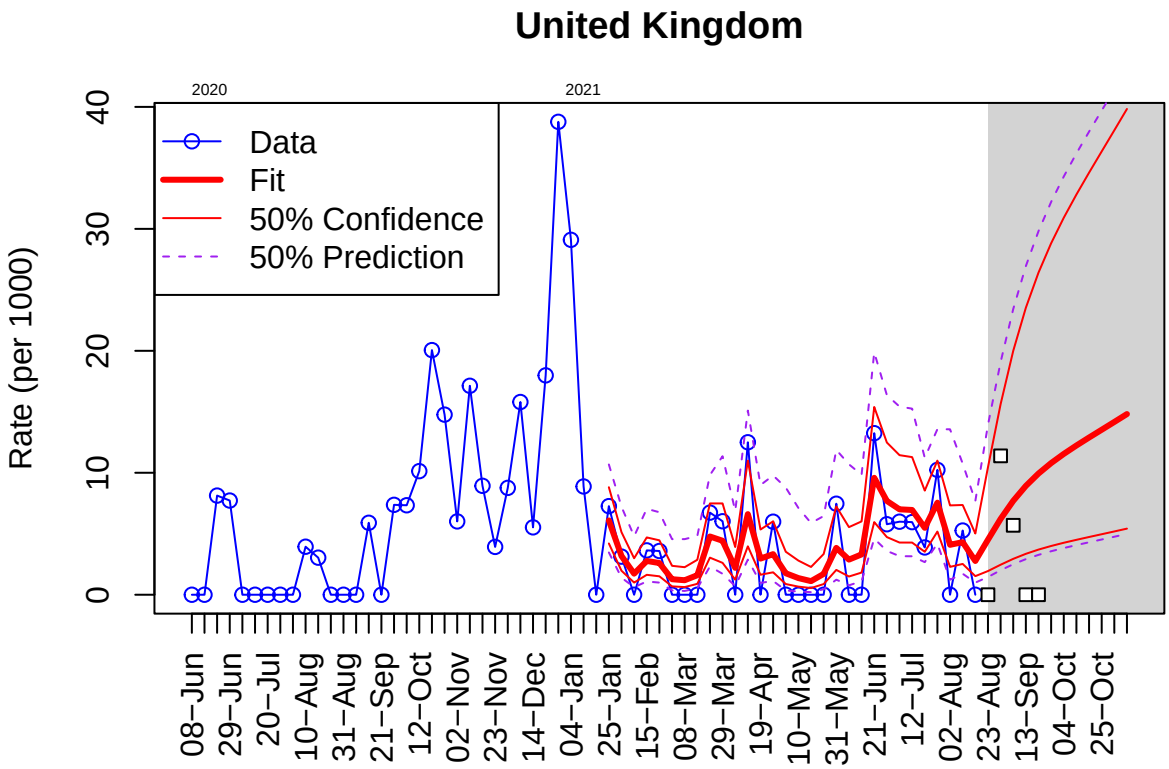


Figure 6: Infection rates among arrivals from the United Kingdom. The fitted rate is shown by the bold red line, with thinner red lines bounding a 50% confidence interval. A 50% prediction interval is shown by the dashed purple lines. Predictions are in the shaded region at right, with actual observations shown as white squares.

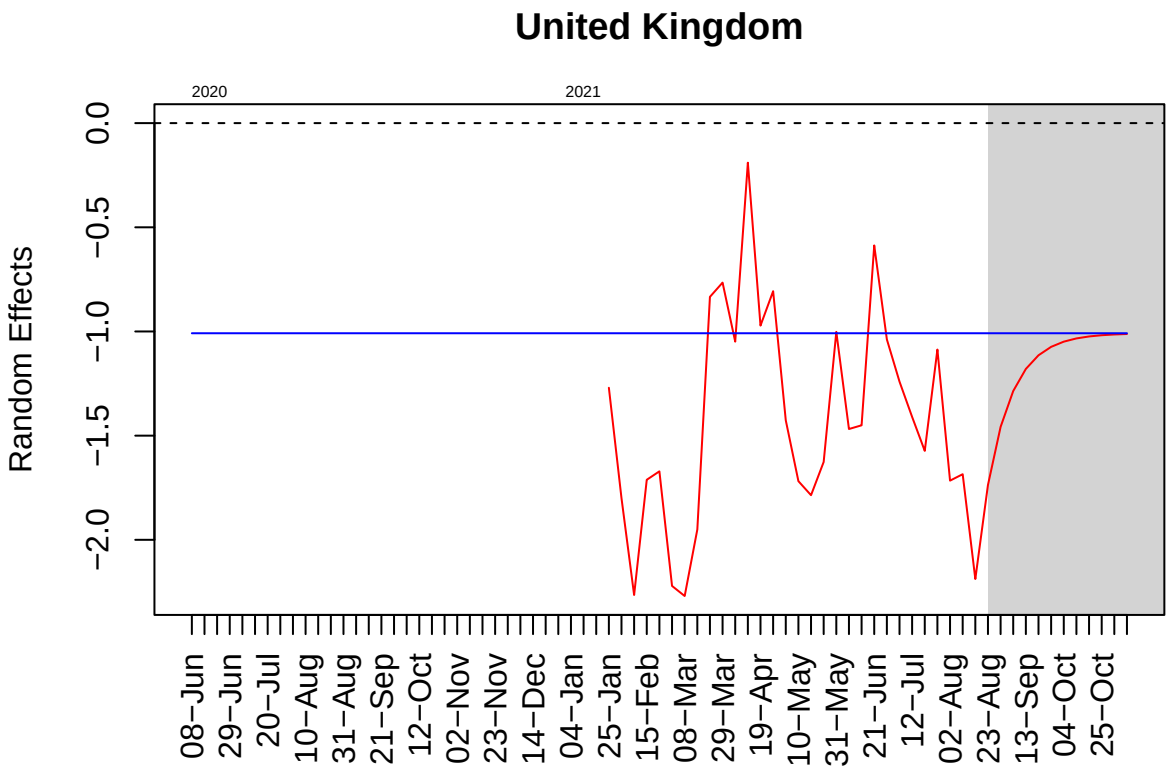


Figure 7: Random Effects for the United Kingdom. The constant random effect u is shown by the horizontal blue line. The time varying random effect v is shown by the red curve.

is modified by factors other than in country incidence. As noted already, there is a multitude of such factors. Firstly, in countries where the pandemic is fully under control, almost all cases will be in managed isolation, and the effective reproduction number R_{eff} will consequently be much less than 1. This might be viewed as an argument for including R_{eff} as a predictor in the model: however the estimation of R_{eff} is not reliable in all countries in the data set.

Similarly, the rates of testing, test positivity rate and vaccination rate are all likely to be associated with greater or lesser risk among arrivals, but the data that inform these measures is too unreliable to include these as predictors in the model.

Also of importance are any factors that mean travellers from a source country differ in their risk profile from the average resident in that country. This difference may be due to factors imposed by policies imposed by the New Zealand government (e.g. mandatory negative COVID tests, mandatory vaccination), but are also due to individual traveller characteristics (place of residence in the source country, ability to isolate from risk etc.).

On top of this, the relationships between these variables and infection risk are likely to be strongly time dependent, highly correlated and thus difficult to estimate separately from one another.

The random effects model absorbs all of these country specific and time dependent effects into measures which learn the degree of autocorrelation and the magnitude of variation of differences between arrival risk and in country disease incidence.

Figure 8 shows some of the other information available about the UK, but not included in the model: death rates, R_{eff}, testing rates, test positivity rates and vaccination rates. This information is useful as a supplement to the predictions of the model.

