Using COVID-19 data to investigate the use of early warning signs to identify epidemic peaks and areas of concern

Joshua Looker[1*], Kat S Rock[2], Louise Dyson[2,3]

1 EPSRC & MRC Centre for Doctoral Training in Mathematics for Real-World Systems, University of Warwick, Coventry, United Kingdom.
2 The Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research, and Mathematics Institute, University of Warwick, Coventry, UK.
3 School of Life Sciences, University of Warwick, Coventry, UK

* Corresponding Author. Email: joshua.looker@warwick.ac.uk

Abstract

The SARS-CoV-2 (COVID-19) pandemic has had catastrophic effects on public health and economies. Around the world, many countries employed modelling efforts to help guide pharmaceutical and non-pharmaceutical measures designed to reduce the spread of the virus. Modelling efforts for future pandemics could use the theory of early warning signals (EWS), which aims to predict 'critical transitions' in complex dynamical systems. In infectious disease systems, such transitions correspond to (re-)emergence, peaks and troughs in infections which can be indirectly observed through the reported case data. There is increasing evidence that including EWS in modelling can help improve responses to upcoming increases or decreases in case reporting. Here, we present both theoretical and data-driven analyses of the suitability of EWS to predict critical transitions in reported case data. We derive analytical statistics for a variety of infectious disease models and show, through stochastic simulations of different modelling scenarios, the applicability of EWS in such contexts. Using the COVID-19 reported case dataset from the United Kingdom, we demonstrate the performance of a range of temporal and spatial statistics to anticipate transitions in the case data. Finally, we also investigate the applicability of using EWS analysis of hospitalisation data to anticipate transitions in the corresponding case data. Together, our findings indicate that EWS analysis could be a vital addition to future modelling analysis for real-world infection data.

Author summary

The application of early warning signals (EWS), which uses statistics to predict changes in real-world data, has increasingly shown promise in successfully anticipating disease (re-)emergence or elimination of infectious diseases. We extend the existing literature through a theoretical study of EWS indicators under a variety of modelling scenarios that could have occurred during the COVID-19 pandemic (and in future pandemics). We show that EWS calculated on hospitalisation data can also accurately anticipate changes in the cases of related regions whilst also showing that spatio-temporal trends in EWS can be used to identify regions of concern.

1 Introduction

The end of 2019 saw the beginning of the SARS-CoV-2 (COVID-19) pandemic in Wuhan, China. A highly transmissible pathogen with the possibility of causing long-lasting and severe complications, COVID-19 was quickly evaluated by the World Health Organization to be a public health emergency of international concern and by 11th March 2020, following case repoting in 114 countries, a pandemic [1]. Many countries, including the United Kingdom (UK), enacted non-pharmaceutical interventions (NPIs), alongside later mass vaccine rollouts, aimed at reducing transmission [2]. Across the globe, countries employed various NPI measures including the closure of schools and workplaces, social distancing, gathering size limits, isolation and quarantine routines, and travel restrictions [3–5]. In the UK, the government took similar intervention measures during the 2020–2021 period. These included national lockdowns (e.g. in November 2020), gathering size restrictions (April 2021) and closures of workplaces and schools (October 2020). Naturally, the timeliness of these interventions directly influenced their effectiveness in decreasing the case numbers. Furthermore, in order to balance pandemic impacts on public health and on the economy, the UK government also undertook programmes to encourage resocialisation and increase spending (such as 'Eat Out to Help Out' in August 2020) and removed previously imposed restrictions, thus likely increasing virus spread [6].

Many governments also sought the help of mathematical modellers to predict the dynamics of COVID-19 with and without control measures, in order to make better-informed decisions on intervention policies. Modellers used a mixture of deterministic and stochastic methods to model the contagion dynamics and related statistics, such as reported cases, hospitalisations, and the effective reproduction number. Perfect mechanistic models could in theory be used to predict the future re-emergence of COVID-19 waves, but are in practice limited by the effects of data quality and human responses (such as the inability to accurately predict the implementation of intervention measures and public response to them). Model-independent methods such as early warning signals (EWS) have shown promise in anticipating critical transitions (in epidemics corresponding to peaks, troughs and (re-)emergence in the reported case time series) irrespective of the data quality and statistical noise [7–9].

In epidemic systems, the change in the expected number of secondary infections from an infectious individual determines a critical transition in the system [10]. Within a fully susceptible population, this value is called the basic reproduction number (R_0) and determines the stability of the endemic equilibrium. When $R_0 < 1$, the disease-free equilibrium is stable and elimination will likely be achieved, but when $R_0 > 1$ the disease-free equilibrium is instead unstable (and a stable endemic equilibrium may be reached, provided there is sufficient replenishment of susceptible individuals). In reality, infection- or vaccination-conferred immunity and intervention measures reduce the susceptible pool available/transmission rate, and the effective reproduction number (R_t) takes this into account. When $R_t > 1$, we expect multiple secondary cases from every infection and the disease is likely to 'take-off', but with $R_t < 1$ we expect less than one secondary case from every infection and the disease will eventually die-out. The critical transition between increasing and decreasing infections (or vice versa) therefore occurs when $R_t = 1$ [7, 11–13].

The theory of critical slowing down (CSD) states that on the approach to these transitions, the system takes longer to return to its equilibrium after small external perturbations, causing predictable changes in the statistics of the time series (such as an increase in variance and the autocorrelation). These expected changes then provide advanced warning (early warning signals) of the upcoming transition, such that appropriate action can be taken to mitigate the resultant effects [14, 15]. CSD-based EWS have also been shown to predict non-stationary periods (when there is growth or

reduction in infections) that do not necessarily correspond to critical transitions —this is also of benefit in predicting further infection waves which are not generally caused by critical transitions [16–18].

Indeed CSD and other EWS analyses have been investigated in the context of the COVID-19 pandemic (primarily using country-level incidence data) with varying levels of success [16, 17, 19–21]. These studies investigated whether EWS could be used to predict both re-emergence and waning. The fast dynamics of the pandemic, however, caused difficulties in reliably predicting these changes and accurately estimating the long-term trend in the observed cases (required to calculate the stochastic perturbations) [17, 20, 21]. Furthermore, recent results in the literature suggest that EWS show spatio-temporal effects that could be more noticeable than only considering the temporal domain [19, 22]. In the context of the COVID-19 pandemic, spatial detrending techniques and spatially-fine-grained datasets have been amongst the more successful EWS studies in anticipating re-emergence and the impacts of NPIs [16, 19].

The UK COVID-19 data has a far finer level of granularity and more consistent reporting than most previous outbreaks. Publicly-available daily case data was published for 315 regions/cities/smaller areas called Lower Tier Local Authorities (LTLAs), with other data (including hospitalisations) also published for the seven National Health Service (NHS) regions. Thus, this infection dataset provides a unique opportunity to investigate whether changes in temporal- (such as the variance) and spatial- (such as the spatial skew) statistics reliably anticipate upcoming critical transitions in a real-world disease system.

In this paper, we investigate whether using EWS analysis alongside other modelling efforts could have impacted decisions made in the June 2020 to December 2021 period of the COVID-19 pandemic in the UK. We first consider, through simulation and theory, the ability of EWS to anticipate the emergence of new variants and the approach of epidemic peaks. Using the publicly-available UK Health Security Agency (UKHSA) reported case and hospitalisation data, we consider the ability of various temporal- and spatial-statistics to anticipate these critical transitions. Finally, we employ the "2-σ detection method" [7] to quantify how early these transitions could be anticipated. This allows us to explore whether this advanced warning would have been enough to have changed policy decisions made during the selected pandemic period.

2 Materials and methods

Here we summarise the methods used, beginning by deriving the expected behaviour of various EWS in infectious disease models on the approach to a critical transition, and detailing the simulation methods used (Section 2.1). We then describe the dataset and pre-processing required, along with a summary of the course of the UK pandemic and how EWS were applied to the data, including investigation of the timing of EWS (Section 2.2).

To calculate the EWS and analyse the results we used Python 3.11 with packages: Numpy (version 1.26.4), Matplotlib (version 3.84), GeoPandas (version 0.14.4), statsmodels (version 0.14.1), epyestim (version 0.1) and Pandas (version 2.2.1). A repository containing the data and code used to conduct this study can be found at `https://github.com/joshlooks/covidews`.

2.1 Early warning signals: the theory

This section outlines the mathematical theory behind EWS upon the approach of a critical transition. We explore possible mathematical models for the (re-)emergence and peaks of COVID-19 waves (due to the easing of restrictions or mutation of new variants)

and the effects of NPIs. Analytical EWS results are derived from the Fokker-Planck equations (FPEs) and corresponding stochastic differential equations (SDEs) for the stochastic deviations from the mean for two common disease models. We selected the susceptible-infectious-susceptible (SIS) model as a baseline to match with the existing literature. The susceptible-exposed-infectious-recovered model (SEIR) is also analysed as it is commonly used to model COVID-19 infection and, to the best of the authors' knowledge, has not yet been theoretically considered under an EWS framework although numerical analyses have been conducted [22, 23]. Finally, we use extensive Gillespie simulations to explore EWS results under a variety of situations (re-emergence, NPI introductions, change of social behaviours) present during the COVID-19 pandemic.

2.1.1 Analytical early warning signals

The theory of critical slowing down expects signals in the statistics of fluctuations from the long-term trend in the infectious time series on approach to critical transitions (such as peaks and troughs). To develop an understanding of the possible behaviour of such statistics, we derive the SDEs and corresponding FPEs for the stochastic deviations from the long-run averages of the infectious compartment in two commonly used disease models: the SIS and SEIR models. Finally, we solve for the moments of the FPEs to analyse their expected behaviour on approach to a critical transition.