One step ahead predictions are ways of assessing the goodness of fit of the model. At each time step we use the current covariates and parameter estimates, but forecast the time-dependent random effect forwards from the previous week. These are shown for cases and rates in Figures 9 and 10 respectively, and can be compared to the fitted values in Figures 5 and 6. (Note that in the prediction region of Figures 9 and 10 we plot the same $k-$step ahead predictions as in Figures 5 and 6.)

As is typical of one step ahead predictions, they appear to slightly lag the true data, with each observation predicting forwards a modified version of its own value.

5.3 Risk classification

Using the estimates for the expected numbers of cases and the expected infection rates, we can assign each country to a risk category using specified thresholds. Here we use the upper bound of a 50% confidence interval for the infection rates to classify countries into four groups. The cut-points between the groups are 3, 8 and 20 cases per thousand arrivals.

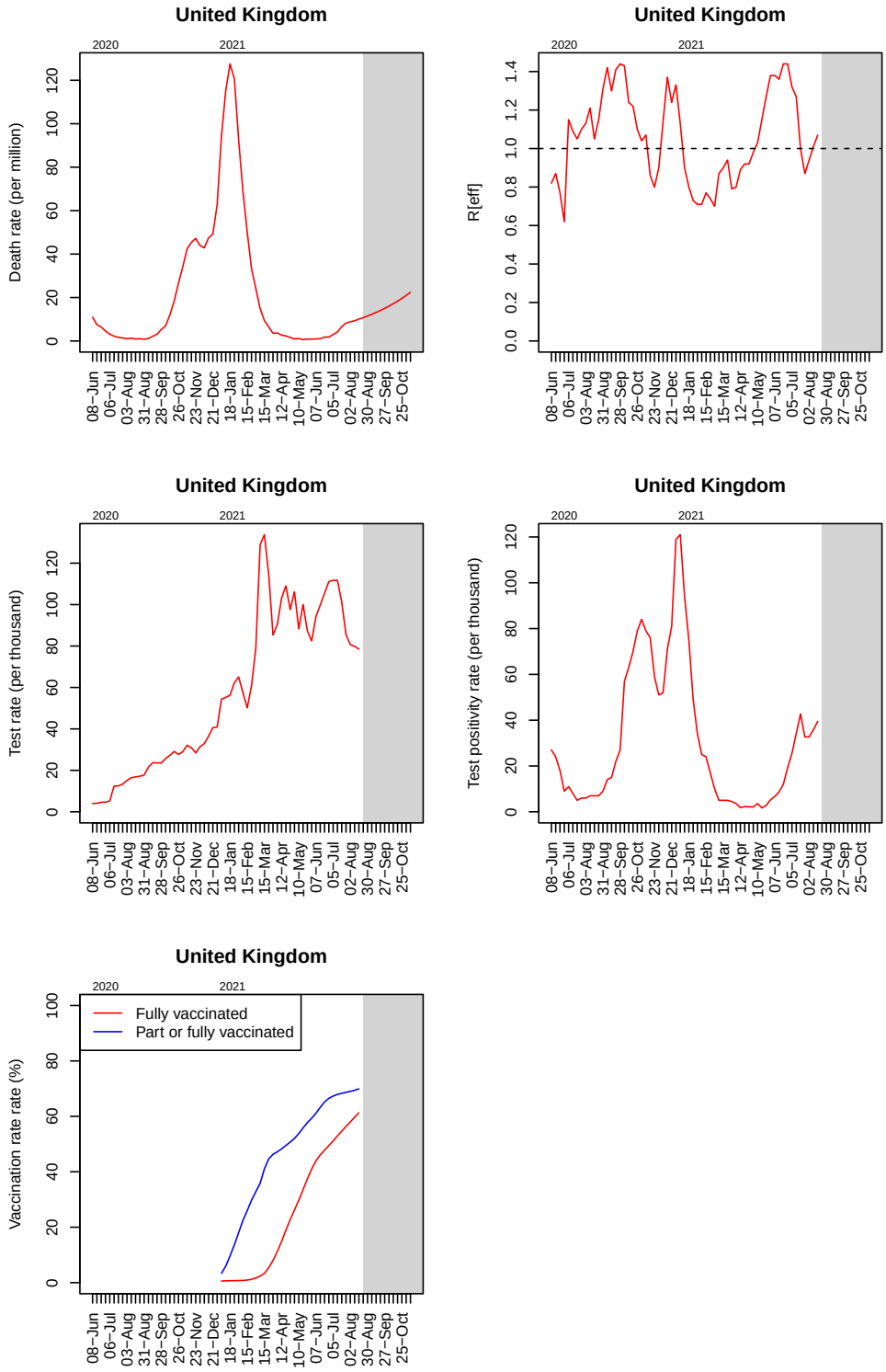


Figure 8: Factors associated with disease risk among arrivals from the United Kingdom

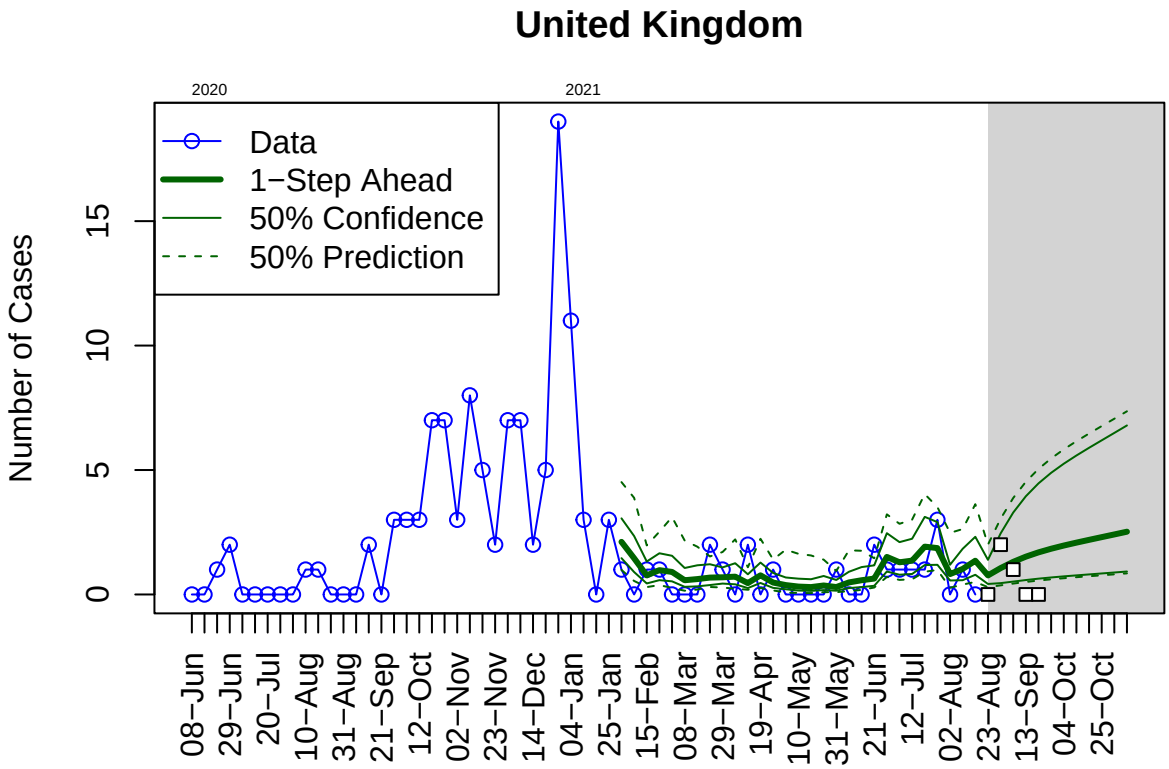


Figure 9: Cases arriving from the United Kingdom. The one step ahead prediction of the number of cases is shown by the bold green line, with thinner green lines bounding a 50% confidence interval. A 50% prediction interval is shown by the dashed lines. The predictions at right are all predicted from data up to 22 August 2021, with actual observations shown as white squares.

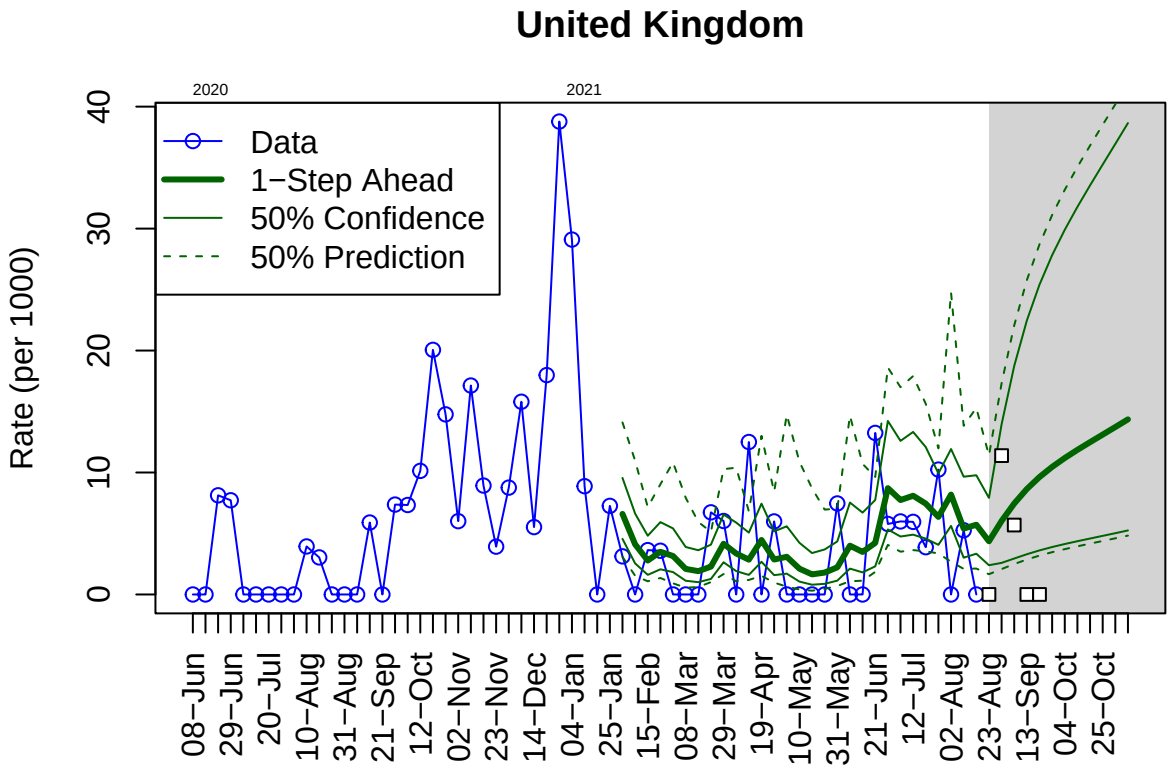


Figure 10: Infection rates among arrivals from the United Kingdom. The one step ahead prediction of the disease rate rate is shown by the bold green line, with thinner green lines bounding a 50% confidence interval. A 50% prediction interval is shown by the dashed lines. The predictions at right are all predicted from data up to 22 August 2021, with actual observations shown as white squares.

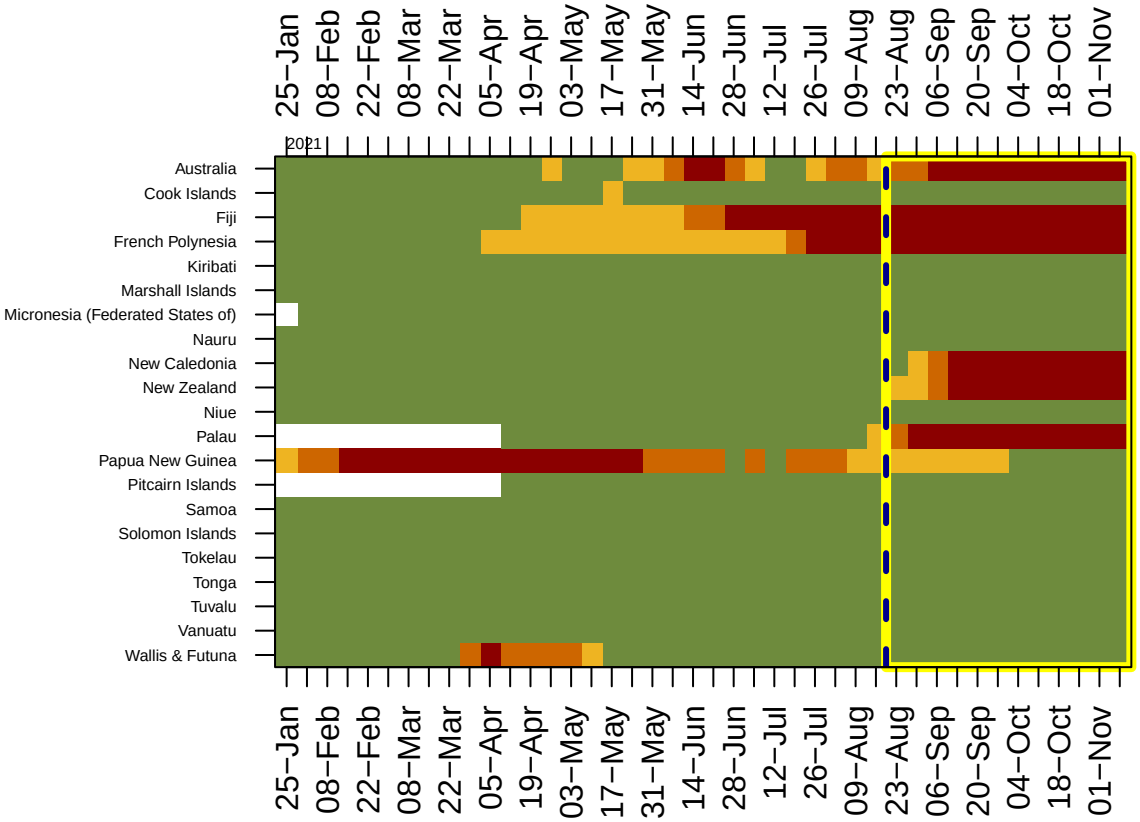


Figure 11: Risk categorisation for countries in Oceania. Green is the lowest risk, dark red is the highest. Forecasts are in the area bounded by the yellow box at right.

Figure 11 shows the changing classification over time for modelled countries in Oceania at the beginning of the Delta variant wave. The forecast region is bounded by a yellow box at right. In this prediction the risk categorisations of New Caledonia and Palau are expected to increase from the lowest category (green) to the highest (dark red) in the near future. New Zealand, though not included in the modelling, is included and is assessed by the same criteria, showing the anticipated increase in disease incidence. Fiji, French Polynesia, and (to lesser extents) Australia and Papua New Guinea are classified as current and continuing risks.

A world risk map is shown in Figure 12 for the week starting 23 August 2021. The risk classification is based on the upper bound of the confidence interval for infection rates on arrival. Countries for which there is insufficient data to calibrate a rate are shown in white.

Figure 13 shows the country level random effects, in decreasing order. Those at the top of the graph have arriving infection rates that are higher than their in country incidence would suggest. Those towards the bottom of the graph have lower rates.