We begin with a simple SIS model of disease (re)-emergence due to the introduction of a new variant (Fig. S1, with full derivations in Section S1) where we considered the effective contact rate (β) between susceptible (S) and infectious (I) individuals as the control parameter. This allowed for similar analysis to be considered for the introduction or easing of NPIs which also impact β. Assuming a constant population size of $N = S + I$, the corresponding system is then one-dimensional with the ordinary differential equation (ODE) for I as:

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta(N - I)I/N - \gamma I. \tag{1}$$

where γ is the recovery rate of infectious individuals. We take the approach in Schnoerr *et al.* [24] in using the system size expansion and linear noise approximation [12, 24, 25] to derive the FPE for the probability distribution ($\Pi(\boldsymbol{\zeta}, t)$) of the stochastic fluctuations in the system:

$$\frac{\partial \Pi(\boldsymbol{\zeta}, t)}{\partial t} = -\sum_i \frac{\partial}{\partial \zeta_i} \sum_j J_{i,j}\zeta_j \Pi + \frac{1}{2}\sum_i \sum_j B_{ij} \frac{\partial^2}{\partial \zeta_i \partial \zeta_j}\Pi + O(N^{-1/2}), \tag{2}$$

where J is the Jacobean of the system (with respect to the average proportion of individuals in each compartment), the subscripts i, j represent each compartment and transition in the model, respectively, ζ_i is the stochastic fluctuations from the average proportion of reactants in compartment i and B is the noise-covariance matrix of the system.

Returning to the SIS system, the FPE and corresponding SDE can then be derived (Eq. (S4)). Finally, we can obtain the differential equation for the analytical variance ($\langle \zeta^2 \rangle$) as in Southall *et al.* [26]:

$$\begin{aligned} \frac{\partial \langle \zeta^2 \rangle_t}{\partial t} &= \int_{-\infty}^{\infty} \zeta^2 \frac{\partial \Pi}{\partial t} d\zeta, \\ &= 2(\beta - \gamma - 2\beta\phi)\langle \zeta^2 \rangle_t + \beta(1-\phi)\phi + \gamma\phi, \end{aligned} \tag{3}$$

where $\phi = I/N$. Previous numerical results for this equation show an increasing variance on approach to critical transitions (as expected from CSD theory) [12, 26, 27].

We wish to understand EWS in the SEIR model with demography (Fig. 1), as similar systems have been used successfully to model COVID-19 [28, 29]. We present a fuller analysis of this system as it has not previously been studied using the LNA in the literature.

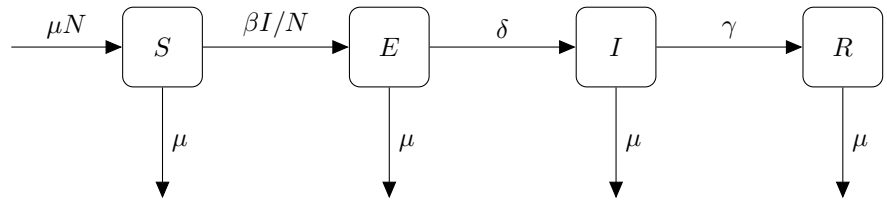

Fig 1. SEIR compartmental diagram showing the infection, recovery, birth/death and symptom onset processes. β is the effective contact rate between susceptible and infectious individuals, γ the rate of recovery of infectious individuals and the expected latent period is $1/\delta$. The birth and death rates (μ) are set equal such that the population size is constant.

The corresponding model equations are:

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= \mu N - \beta SI/N - \mu S, \\ \frac{\mathrm{d}E}{\mathrm{d}t} &= \beta SI/N - (\delta + \mu)E, \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \delta E - (\mu + \gamma)I, \end{aligned} \tag{4}$$

where $1/\delta$ is the expected latent period and μ is the natural death/birth rate. We note that by equating the birth and death rates, the average population size is constant such that $S + E + I + R = N$ and the ODE system is three-dimensional. The SEIR model also has two fixed points, a disease-free equilibrium ($S^* = N$) and an endemic equilibrium ($S^* = N/R_0 = N\frac{(\gamma+\mu)(\delta+\mu)}{\beta\delta}, E^* = N\frac{\mu(\mu+\gamma)}{\beta\delta}(R_0 - 1), I^* = N\frac{\mu}{\beta}(R_0 - 1)$). It is well known that the endemic equilibrium is stable for $R_0 > 1$ (and unstable for $R_0 < 1$), while the disease-free equilibrium is unstable for $R_0 > 1$ (and stable for $R_0 < 1$) [10].

Next, we use the linear noise approximation to express the number of susceptible, exposed and infectious individuals (S, E, I) in terms of the expected proportion of the population who are susceptible, exposed and infectious (ϕ_1, ϕ_2, ϕ_3, respectively) and the deviations from these means $(\zeta_1, \zeta_2, \zeta_3)$ as:

$$\begin{aligned} S &= N\phi_1(t) + N^{1/2}\zeta_1, \\ E &= N\phi_2(t) + N^{1/2}\zeta_2, \\ I &= N\phi_3(t) + N^{1/2}\zeta_3, \end{aligned} \tag{5}$$

where again ϕ_i, and ζ_i are the expected proportion and stochastic deviations. To first order (in N), Eq. (4) is then given as:

$$\begin{aligned} \phi_1 &= \mu - \beta\phi_1\phi_3 - \mu\phi_1, \\ \phi_2 &= \beta\phi_1\phi_3 - (\delta + \mu)\phi_2, \\ \phi_3 &= \delta\phi_2 - (\gamma + \mu)\phi_3. \end{aligned} \tag{6}$$

Finally, we need to find the Jacobean and noise-covariance matrices to express the Fokker-Planck equation of the system $(\zeta_1, \zeta_2, \zeta_3)$. The Jacobean is:

$$J = \begin{bmatrix} -\beta\phi_3 - \mu & 0 & -\beta\phi_1 \\ \beta\phi_3 & -(\delta+\mu) & \beta\phi_1 \\ 0 & \delta & -(\gamma+\mu) \end{bmatrix}, \tag{7}$$

and the noise-covariance matrix is:

$$B = \begin{bmatrix} \beta\phi_1\phi_3 + \mu\phi_1 + \mu & -\beta\phi_1\phi_3 & 0 \\ -\beta\phi_1\phi_3 & \beta\phi_1\phi_3 + (\delta+\mu)\phi_2 & -\delta\phi_2 \\ 0 & -\delta\phi_2 & \delta\phi_2 + (\gamma+\mu)\phi_3 \end{bmatrix}. \tag{8}$$

To calculate the variance (the main EWS used in the literature [7, 8, 22, 30]), we take the approach used in Southall *et al.* [26] by numerically solving the Lyapunov equation for the evolution of the covariance matrix $\Theta_{ij} = \langle\zeta_i\zeta_j\rangle - \langle\zeta_i\rangle\langle\zeta_j\rangle$:

$$\frac{\partial\Theta}{\partial t} = J\Theta + \Theta J^T + B \tag{9}$$

where the variance of the infectious fluctuations is given by Θ_{33}. We note that this (linear noise) assumption then also leads to the number of individuals in each compartment ($X = (S, E, I)$) following a multi-variate normal distribution given by:

$$X \sim \mathbf{N}(N\phi, N\Theta). \tag{10}$$

We have successfully derived the distribution of stochastic fluctuations in the SEIR model. This allows us to next analyse expected behaviour of EWS on approach to critical transitions.

2.1.2 Gillespie simulations

We next investigate the behaviour of EWS for the SEIR model for four scenarios: with fixed effective contact rate; with an increasing effective contact rate (due to the relaxing of NPIs/change in behaviour/introduction of a new variant); with continuous decrease in effective contact rate (due to NPIs); and with a step-wise decrease in effective contact rate (due to an immediate NPI introduction) to analyse whether a critical transition can be predicted/bifurcation delays can be observed (Fig. 2).

The model parameters, shown in Table 1, are informed by previous SEIR-based modelling of the wild-type and Alpha variants of COVID-19, and biologically relevant distributions (such as the latency period) [28, 29, 31–34], and differ in the form of $\beta(t)$ (which we currently consider to have a different time-dependent form depending on the modelling scenario). The basic reproductive number is assumed to be $R_0 = 4$ for the Alpha-variant, and $R_0 = 3.1$ for the wild-type variant, when setting the base effective contact rate (with the corresponding base effective contact rates similar in value to those used in [28, 29]). Note that these simulations are not designed to perfectly capture the spread of the two variants considered within the UK but to test the identifiability of EWS for a wild-type-like and an Alpha-like contagion undergoing SEIR dynamics.

We use the Gillespie method [35] to run 10,000 simulations of each of the four modelling scenarios, for the two different COVID-19-like variants, totalling eight sets of simulations. Event and transition rates for the four scenarios are given in Table 2. Each simulation has a population size of $N = 1,000,000$, an initial infectious population of 10 individuals, with the remaining 999,990 individuals being susceptible and a maximum simulation time of 200 days.