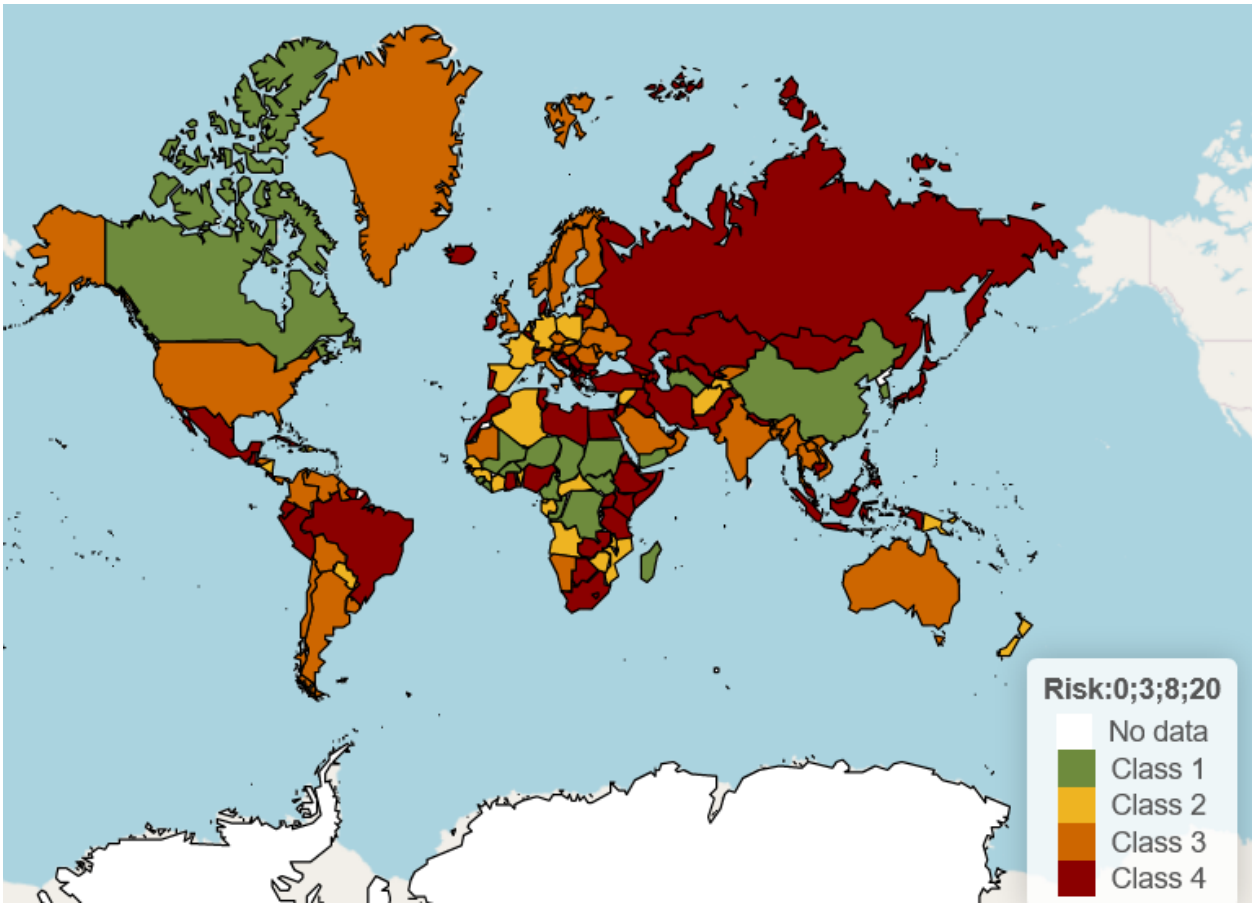


Figure 12: Map of border risk for countries across the world for the week starting 23 August 2021. Green is the lowest risk, dark red is the highest.

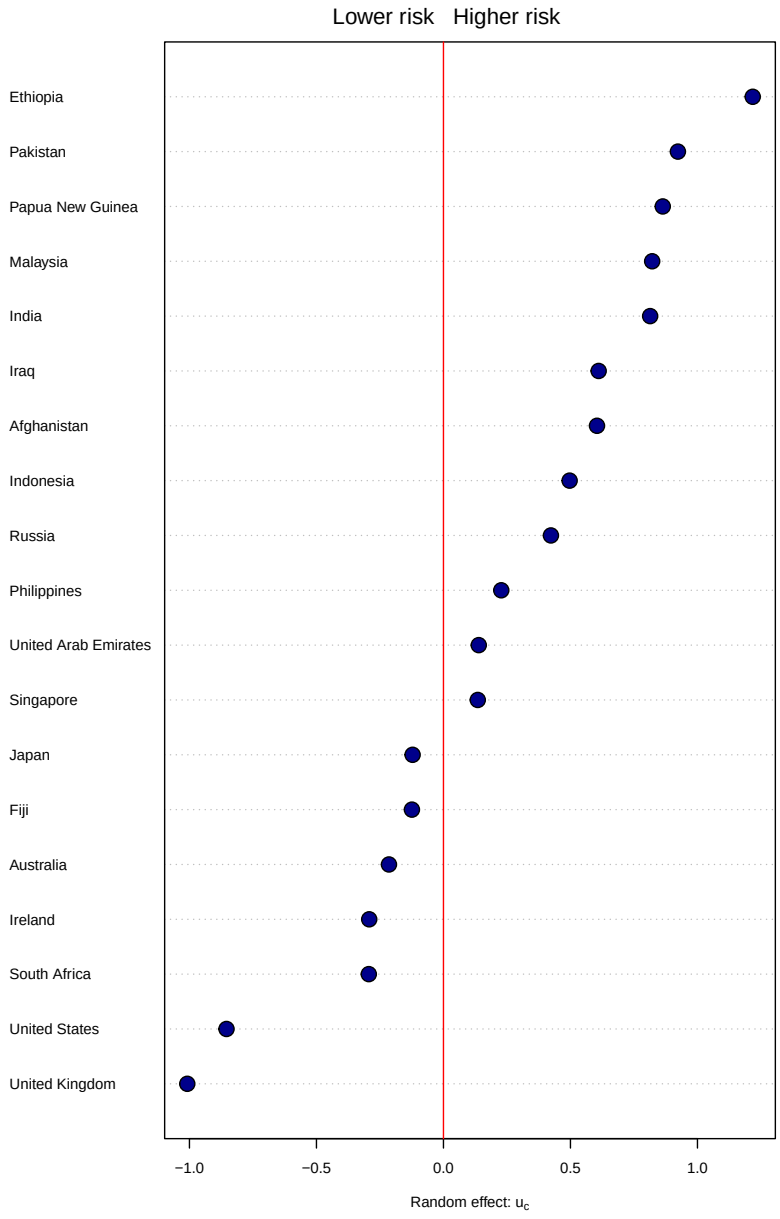


Figure 13: Country level random effects, u_c, for the 19 fully modelled countries using data from the period 25 January-22 August 2021