We analyse the simulation results in a similar way to the COVID-19 data (results can be seen in section Section 3.1), which will be explained in Sections 2.2.2 and 2.2.3. The main difference is that when using the Gillespie simulations the prevalence time series is detrended by subtracting the time-dependent average of all realisations (of that

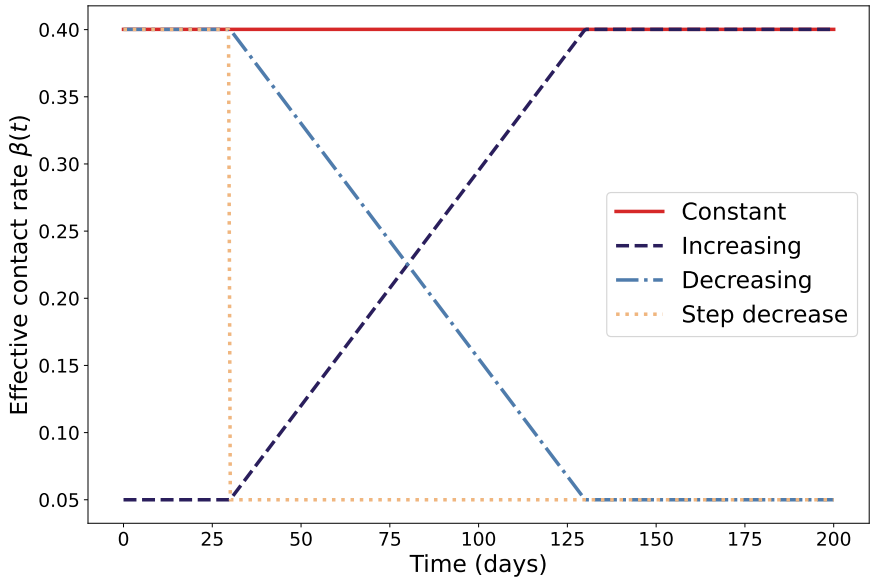

Fig 2. Effective contact rate over time for the different modelling scenarios considered: fixed, continuously increasing, continuously decreasing and with a step-decrease.

modelling scenario). Normalised statistics of the detrended time series are then analysed using the 2-σ algorithm (see section Section 2.2.3) to determine when the upcoming transition was detected. The estimated location of the critical transition is found by taking the average time $R_t = 1$ from the simulations in each scenario. EWS theory is concerned with the prevalence of infectious individuals, in a best-case scenario of "perfect information" and so we analyse this time series (as opposed to the incidence which is present in real-world case data).

Parameter	**Biological Interpretation**	**Value**
β_0	Effective contact rate (baseline)	0.4002, 0.3102 per day (Alpha-like, wild-type-like) [31, 36, 37]
μ	Natural mortality/birth rate	$1/(80 * 365)$ per day (both pathogens)
$1/\delta$	Latent period	5 days (both pathogens) [28, 29]
$1/\gamma$	Infectious period	10 days (both pathogens) [28, 29]
β_{low}	Effective contact rate (lower due to NPIs or similar)	0.05 per day (both pathogens)
p	Change in effective contact rate (due to NPIS or similar)	$(\beta_0 - \beta_{low})/100$ per day (both pathogens)
τ	Day $\beta(t)$ starts to change (for scenario two and four)	30 days (both pathogens)

Table 1. Parameters and values for each set of simulations run. The parameters are chosen to give similar contagion dynamics to both the wild-type and Alpha variants of COVID-19 in populations of the same magnitude as the three LTLAs considered in the UKHSA reported case results (Maidstone, York and Camden).

Event	Transition	Rate
All simulations		
Birth in S	$T(S+1\|S,E,I,R)$	μN
Death in compartment X	$T(X-1\|S,E,I,R)$	μX
Symptom onset	$T(E-1,I+1\|S,E,I,R)$	δE
Recovery	$T(I-1,R+1\|S,E,I,R)$	γI
Constant β		
Infection	$T(S-1,E+1\|S,E,I,R)$	$\beta_0 SI/N$
Re-emergence/restrictions easing		
Infection	$T(S-1,E+1\|S,E,I,R)$	$\beta_2(t)SI/N = [\beta_{low} + p(t-\tau)H(t-\tau)]SI/N$
Continuous NPI introduction		
Infection	$T(S-1,E+1\|S,E,I,R)$	$\beta_3(t)SI/N = [\beta_0 - p(t-\tau)H(t-\tau)]SI/N$
Step-change NPI introduction		
Infection	$T(S-1,E+1\|S,E,I,R)$	$\beta_4(t)SI/N = [\beta_0(1-H(t-\tau)) + \beta_{low}H(t-\tau)]SI/N$

Table 2. Events and transition rates for the four sets of Gillespie simulations run. Here $T(Y|X)$ indicates a transition from a system state of X to a system state of Y. The first set has a constant effective contact rate. The second models the re-emergence of infection due to the easing of restrictions or the introduction of a new variant and has $\beta_2(t) = \beta_{low} + p(t-\tau)H(t-\tau)$. The third and fourth model the introduction of NPIs continuously ($\beta_3(t) = \beta_0 - p(t-\tau)H(t-\tau)$) and as a step-change ($\beta_4(t) = \beta_0(1-H(t-\tau)) + \beta_{low}H(t-\tau)$) respectively.

2.2 Early warning signals: the data and methods

In this section, we summarise the pre-processing conducted on the UKHSA dataset to improve the prediction power of the EWS considered and to standardise the dataset into an interpretable form. We also describe the exact EWS considered in the study and how they are numerically calculated.

2.2.1 UK COVID-19 data: incidence and hospitalisations

Although COVID-19 infection data is still being published today, we focus on the period from June 2020 to December 2021 as it contained multiple critical transitions that could possibly have been anticipated by EWS analysis. This data, along with the timing of key events is shown in Fig. 3. This period saw the introduction of the 'Eat Out to Help Out' campaign in August 2020 by the UK government to support the hospitality sector [6]. Coupled with the easing of social restrictions in June and July, these changes led to the start of a COVID-19 wave in September 2020 [6]. After the introduction of additional measures (including enforced work from home) in September and October, there was another period of non-stationarity in the reported case time series. Finally, the dominance of the Alpha-variant in the December 2020 to January 2021 time period (Fig. S4) led to a much larger COVID-19 wave in the 2020 Christmas and New Year period. We note that a vaccination campaign against COVID-19 in the UK began in the January–February period of 2021, with the majority of the UK receiving a first dose within the first few months. However, since less than 65% of the population received a second dose, we do not expect the vaccination to noticeably reduce the last transition that we consider (in July 2021). Therefore the vaccine was not an indicator of a critical transition [38, 39]. We also note that the data between June and December 2021 shows periods of non-stationarity likely caused by the emergence and dominance of the Delta variant.

Two types of COVID-19 tests were commonly used in the UK: polymerase chain reaction-only (PCR) tests, which test for the presence of viral ribonucleic acid, and lateral flow device (LFD) tests, which test for viral antigens. The potential usable

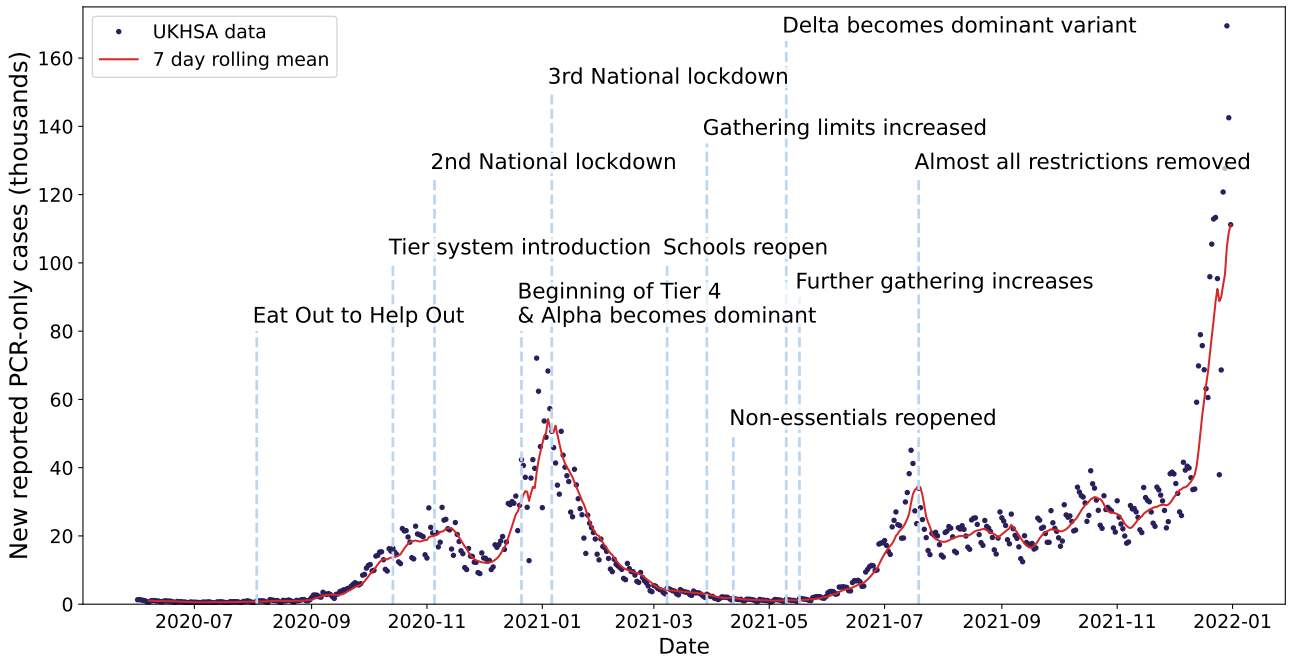

Fig 3. Timeline of the COVID-19 pandemic in the United Kingdom between June 2020 and December 2021. Daily reported polymerase chain reaction-only case data is displayed as dark dots, and the red line shows the rolling seven-day average. Key NPIs or changes in restrictions are noted: Eat Out to Help Out [6], Tier system introduction [40], 2nd National lockdown [40], Beginning of Tier 4 & Alpha becomes dominant [41], 3rd National lockdown [40], Schools reopen [42], Gathering limits increased [43], Non-essentials reopened [43], Delta becomes dominant variant [44], Further gathering increases [40], Almost all restrictions removed [40].

datasets are then: PCR-only, PCR-confirmed LFD, and LFD only (collected by UKHSA for their, now decommissioned, COVID-19 dashboard [45]). As during the time period considered almost all reported tests were PCR tests, we therefore chose to use the PCR-only dataset. There are also two potential routes for PCR test: 'Pillar 1' (healthcare-related testing) or 'Pillar 2' (mass community testing). The dataset used is primarily composed of Pillar 2 testing.

To help smooth the data, we aggregate the daily incidence of newly reported PCR-only cases to seven-day rolling sums. With a high daily variance in testing rates, this reduces the external noise in the resultant reported case time series. Using this summation also removes the 'weekend' effect seen (whereby testing and reporting rates were lower during the weekend [16, 17]). The daily reported case time series is analysed at the level of the 315 LTLAs (the smallest statistical area considered by the Office for National Statistics) within the UK, with the EWS calculated for each LTLA individually.

We focus on three LTLAs in particular for individual reporting: Maidstone, York and Camden. These LTLAs are spread across the UK with Maidstone in the South East, Camden in London and York in the North East. With the Alpha-variant hypothesized to have originated in East London or Kent [46], we also expect the critical transition times to differ between these LTLAs, particularly in late 2020. Further, the LTLAs selected have different population sizes so the impacts of population size on EWS, and the level of noise in the statistics, can also be examined (Maidstone with 173,132, Camden with 279,516 and York with 211,012 inhabitants according to the 2021 census [47]).

In addition to the PCR-only case reporting data, we also consider temporal statistics calculated on both the hospitalisation incidence and occupancy time series recorded by UKHSA [45]. These were collected for each of the seven NHS regions in the UK and are also smoothed by aggregating to seven-day rolling sums.

For each of the aggregated time series considered, the values are normalised by dividing by the population size of the corresponding region (LTLA or NHS region). This allows for the comparison of values across the different areas and for spatial/metapopulation detrending to be conducted (see Section 2.2.2).

2.2.2 Calculating early warning signals

This section outlines the numerical calculation of the statistics considered (both temporal and spatial) in the EWS analysis of COVID-19 reported case time series, including the window sizes used in the estimations. We also explain the detrending techniques used to separate stochastic fluctuations in the time series from the long-term trend.

The theory of EWS identifies expected trends in the statistics of stochastic fluctuations away from the long-term trend of the time series. We therefore take a metapopulation approach to detrend the smoothed time series (shown to be more accurate than purely temporal methods such as Gaussian or rolling-window detrending, [19, 27]). For the LTLA reported case time series, the average value for the corresponding NHS region is subtracted from each LTLA within that region. For the NHS regional hospitalisation incidence and hospitalised occupancy time series we instead subtract the national average.

After aggregating, normalising by the population size and detrending the time series, we calculate EWS on the resulting residuals. For each of the temporal statistics considered, we take an online approach: calculating the statistics on each day using only previously occurring data, mirroring how EWS would be applied in a real-world setting. The statistics are calculated using a back window of seven days, for those calculated on case incidence and hospital occupancy, or fourteen days, for those calculated on hospitalisation incidence. These window sizes were set using a sensitivity analysis (Figs. S5 and S6) to account for the differing noise levels. For temporal statistics, we considered both the variance and first difference in variance for each of the LTLAs. Other potential statistics were calculated but did not lead to robust EWS of upcoming transitions (Figs. S5 and S6) [17, 20]. The calculations are done using the Python package 'Pandas' which uses unbiased population estimates of statistics (for example the variance and skew). The variance ($V(t)$) and first difference ($FD(t)$) are calculated as:

$$V(t) = \frac{1}{\delta - 1} \sum_{i=t-\delta}^{t} (y_i - \bar{y_i})^2,$$
$$FD(t) = V(t) - V(t-1),$$

where $\delta \in \{7, 14\}$ is the window size used and y_i and $\bar{y}$ are the residual and mean at time point i, respectively.

The two types of spatial statistics considered are spatial variance and skew. These are calculated at the NHS regional level by considering the variance and skew across the residuals of all LTLAs within an NHS region. Explicitly, the spatial variance (SV) and

spatial skew (SS) are calculated as

$$SV(t) = \frac{1}{|Z|} \sum_{i \in Z} (y_t^i - \bar{z}_t)^2,$$

$$SS(t) = \frac{\sqrt{|Z|(|Z|-1)}}{|Z|-2} \frac{\frac{1}{|Z|} \sum_{i \in Z} (y_t^i - \bar{z}_t)^3}{(\frac{1}{|Z|} \sum_{i \in Z} (y_t^i - \bar{z}_t)^2)^{3/2}},$$

where y_t^i is the residual of LTLA i at time point t, $\bar{z}_t$ is the average of all LTLAs in the NHS region z at time point t, Z is the set of all LTLAs in NHS region z.

We also note that the normalised versions of these statistics are calculated to be used alongside the 2-σ method to detect outlying values (explained in Section 2.2.3). To normalise the statistics, we subtract the expanding mean and then divide by the expanding standard deviation (Fig. 4 bottom left). The rolling mean and rolling standard deviation were also tested but did not affect the results.

2.2.3 Detecting a critical transition

The 2-σ method from Drake and Griffen [7] proposes that a normalised EWS is significant if its instantaneous value exceeds two long-run standard deviations on either side of its long-run mean value. Following Southall *et al.* [48], we take the time-of-detection for the upcoming critical transition to be when at least three consecutive time points cross the threshold. The 2-σ threshold is analogous to a 95% confidence interval for the statistic value. This method has also previously been used in the context of COVID-19 [17].

Similarly to Dablander *et al.* [16], we need a method to determine where a critical transition ($R_t = 1$) occurs. This allows us to then analyse the effectiveness of different statistics in anticipating these transitions and also to refit the 2-σ method after each transition. Without refitting, time series statistics could be masked by previous transitions [16]. We make use of the method outlined in Huisman *et al.* [49] which builds on the method from Cori *et al.* [50]. The method smooths the reported case time series using local polynomial regression with a smoothing parameter that we set to 63 days at the LTLA level and 35 days at the NHS level (again set through a sensitivity analysis). This smoothed case time series is then used to estimate the number of infections and R_t using Bayesian inference techniques.

Although this may seem to combine online (*i.e.* real-time) learning and retrospective analysis, we note that in practice our method would only be used to predict the next upcoming critical transition. As stated, changes in the statistics due to previous transitions can reduce the ability to anticipate future transitions. Therefore, once the transition has definitely occurred, the expanding mean and standard deviation calculations would be cut and recalculated to anticipate the next transition (in the same manner that we conduct retrospectively). Further, similar methods of calculating R_t were also conducted in real-time in modelling efforts during the COVID-19 pandemic, and so could be used alongside our method in the event of a future pandemic [49–52]. Finally, the EWS analysis method conducted (in terms of predicting the next upcoming transition) does not rely on fitting R_t and the fitting is only necessary to retrospectively evaluate how including EWS analysis may have impacted modelling efforts and pandemic response.

The complete method from reported case time series to time-of-detection is summarised in Fig. 4.

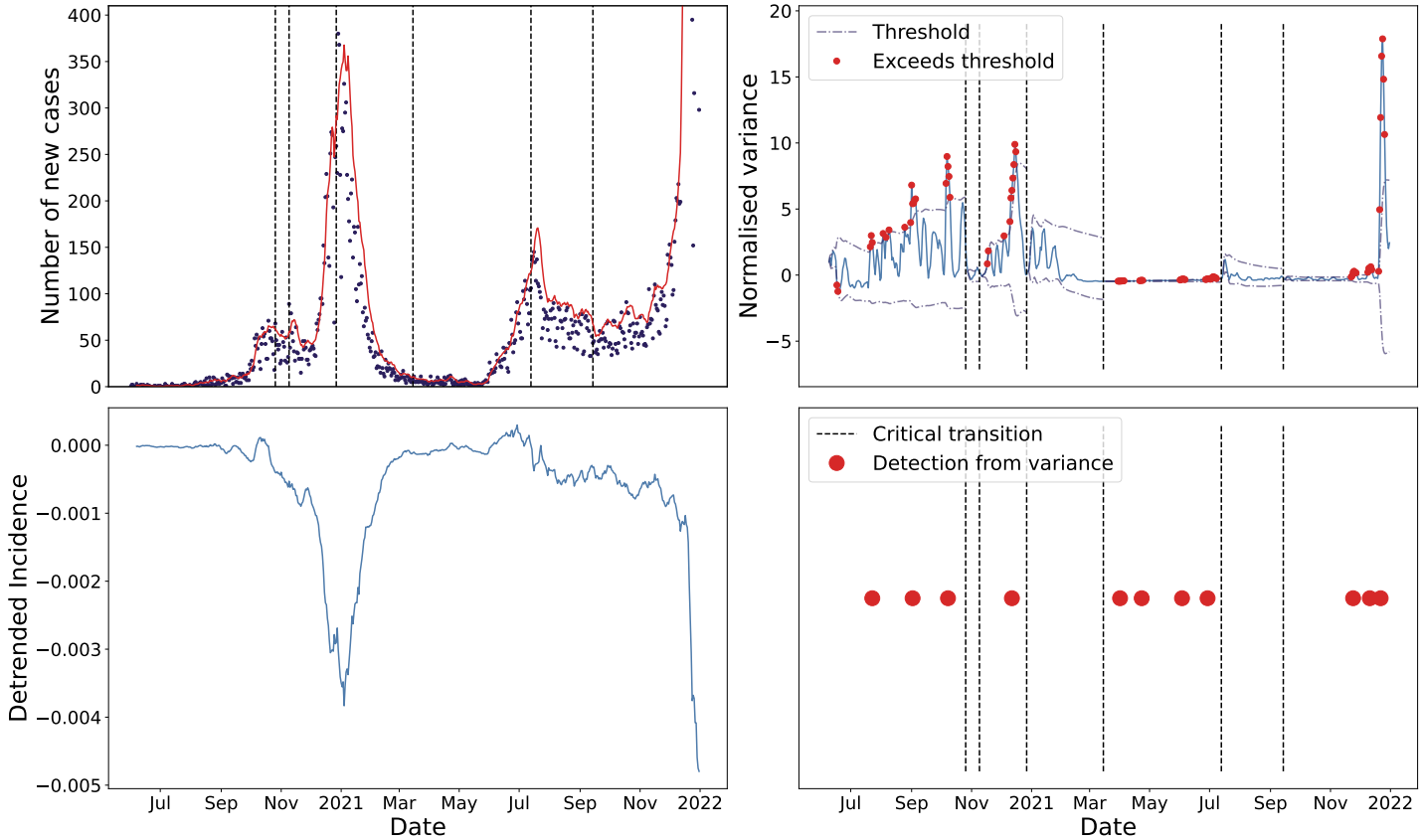

Fig 4. Full method of detecting EWS for the COVID-19 reported case time series in Camden. Top left, number of reported PCR-only COVID-19 cases (black dots) with the seven-day rolling sum overlayed. The vertical dotted lines correspond to the approximate date that $R_t = 1$. Bottom left, the detrended residuals in the reported case time series with the average incidence of the corresponding NHS region used for detrending. Top right, the normalised variance in residuals, with the dashed lines displaying the 2-σ detection threshold, the red dots signifying exceedances of this threshold and the dotted vertical lines again showing the critical transitions (whereon the threshold is reset). Bottom right, the resultant time-of-detection timeline showing the time-of-detection for upcoming critical transitions as the red dots (with each corresponding to three consecutive threshold exceedances in the subplot above).