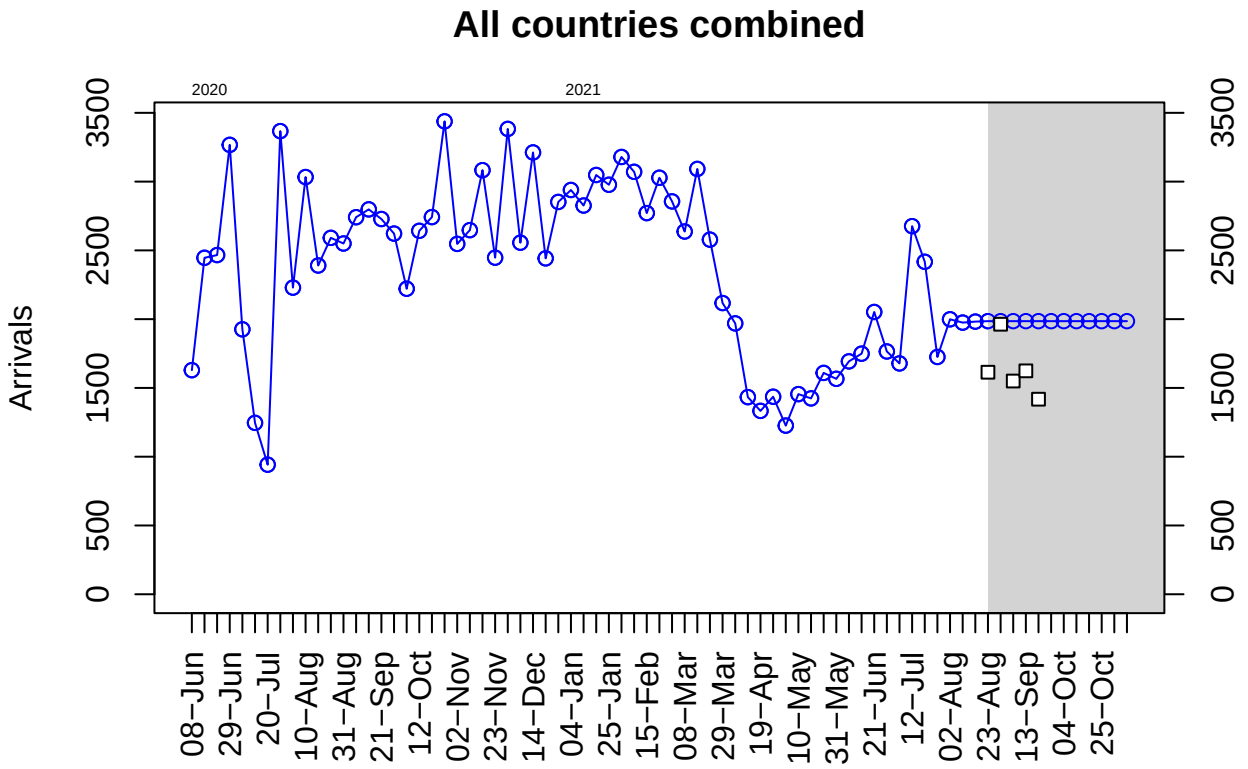


Figure 14: Total arrivals from all countries at the New Zealand border. Actual arrivals are shown up to 22 August 2021, modelled arrivals are shown in the grey area at right, with the actual arrivals for the first five weeks of the forecast period shown as white squares.

5.4 Aggregate risk estimates

The country level estimates can be aggregated over all countries to create estimates of the total numbers of cases that are expected at the border each week. These estimates include all fully modelled countries as well as those with low and zero case counts.

Figures 14, 15 and 16 display these total results in formats analogous to the country specific Figures 4, 5 and 6: showing the data, the fitted model, confidence and prediction intervals, and the true data for the first 5 weeks beyond the end of the data set. The opening and closing of a period of quarantine free Australian travel of April-July 2021 is clearly visible in Figure 14.

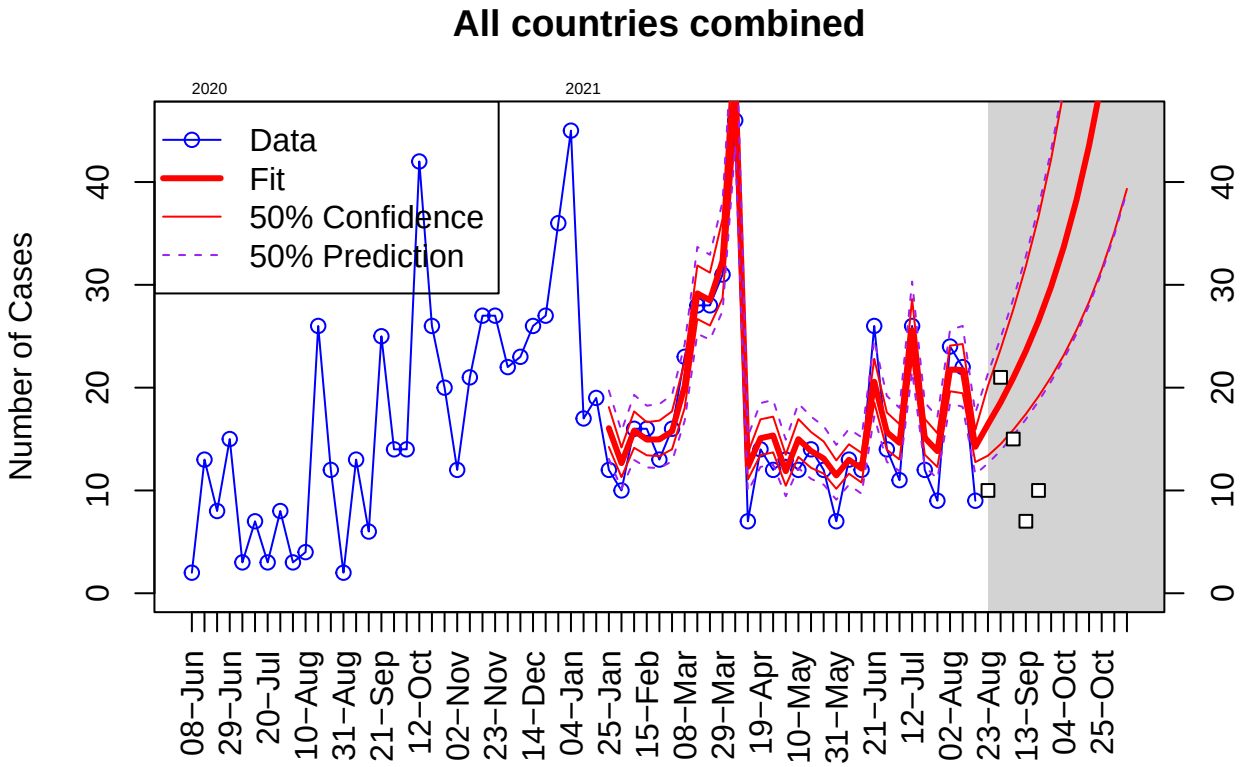


Figure 15: Total cases arriving at the border. Predictions are in the shaded area at right, and actual observations shown as white squares.

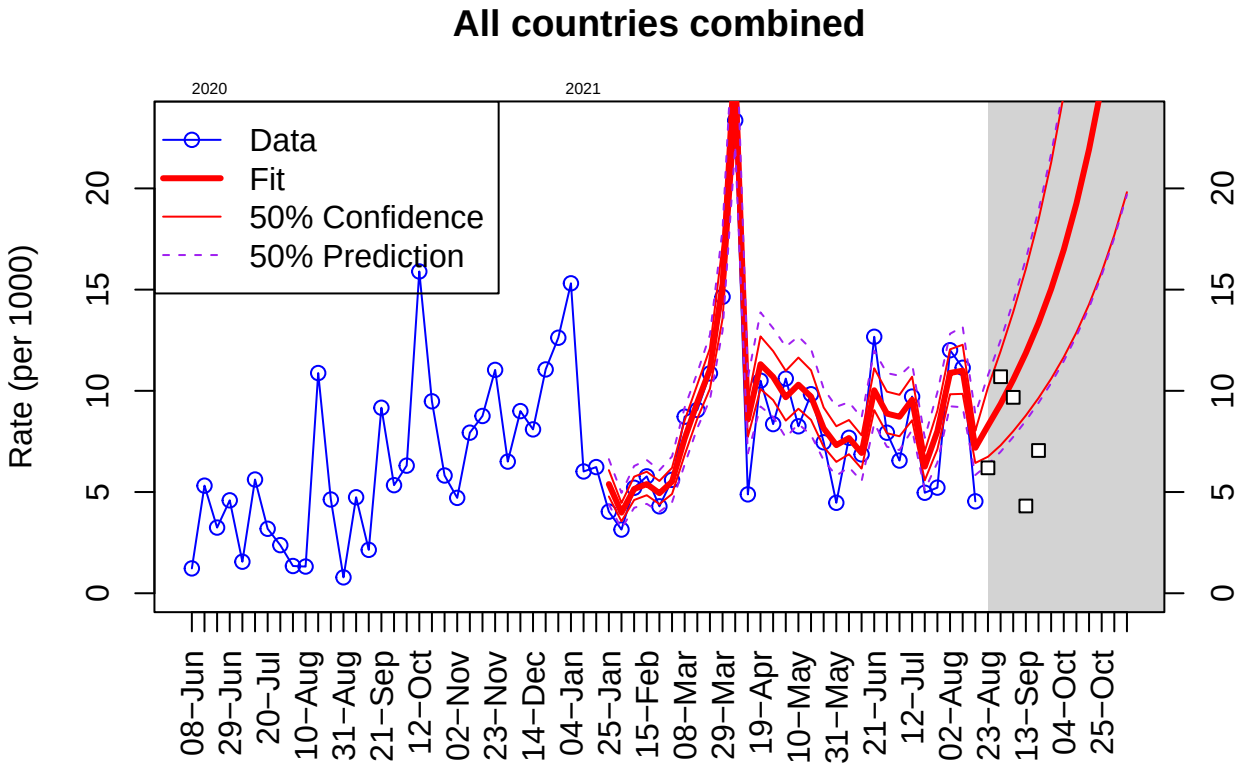


Figure 16: Infection rates among arrivals at the border. Predictions are in the shaded area at right, and actual observations shown as white squares.

6 Discussion

The risk assessment model presented here is a robust and practical tool for forecasting the disease arrival risk at the New Zealand border. It relies only on readily available data (reported incidence rate in the source countries I_{ct}, together with counts of arrivals N_{ct} and cases Y_{ct} by week and country). The forecasts are in the form of expected rates and counts by country, and are accompanied with suitable measures of uncertainty. Combined with other information (effective reproduction number, death rates, testing rates etc) policy makers can make informed decisions about disease risk among source countries.

The COVID19 pandemic has shown very rapid changes, and there have been very diverse policy responses across the world. This led us to a modelling approach which was simple (using incidence as the only covariate, with three lags to measure disease trajectory), with a random effects structure that absorbs many of the differences between countries that are impossible to model adequately. These differences include the differing testing rates between countries which mean that the reported incidence in some countries may differ strongly from the true incidence. Any systematic differences between the resident population (which generates the incidence statistics) and the travelling population (to which New Zealand's borders are exposed) are also confounded with potentially unreliable reported incidence. Travellers are likely be healthier and wealthier than the general population of the country from which they come. The random effects structure, including a fixed effect over a 30 week time window for each country, as well as an autoregressive effect to allow for temporal variations in these same differences, allows us to estimate the contrast between in country incidence and disease risk at the New Zealand border.

Changing risks due to differences in the properties of the variants of COVID19, and the changing levels of immunity due to vaccination programmes (including their waning of effectiveness over time) do not need to be modelled explicitly. Instead the random effects structure learns from the data how the risk from each country is changing over time.

Air travellers from different countries share airports and aircraft with each other, and in-transit risks may pose significant risks to individual travellers from low risk countries as they mix with travellers from higher risk countries. If travellers from a particular country tend to use the same routes, and mix with passengers from the same set of other countries, then to some extent this inflight risk is incorporated in the random effects structure as a component of the difference between the travelling population at the at home population from any given country. Beyond this observation, and the fact that air travel to New Zealand currently requires passengers to be vaccinated, we cannot see a reliable source of data by which we could incorporate inflight transmission into our model.

As noted above, the outputs of the model are not on their own sufficient to provide an automatic risk classification that should be used without consideration of other information such as reports of sudden changes in disease rates, or the emergence of

a new variant, in a source country.

Compared to countries with land borders, or those at short sea distances from their neighbours, the geographical isolation of New Zealand puts it in an excellent position to be able to monitor and forecast COVID19 border risk. Models created in other countries generally focus on the risks that imported cases pose to the pandemic within that country (e.g. [26, 27]), and this question has also received consideration in New Zealand (e.g. [28]). However, specific models designed for the quantitative assessment of risks are rare. Lee et al. [29], for example, created a country risk model for arrivals to South Korea where the arrival risk was simply proportional to the monthly in-country incidence.

Other modelling approaches exist as well, with Wang et al. [30] aggregating risk along the route taken by an arriving ship based on the current case numbers and rates of change at the ports visited. Zhang et al. [31] use a model incorporating the connectivity of the international air travel network to assess border risk at provincial level in China. Quilty et al. [32] investigate the aggregate risk of arrivals from all origins, calibrated using flight data, with an interest in optimal testing policies for international arrivals. Their study uses the methods of Russell et al. [33, 34] to account for in-country case underascertainment using death rates rather than reported incidence or prevalence. Their methodology allows for underascertainment (particularly of mild cases), assuming a case fatality rate of 1.4%, modified for reporting delays.

The International Civil Aviation Organisation (ICAO) has published suggestions for creating a country risk classification [35]. They propose a four level classification system based on the percentage of non-immune persons, the 7 day prevalence, the test positivity rate that the testing rate. Classification is based on a set of thresholds of these measures. Where countries have imposed differential treatment of arrivals from source countries, as opposed to treating all arrivals in the same way, it is likely that some version of rules based on these measures has been applied.

Our methodology provides a finer calibration of the border risk posed by arrivals from every country by incorporating the additional information gained from observing recent arrivals. Our methodology does not rely on assumptions such as a fixed case-fatality ratio, nor a fixed level of case underascertainment. Such assumptions need constant revision when case monitoring and registration practices change, when new variants emerge, when new treatments are made available, and as vaccination coverage increases and then wanes in effectiveness. Such changes affect border risk, but cannot be easily or separately estimated. Since our method is directly calibrated by actual arrivals we estimate the combined effect of these changes in the random effects structure.

Our approach is of course at risk of missing rapid changes in source countries since it takes time for the model to adjust to abrupt shifts in the level of risk, and we reiterate our view that border risk assessment decisions need to rely on a wider range of data sources. Thus we agree with the advice given in the ICAO report that '[a]lthough data-driven decision making is encouraged, the current scenario may require a qualitative approach, as validated data and information is incomplete' [35].

Acknowledgements

The authors acknowledge the support of the New Zealand Ministry of Health, StatsNZ, and the Institute of Environment Science and Research in supplying data in support of this work. The authors are grateful to Samik Datta, Melissa McLeod, Fraser Morgan, Nigel French, Anja Mizdrak and Markus Luczak-Roesch, and the COVID-19 Modelling Government Steering Group for feedback on earlier versions of this report. This work was funded by the New Zealand Department of Prime Minister and Cabinet.

A Appendix: Model Selection

In this appendix we briefly describe details of the model selection procedure, to supplement the model description given in Section 3.

The available predictor variables are, for time period t and country c:

- Death rate, D_{ct}
- Incidence rate, I_{ct}
- Prevalence rate, P_{ct}
- Test rate, TR_{ct}
- Test positivity rate, TP_{ct}
- Partial vaccination rate, PV_{ct}
- Fully vaccinated rate, FV_{ct}
- Effective Reproduction number, R_{ct}

All are available at each time period. Where daily estimates are available we aggregate to weekly values by taking the weekly median.