3 Results

In this section, we outline the numerical results obtained from analysing the ability of EWS to reliably predict the re-emergence and die-out of subsequent COVID-19 waves between June 2020 and December 2021. We begin with theoretical results obtained from simulations designed to capture different behaviour from the pandemic. The data-driven analysis begins with the temporal statistics and time-of-detection results, for incidence-based EWS at the LTLA level before reporting the corresponding analysis for the spatial statistics at the NHS region level. Finally, the EWS results calculated from hospitalisation data are also reported.

3.1 Simulation results

The first section of results is from the sets of 10,000 Gillespie simulations run for the four different modelling scenarios considered, described in Section 2.1.2. All results were

made using the Alpha-variant-like parameters (with results for wild-type-like parameters showing similar qualitative trends and presented in Figs. S2b and S3). To compare the Gillespie simulation outputs (at consistent time intervals), we aggregate the proportion of individuals in SEIR to daily counts from the start of each day. We can then calculate the infectious proportion over time for each modelling scenario considered (single SEIR wave, increasing effective contact rate, gradually decreasing effective contact rate, and step-decrease in effective contact rate: Fig. S2). We analyse how the 'simulation' variance in the proportion infected (variance between each simulation's proportion infectious at each time point) and the time-of-detection distribution differ between scenarios (Fig. 5). As expected from CSD theory, the variance increases upon approach to the critical transitions in all scenarios tested (Fig. 5, third column). There is a second spike in variance after the critical transition for all scenarios bar the step-decrease in β possibly signalling the upcoming non-linear decrease in prevalence. Further, the variance in simulations, for both pathogens and all scenarios except increasing β, shows good concordance with the theoretical variance. Any differences are due to simulation stochasticity; plotting more simulations results in even closer concordance. For the 'increasing β' simulations we observe a difference in magnitude between the theoretical and simulation-based variance. This is because we specifically selected simulations that did not fade out, thus increasing the correlation between the simulation results and reducing the variance between simulations (compared to theory). Nonetheless, the qualitative shape of the theory and simulation-based variances are the same which suggests that an increase in variance is still a good EWS of (re-)emergence.

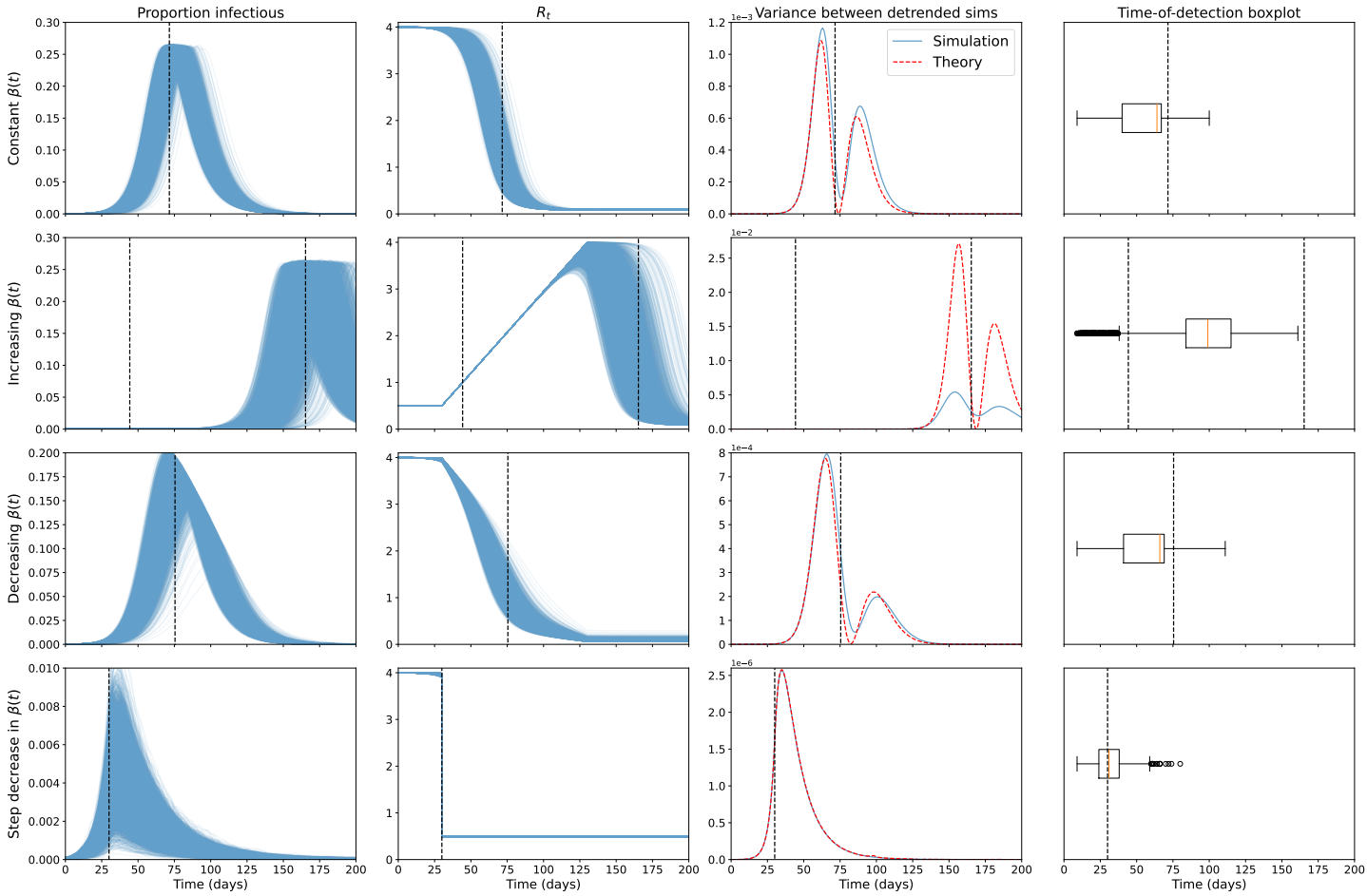

Fig 5. Columns show proportion infectious over time, effective reproduction number, variance between the mean-detrended simulations and time-of-detection distribution for the 10,000 simulations run for each of the four modelling scenarios (rows, constant $\beta(t)$, increasing $\beta(t)$, decreasing $\beta(t)$ and a step-decrease in $\beta(t)$) for the Alpha-like pathogen

The time-of-detection distribution (three consecutive 2-σ exceedances) in all simulations (for the Alpha-variant, Fig. 5 fourth column) shows the successful anticipation of the upcoming peak in the single SEIR wave set of simulations (Fig. 5 top row, most of the distribution is situated days before the transition). A similar trend is seen in the set of simulations that had a gradual decrease in β (third row), with almost all of the distribution mass occurring over five days before the critical transition. In contrast, simulations with an increasing effective contact rate (second row) have the majority of the distribution mass occurring after the re-emergence critical transition, possibly showing a characteristic bifurcation delay seen in a previous COVID-19 analysis [16]. It is unclear whether this is a delayed detection of the critical transition occurring or an early, successful anticipation of the upcoming peak. Finally, simulations with a step-decrease in β (fourth row) behave as expected with the 2-σ method showing a delayed detection of the instantaneous forcing through the critical transition. There is no gradual approach to the transition (β decreases instantaneously), so critical slowing down should not occur. Similar trends were also seen in the wild-type-like pathogen results, however the lower and later infection curves possibly led to the observed later and more spread-out time-of-detection distributions (Fig. S3).

3.2 LTLA regional early warning signals

In this section we outline the results using temporal EWS at the LTLA level to predict critical transitions in the COVID-19 cases time series between June 2020 and December 2021. We conduct a sensitivity analysis of the rolling window size used to calculate the normalised statistics and also compare the qualitative performance of the normalised signals in providing information of upcoming transitions (see Figs. S5 and S6), focusing on the detrended incidence residuals of the Maidstone, Camden and York LTLAs. We also analyse the spatio-temporal trends in early warning signals on approach to the peak in cases seen in many areas at the end of 2020.

We find that the skew, kurtosis and autocorrelation are noisy for all the LTLAs. Conversely, the index of dispersion and coefficient of variation have too few spikes to provide much predictive power. These observations agree with previous findings that the variance (and related statistics) generally perform best in anticipating transitions [9]. Although using a 14-day window makes the resultant time series less noisy, it also smooths out some of the possible spikes corresponding to a transition detection Figs. S5 and S6. We therefore take a rolling window size of seven days and focus on the variance and first difference in the variance in the rest of our analysis of temporal statistics at LTLA-level case data.

Using the previously outlined method, in Section 2.2.3, we estimated the R_t time series of all LTLAs, and now test the performance of the 2-σ method over the COVID-19 pandemic time period considered in terms of anticipating upcoming transitions (Fig. 6). There are clearly many 2-σ exceedances leading up to the Christmas 2020 peak in the three selected LTLAs as well as the peak in July 2021. There are very few and delayed exceedances possibly corresponding to the return to endemicity around March 2021 in all the LTLAs. The variance appears to have more consistently positive exceedances than the first difference. This is expected as CSD theory predicts increasing variance (and thus more positive exceedances), which would then lead to both positive and negative exceedances in the first difference in variance. The near-zero number of cases in the middle of 2021 also seems to reduce the efficacy of the method (there are far fewer exceedances).