We model data only for 30 week periods in order to avoid very long term changes in the pandemic.

We show the model selection process here for a data set restricted to arrivals in the 30 week period starting 26 April 2021 and ending on 21 November 2021. In this period there were 388 cases among 53275 arrivals from 185 countries.

We further restrict modelling to countries where there were at 5 or more cases arriving during the period. This reduces the data to 298 cases among 32825 arrivals from 17 countries. These are 62% of the total number of arrivals, and 77% of the cases. (The list of modelled countries is Afghanistan, Australia, Fiji, India, Indonesia, Iraq, Japan, Malaysia, Philippines, Qatar, Russia, Singapore, South Africa, Sri Lanka, United Arab Emirates, United Kingdom, United States.)

When comparing time series models where prediction is the goal prediction errors are typically used rather than standard (relative) goodness of fit measures such as AIC. The use of AIC risks overfitting the model to the observed sample data, rather than minimising the error of predictions.

Here we use the mean absolute deviation (MAD) of one step ahead prediction errors to measure goodness of fit, and as the criterion for model selection when determining which predictors to include in the final model.

When fitting the model we estimate the linear predictor

$$\widehat{\eta}_{ct} = \mathbf{x}_{ct}^T\widehat{\boldsymbol{\beta}} + \widehat{u}_c + \widehat{v}_{ct}$$

at each time point for each country, along with its standard error $\widehat{S}_{ct}$. The one step ahead forecast is

$$\widehat{\eta}_{ct|t-1} = \mathbf{x}_{ct}^T\widehat{\boldsymbol{\beta}} + \widehat{u}_c + \widehat{\rho}\widehat{v}_{c,t-1}$$

with approximate estimation standard error $\widehat{S}_{ct|t-1} = \sqrt{\widehat{S}_{ct}^2 + \widehat{\sigma}_e^2}$.

On the observation scale the fitted value and one step ahead fitted value are

$$\begin{aligned} \widehat{Y}_{ct} &= N_{ct}\text{expit}(\widehat{\eta}_{ct}) \\ \widehat{Y}_{ct|t-1} &= N_{ct}\text{expit}(\widehat{\eta}_{ct|t-1}) \end{aligned}$$

The mean absolute deviation is then defined

$$\text{MAD} = \frac{1}{CT}\sum_{ct}|Y_{ct} - \widehat{Y}_{ct|t-1}|$$

where p is the number of parameters in the model, C is the number of countries and T is the number of weeks.

We omit from the sum (and adjust the $1/(CT)$ denominator) any observations where $N_{ct} = 0$, and for which therefore $\widehat{Y}_{ct|t-1} = 0$.

In order to compare models we need a subset of the data for which all the potential covariates are present. There is a substantial amount of missing data in vaccine coverage and population testing, with some countries not reporting these data consistently, and in some cases not all.

We have tested two data sets in selecting the model: each one 30 weeks long. The first contains weeks 38 to 67 (28 Sep 2020-25 Apr 2021) and the second weeks 68-97 (26 Apr 2021-21 Nov 2021). In each data set we restricted to countries with at least 50 arrivals (22 countries in the first data set, and 17 in the second). In any instances country/week combinations where any covariates were missing we set the number of arrivals and cases for that country in that week to zero. This means that those rows did not influence the estimation process.

We carried out forward and backwards model selection, analogous to model selection by an information criterion such as AIC.

In the forward selection procedure we started with a model which included no covariates and then added covariates to the model one by one, as long as the MAD value decreased. At each step we selected the covariate that resulted in the greatest reduction in MAD. The backwards model selection proceeded similarly, but starting with

Table 3: Model selection by MAD. A blank indicates that the relevant covariate was not included in the final model. Data Set 1 covers 28 Sep 2020-25 Apr 2021 and Data Set 2 covers 26 Apr 2021-21 Nov 2021. ($\delta = 10^{-7}$ is a small offset to prevent taking the log of zero.)

	Data Set 1; forwards		Data Set 1: backwards		Data Set 2: forwards		Data Set 2: backw	
Parameter	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
(Intercept)	-7.80	1.6731	-7.80	1.6731	-2.287	0.998	-2.522	1.083
logit(δ + $I_{c,t}$)	1.11	0.6721	1.11	0.6721			-1.461	1.556
logit(δ + $I_{c,t-1}$)	-0.54	0.8885	-0.54	0.8885			2.407	1.393
logit(δ + $I_{c,t-2}$)							-1.366	1.239
logit(δ + $I_{c,t-3}$)	0.81	0.4672	0.81	0.4672	0.434	0.132	0.848	0.783
logit(δ + $D_{c,t}$)	-1.11	0.2540	-1.11	0.2540				
logit(δ + PV_{ct})							0.255	0.370
logit(δ + FV_{ct})	-0.0493	0.0421	-0.0493	0.0421	-0.341	0.113	-0.568	0.378
log(R_{eff})							1.049	2.706
σ_u	9.33e-05	0.2975	9.33e-05	0.2975	0.404	0.312	0.308	0.536
σ_e	1.14	0.1771	1.14	0.1771	1.266	0.194	1.283	0.242
ρ	0.496	0.1270	0.496	0.1270	0.430	0.163	0.438	0.172
AIC	355		355		666		672	
MAD	0.971		0.971		0.620		0.614	

the model with all covariates, successively deleting the one which led to the greatest reduction in MAD, and terminating when no further removals reduced the MAD.

The results of the forward and backward model selection algorithms applied to the two data sets are shown in Table 3. The forward and backward methods select the same model for Data Set 1, whereas there are some differences for Data Set 2, with the backwards method preferring a more complex model.

For comparison in Table 4, the parameter estimates, AIC and MAD values are shown for a model including only incidence at lags 0, 1 and 2, and no other covariates. The MAD values differ very little from any of those in Table 3, and are therefore just as good as those more complex models from the point of view of one step prediction.

The incidence only models also have the advantage that they rely only on covariates that are readily available, and are the most complete both across countries and across time. For this reason we have opted to use the three lag incidence modelling for our border risk assessment.

Table 5 shows the effect on the values of AIC and MAD of successively adding country level and the autoregressive random effects to a model with lagged incidence as the only predictors. The inclusion of random effects at country level is strongly supported by both AIC and MAD. The justification for serial correlation is weaker using the MAD criterion: being weakly supported in Data Set 2, but not in Data Set 1. We have nevertheless retained the serial correlation in the main model in order to allow for short term excursions in infection risk.

Table 4: Parameter estimates for a model with three lags of incidence as the only predictors to the two data sets.

	Data Set 1		Data Set 2	
Parameter	Estimate	Std. Err.	Estimate	Std. Err.
α	-1.779	1.448	-2.444	1.109
β_0	1.081	0.724	-0.756	0.687
β_1	-0.936	1.055	2.034	1.222
β_2	0.267	0.708	-0.938	0.715
σ_u	0.358	1.931	0.771	0.247
σ_e	1.426	0.499	1.173	0.186
ρ	0.785	-0.256	0.485	0.161
AIC	367.311		672.092	
MAD	1.073		0.637	

Table 5: Goodness of fit for models from Table 4 with and without random effects

	Data Set 1		Data Set 2	
Model	AIC	MAD	AIC	MAD
No Random Effects	656	1.38	928	0.844
+ Country Level RE	440	1.04	727	0.663
+ AR(1) Correlation	367	1.07	672	0.637

References

[1] Wikipedia. COVID-19 pandemic in New Zealand;. Available from: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_New_Zealand.