Next, following the sensitivity analysis in Southall *et al.* [48], we define the detection time as the time point following at least three consecutive exceedances of the threshold, and calculate all of these occurrences between June 2020 and December 2021 for the three LTLAs of interest (Fig. 7).

Using the variance leads to more robust predictions of the upcoming peaks compared to the first difference in variance, especially for Maidstone, which had a much sharper peak than the other two LTLAs considered. This could be due to the more distinctly positive and consecutive trend in the variance.

Finally, we consider how well the variance anticipates the Christmas 2020 peak in all 315 LTLAs in England. For each LTLA, we consider the times-of-detection that occurred before the Christmas peaks. We discount any that occur before the previous critical transition, which mainly corresponds to that wave's take-off point, or a chosen cut-off point of the end of October as detection before then is likely signalling the take-off, not the peak—e.g. in Camden and York. We now analyse our novel spatio-temporal results for how early each LTLA detected the upcoming peak (Fig. 8a), and how soon after the first detection in an LTLA did the other LTLAs signal (Fig. 8).

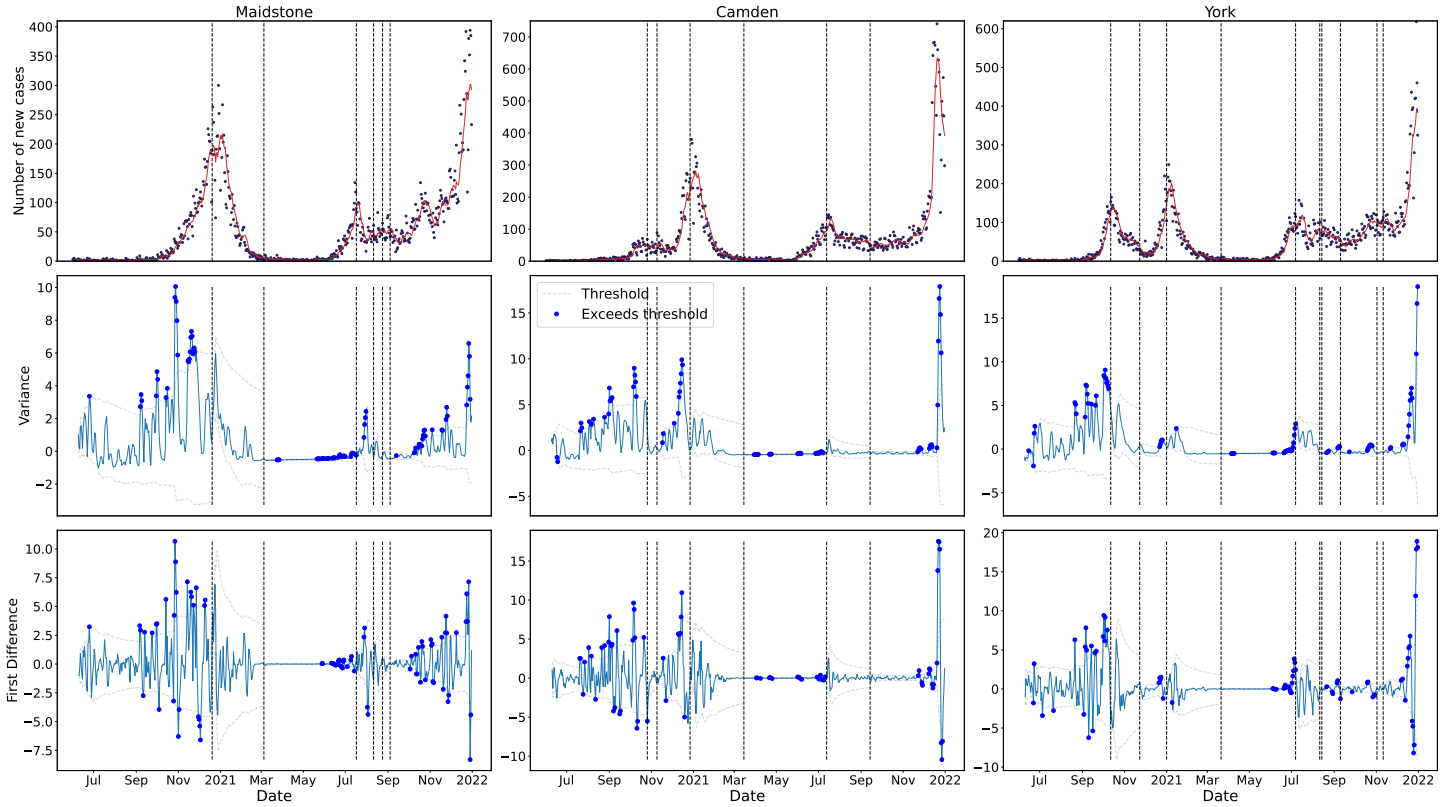

Fig 6. Time series of reported PCR-only COVID-19 cases and selected, normalised time series statistics for a sample of LTLAs between June 2020 and December 2021. The dotted lines show the 2-σ threshold, with dots corresponding to time points where the threshold is exceeded. The threshold is recalculated after the approximate times that $R_t = 1$, shown as the dotted vertical lines. Statistics are calculated using a rolling seven-day window.

The East and South East of England tend to anticipate the upcoming peak much earlier than the rest of England. Of the 44 LTLAs that did not signal before the peak, most saw 'flat' peaks in their reported case time series (or many fluctuations around $R_t = 1$ at Christmas). The last detecting LTLA (North Warwickshire) occurred 74 days after the first LTLAs that signalled (Milton Keynes, Uttlesford, Gosport, Wellingborough, Runnymede, Mid Sussex and Stevenage).

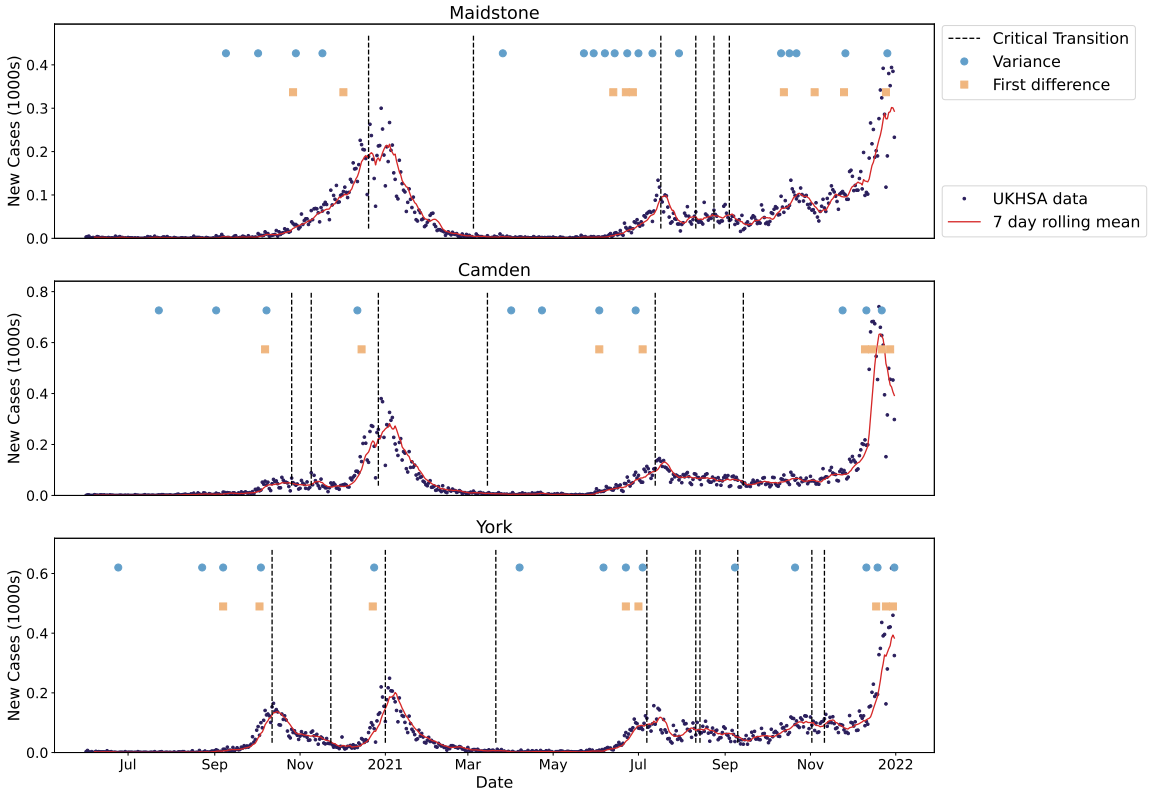

Fig 7. Illustration of the expected critical transitions for a sample of LTLAs between June 2020 and October 2021 alongside the times-of-detection for three consecutive threshold exceedances for the top performing EWS (vertical positioning of the signals does not matter). The vertical dotted lines show the expected dates when $R_t = 1$ and the circles and squares show the time-of-detection for variance and first difference in variance EWS. The incidence of new cases and seven-day rolling average are also overlayed.

3.3 NHS regional early warning signals

Having looked at the ability of EWS to predict the critical transitions at the LTLA level, we next investigate their usage for NHS regions. In this analysis, we test the ability of the spatial variance and spatial skew across the LTLAs in each NHS region to anticipate upcoming peaks and wanes.

Before conducting a corresponding spatio-temporal analysis of NHS regional statistics on the approach to the Christmas 2020 period peaks, we applied the 2-σ method to the spatial variance and skew (across each NHS region). The variance clearly exceeds the 2-σ threshold more reliably than the skew on approach to critical transitions, full results in Fig. S7. Interestingly, the spatial variance also seems to anticipate, or only slightly lag, the re-emergence wave in mid-2021 (which the temporal EWS at the LTLA level did not reliably identify).

In another set of novel analyses, we tested the ability of the temporal variance in both the new hospitalisations and hospital occupancy time series to anticipate the transitions in the NHS regional cases—to our knowledge, no analysis using hospitalisation data exists in the EWS literature. We apply the 2-σ method on the normalised variance to analyse trends in the threshold exceedances (Fig. 9, full results in Fig. S8).

Although hospitalisations and related time series lag infections [53], there are many exceedances upon approach to the critical transitions in the NHS-regional level reported

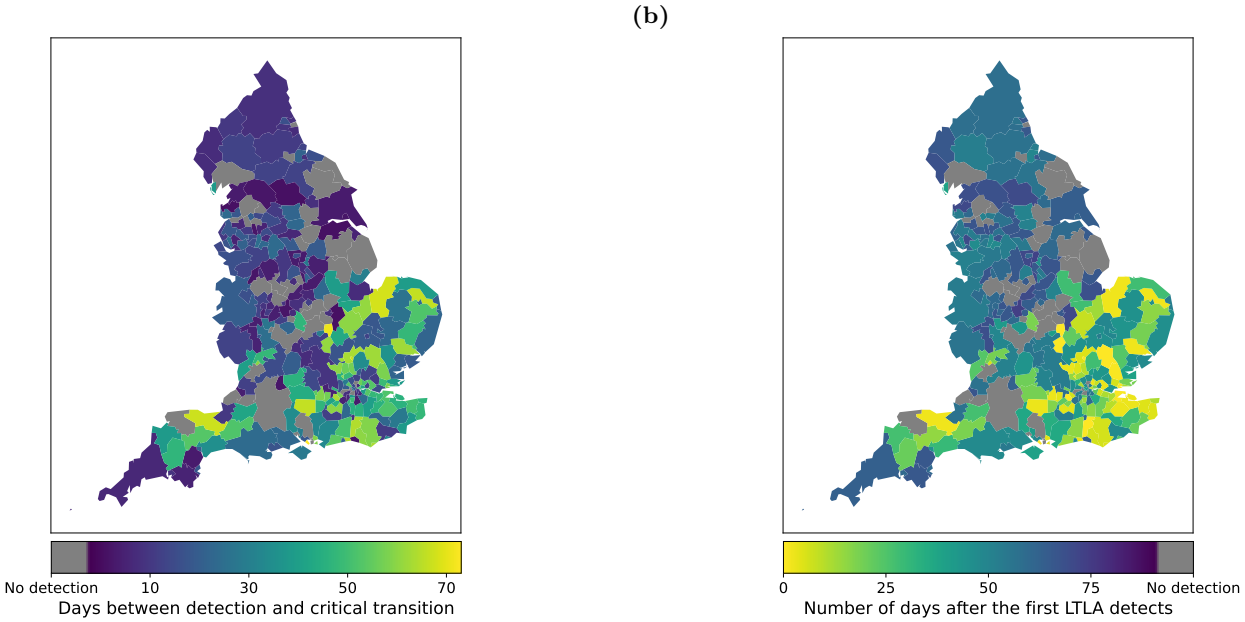

Fig 8. Detection date distributions (relative to the Christmas 2020 peaks) for all 315 LTLAs in England. **(a)** Number of days before the Christmas 2020 peak (for each LTLA) that the upcoming peak is detected (using the variance). **(b)** Number of days after the first day of detection that the upcoming peak is detected for each LTLA.

case time series. This suggests hospitalisation-based EWS still have predictive power for transitions in the corresponding reported cases time series and could be useful in future epidemic applications.

We conduct a similar timeline analysis into how each of the selected EWS perform in anticipating each of the critical transitions identified in the reported case time series for the seven NHS regions (Fig. 10).

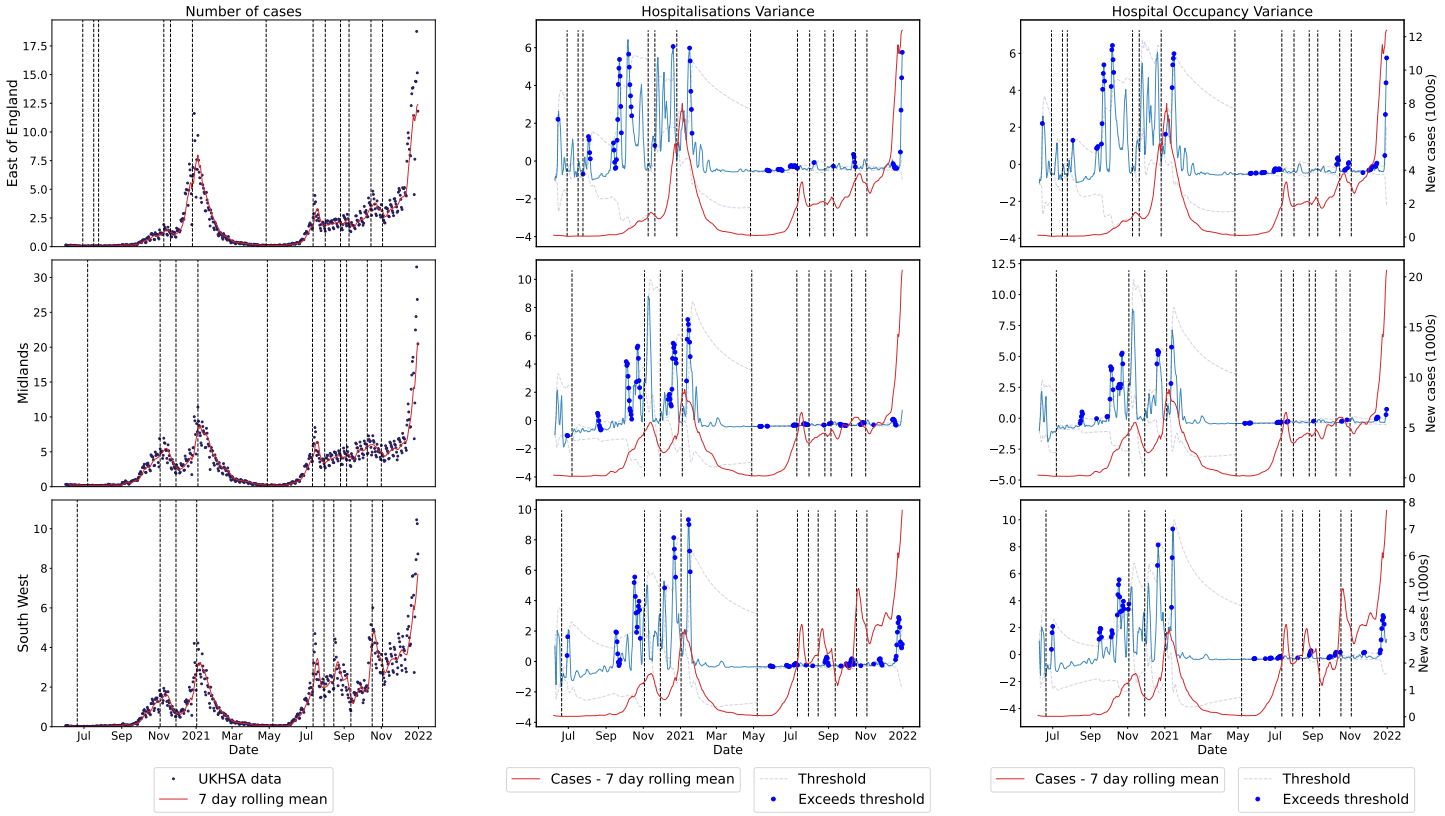

Fig 9. Reported PCR-only COVID-19 cases and selected, normalised temporal statistics calculated over the spatially detrended hospitalisation incidence and hospital occupancy data for the three selected NHS regions between June 2020 and December 2021. The dotted lines show the 2-σ threshold, with dots corresponding to time points where the threshold is exceeded. Incidence data is shown as the dark dots, with the red line being the rolling seven-day average.

Although the spatial variance signals less than the temporal variance in both the new hospitalisations and hospital occupancy, it seems to more reliably anticipate the upcoming peaks. Considering the peaks near the start of November 2020, January 2021 and July 2021 for all NHS regions, the sensitivity of the spatial variance was highest (at one), followed by the variance in incidence (0.857) and then variance in hospital occupancy (0.810). The spatial variance also seems to signal earlier than the temporal metrics, although the efficacy of each differs by region. Nonetheless, all EWS show multiple detections upon approach to almost all of the peaks and troughs in each region.

Finally, we consider how well the spatial variance anticipates the Christmas 2020 peak in each NHS region using the same method as the temporal variance with the LTLAs (Fig. 11).

The East of England is the earliest NHS region to signal with a time-of-detection 23 days before its Christmas critical transition. Although the Midlands signals last, it still anticipates the peak by four days. Regions where the corresponding LTLAs had a very early time-of-detection (Fig. 8b) are also quick to signal. However, the South West, which had some poor predictions at the LTLA level, also signals first.