[2] Tian D, Sun Y, Xu H, Ye Q. The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 Omicron variant. Journal of Medical Virology. 2022;Online Early.

[3] RNZ. NZ to close doors on foreign travellers from China; 2020. Available from: https://www.rnz.co.nz/news/national/408675/nz-to-close-doors-on-foreign-travellers-from-china.

[4] Cooke H. Coronavirus: Government shuts borders to all but citizens and residents; 2020. Available from: https://www.stuff.co.nz/national/health/coronavirus/120423903/coronavirus-prime-minister-to-make-announcement-on-border-controls.

[5] Ministry of Health. History and origins of MIQ; 2022. Available from: https://www.miq.govt.nz/about/history-and-origins-of-miq.

[6] One News. Isolation hotel vouchers now required to enter NZ, with pre-Christmas period completely booked; 2020. Available from: https://www.tvnz.co.nz/one-news/new-zealand/isolation-hotel-vouchers-now-required-enter-nz-pre-christmas-period-completely-booked.

[7] RNZ. Covid-19: UK and US travellers to need pre-departure tests before departing to NZ; 2021. Available from: https://www.rnz.co.nz/news/national/433968/covid-19-uk-and-us-travellers-to-need-pre-departure-tests-before-departing-to-nz.

[8] NZ Herald. Covid 19 coronavirus: Government announces new Covid testing rules; 2021. Available from: https://www.nzherald.co.nz/nz/covid-19-coronavirus-government-announces-new-covid-testing-rules/C4QQPKH2I6J7HZC3ODHBJBNTOY.

[9] Early M. Covid-19: More people now eligible for emergency slots in managed isolation; 2021. Available from: https://www.stuff.co.nz/national/health/coronavirus/300267231/covid19-more-people-now-eligible-for-emergency-slots-in-managed-isolation.

[10] Newshub. Coronavirus: Jacinda Ardern announces suspension of arrivals from India for two weeks; 2021. Available from: https://www.newshub.co.nz/home/politics/2021/04/coronavirus-stopping-flights-from-india-not-ruled-out-as-cases-in-miq-from-asian-nation-jump.html.

[11] NZ Herald. New Zealand and Australia travel bubble to begin on April 19 - Prime Minister Jacinda Ardern; Air NZ 'run off our feet'; 2021. Available from: https://www.nzherald.co.nz/nz/new-zealand-and-australia-travel-

bubble-to-begin-on-april-19-prime-minister-jacinda-ardern-air-nz-run-off-our-feet/CE5FN7UHT23NEY2VVKQZWANFQI/.

[12] Ardern J, Hipkins C. Quarantine Free Travel with Australia suspended; 2021. Available from: https://www.beehive.govt.nz/release/quarantine-free-travel-australia-suspended.

[13] Hipkins C. Making the border stronger; 2021. Available from: https://www.beehive.govt.nz/release/making-border-stronger.

[14] Immigration New Zealand. Restriction on travel from very high risk countries (28 Aug 2021); 2021. Available from: https://web.archive.org/web/20210828043928/https://www.immigration.govt.nz/about-us/covid-19/border-closures-and-exceptions/restrictions-on-travel-from-very-high-risk-countries.

[15] NZ Herald. Covid 19 coronavirus: Cook Islands travel bubble to begin on May 17 - PM Jacinda Ardern; 2021. Available from: https://www.nzherald.co.nz/nz/covid-19-coronavirus-cook-islands-travel-bubble-to-begin-on-may-17-pm-jacinda-ardern/I5KFWNXG57LFITF3U46HBDUQ5Y.

[16] Hipkins C. Fiji and Indonesia designated very high risk countries due to COVID-19 risk; 2021. Available from: https://www.beehive.govt.nz/release/fiji-and-indonesia-designated-very-high-risk-countries-due-covid-19-risk.

[17] Manch T. Government will require non-NZ citizens to have Covid vaccine before entering New Zealand; 2021. Available from: https://www.stuff.co.nz/national/health/coronavirus/126566572/government-will-require-nonnz-citizens-to-have-covid-vaccine-before-entering-new-zealand.

[18] Ministry of Health. COVID-19 Protection Framework (traffic lights); 2021. Available from: https://web.archive.org/web/20211129070056/https://covid19.govt.nz/traffic-lights/covid-19-protection-framework.

[19] RNZ. Covid-19: Major MIQ changes from early next year; 2021. Available from: https://www.rnz.co.nz/news/national/456430/covid-19-major-miq-changes-from-early-next-year.

[20] Global Change Data Lab. Our World in Data - COVID19 data set;. Available from: https://github.com/owid/covid-19-data/raw/master/public/data/.

[21] Funk S, Endo A, Robert A, Gruson H, Munday J, Sherratt K, et al.. Epiforecasts: Real-time modelling and forecasting of infectious disease dynamics;. Available from: https://epiforecasts.io/.

[22] Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal. 2017;9(2):378–400. Available from: https://journal.r-project.org/archive/2017/RJ-2017-066/index.html.

[23] Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM. TMB: Automatic Differentiation and Laplace Approximation. Journal of Statistical Software. 2016;70(5):1–21. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v070i05.

[24] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2019. Available from: https://www.R-project.org.

[25] Kristensen K, Bell B, Skaug H, 9 others. R Package: Template Model Builder: A General Random Effect Tool Inspired by ADMB;. Available from: https://kaskr.github.io/adcomp/Introduction.html.

[26] Koutsellis T, Nikas A. A predictive model and country risk assessment for COVID-19: An application of the Limited Failure Population concept. Chaos, Solitons and Fractals. 2020;140:110240.

[27] Zhu Z, Weber E, Strohsal T, Serhan D. Sustainable border control policy in the COVID-19 pandemic: A math modeling study. Travel medicine and infectious disease. 2021;41:102044.

[28] Steyn N, Plank MJ, James A, Binny RN, Hendy SC, Lustig A. Managing the risk of a COVID-19 outbreak from border arrivals. Journal of the Royal Society Interface. 2021;18:20210063.

[29] Lee H, Kim Y, Kim E, Lee S. Risk Assessment of Importation and Local Transmission of COVID-19 in South Korea: Statistical Modeling Approach. JMIR Public Health Surveill. 2021 Jun;7(6):e26784. Available from: https://publichealth.jmir.org/2021/6/e26784.

[30] Wang Z, Yao M, Meng C, Claramunt C. Risk Assessment of the Overseas Imported COVID-19 of Ocean-Going Ships Based on AIS and Infection Data. ISPRS International Journal of Geo-Information. 2020 May;9(6):351. Available from: http://dx.doi.org/10.3390/ijgi9060351.

[31] Zhang L, Yang H, Wang K, Zhan Y, Bian L. Measuring imported case risk of COVID-19 from inbound international flights — A case study on China. Journal of air transport management. 2020 October;89:101918. Available from: https://europepmc.org/articles/PMC7455240.

[32] Quilty BJ, Russell TW, Clifford S, Flasche S, Pickering S, Neil SJ, et al. Quarantine and testing strategies to reduce transmission risk from imported SARS-CoV-2 infections: a global modelling study. medRxiv. 2021;Available from: https://www.medrxiv.org/content/early/2021/06/14/2021.06.11.21258735.

[33] Russell TW, Golding N, Hellewell J, Abbott S, Wright L, Pearson CAB, et al. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. BMC Med 2020. 2020;18:332.

[34] Russell TW, Wu JT, Clifford S, Edmunds WJ, Kucharski AJ, Jit M. Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. Lancet Public Health. 2021;6:e12–20.

[35] International Civil Aviation Organization. Doc 10152: Manual on COVID19 cross-border risk management; 2021. Available from: https://www.icao.int/covid/cart/Documents/10152_manual_3rd_edition.en.pdf.