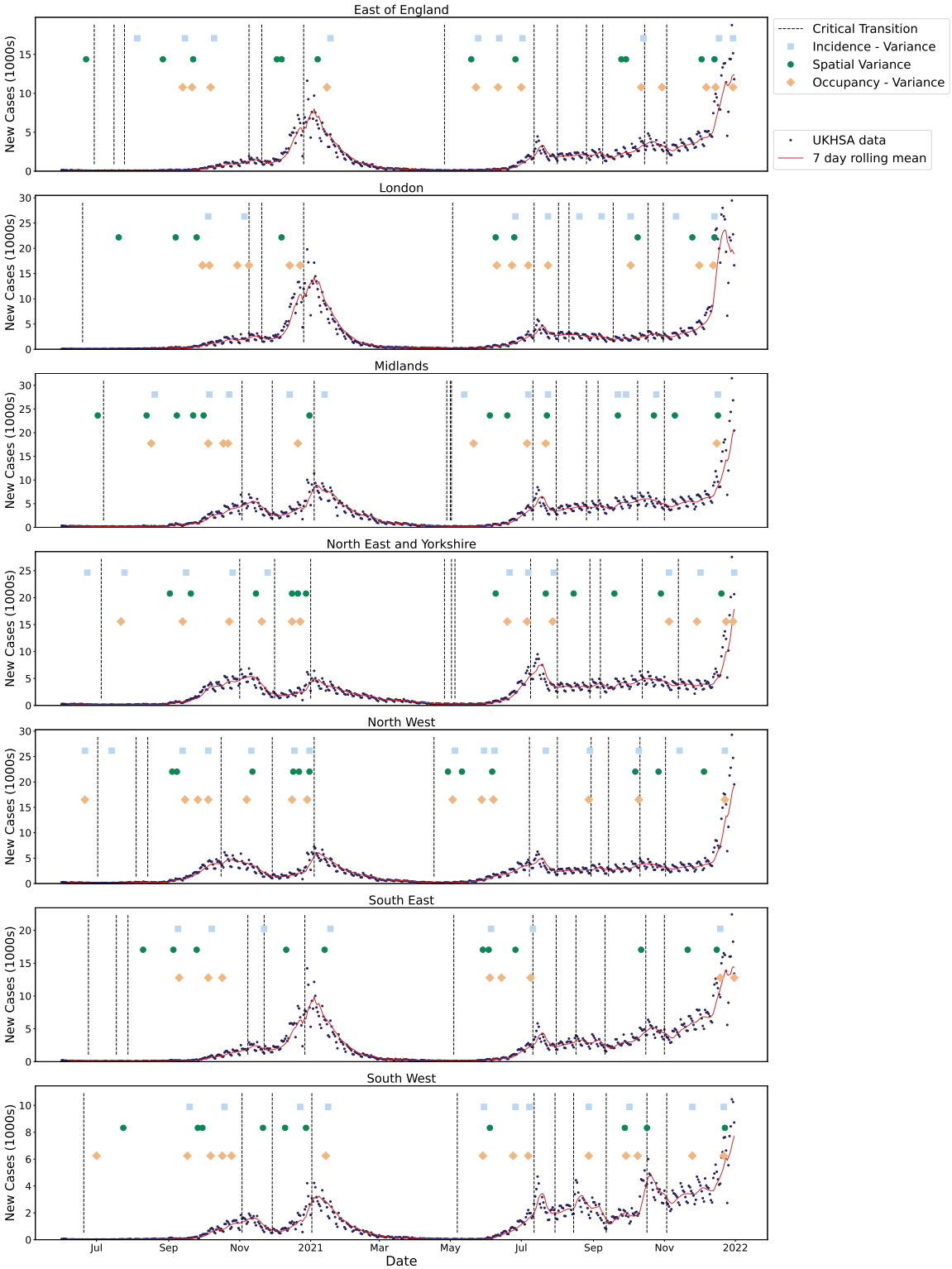

Fig 10. Illustration of the expected critical transitions for each of the seven NHS regions between June 2020 and December 2021 alongside the times-of-detection for three consecutive threshold exceedances. The vertical dotted lines show the expected dates when $R_t = 1$ and the circles, squares and diamonds indicate the time-of-detection for the spatial variance, and variance in hospitalisation-based data. The incidence of new cases and seven-day rolling average are also overlayed.

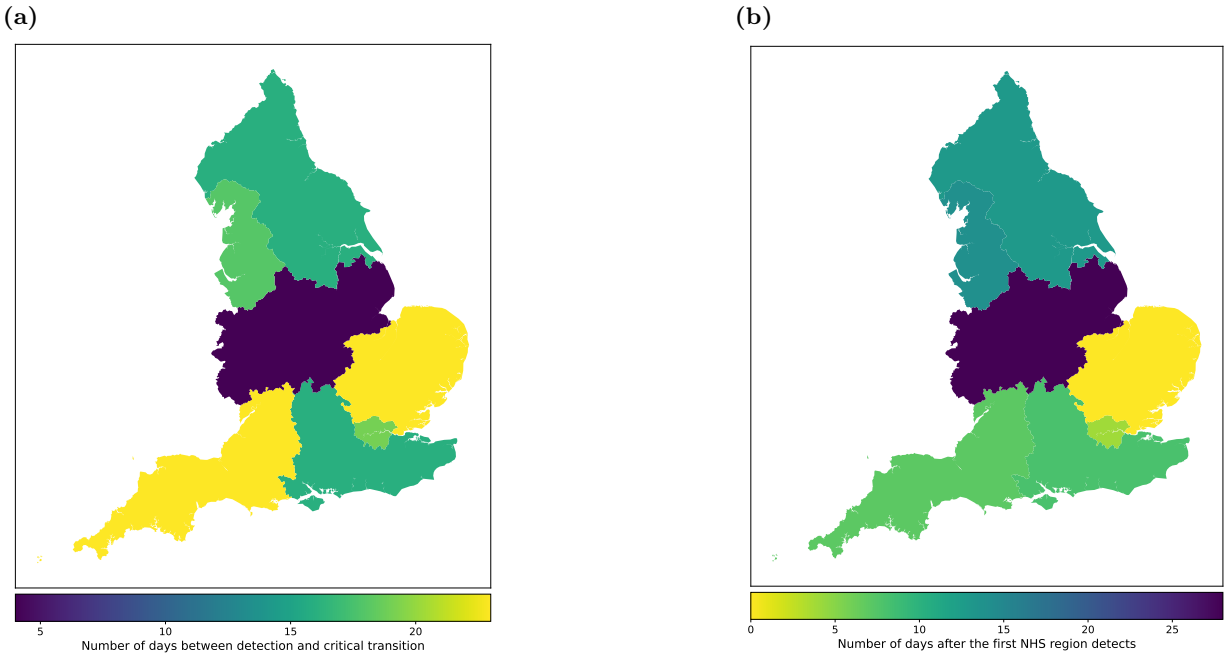

Fig 11. Detection date distributions (relative to the Christmas 2020 peaks) for all seven NHS regions in England. **(a)** Number of days before the Christmas 2020 peak (for each NHS region) that the upcoming peak is detected (using the variance). **(b)** Number of days after the first day of detection that the upcoming peak is detected for each NHS region.

4 Discussion

This paper analyses the ability of early warning signals to predict non-stationary periods of the reported cases time series in England for the June 2020 to December 2021 period of the COVID-19 pandemic. Our analysis considers both temporal- and spatial-statistics of LTLA and NHS regional case data. In addition, we extend the literature on EWS under SEIR-like contagion dynamics through a theoretical and simulation-based analysis. Finally, we also investigate using EWS analysis of hospitalisation time series to identify the critical transitions in the reported case data.

Our Gillespie simulation results (Figs. S3 and 5) provide validation that EWS could be used to anticipate critical transitions in real-world data of diseases with a latency period. At the time of writing, this mirrors the only previous EWS analyses of the SEIR model in the literature (Brett *et al.* [23] and Dablander *et al.* [16]). Brett *et al.* [23] established that SEIR-based disease models of increasing complexity show signs of CSD and Dablander *et al.* [16] considered how EWS of re-emergence may be subject to bifurcation delays, depending on how $\beta(t)$ was increased [54]. Nonetheless, our results are novel in using $\beta(t)$ as the bifurcation parameter in modelling scenarios of a constant, decreasing and step-decreasing transmission rate. To the best of the authors' knowledge, we are also the first to compare the analytical time-dependent variance to numerical results for the SEIR model.

We robustly anticipated transitions under the baseline SEIR and decreasing β scenarios. The second peak in variance seen agrees with previous results showing that CSD can be observed before non-linear periods (not just critical transitions) [16–18]. This suggests that EWS has the ability to anticipate a return to an endemic equilibrium or waning. Future work could investigate this, though it would likely require a method to separate the two peaks in variance/reduce the effects of the bifurcation delay. Furthermore, the simulations modelling a step-decrease in β show a successful (albeit delayed) detection of the critical transition. This suggests that EWS could also analyse, alongside modelling efforts, whether an implemented intervention measure has successfully decreased the transmission rate.

Conversely, only 15.7% of the (re-)emergence simulations signalled before the emergence transition. Further, considering the proportion infectious (Fig. S2a), these are unlikely to be true detections and are more likely caused by the non-linearity and stochasticity of decreasing infections [16]. Most of the time-of-detection mass occurred between the two critical transitions (re-emergence and the peak, Figs. S3 and 5 fourth columns), which could be caused by a delayed detection of the emergence or a successful anticipation of the peak. This lack of interpretability may hinder EWS analysis in real-world modelling. Although EWS may anticipate/detect a critical transition, they provide no information about the time or type of the upcoming transition or indeed if the detection was delayed and the transition has instead already occurred. In practice, if EWS analysis is used alongside other modelling efforts or data sources, these limitations are likely to be removed by the additional context provided. Finally, Brett *et al.* [23] used a different bifurcation parameter (a decrease in the vaccination rate) and successfully anticipated the re-emergence transition. Future work should investigate whether this discrepancy is due to the difference in the bifurcation parameter or in the time scales modelled.

The successful detection of the Christmas peak in over 86% of the LTLAs clearly demonstrates the usefulness of including early warning signals, especially trends in the temporal variance, in future pandemic modelling analyses (Fig. 8). There was widespread uncertainty in predicting when the peak of the Christmas 2020 wave was going to occur, and whether intervention measures were successful [55], during which our proposed EWS analysis might have helped guide policy decisions on the duration and location of NPIs. However, Fig. 7 indicates that EWS are less robust in

anticipating a return to endemicity, again highlighting that EWS are best used alongside other modelling efforts. Still, there are plenty of detection signals weeks before the increases in cases in Maidstone and Camden in July 2021. Using the 2-σ method at the time may have led to advanced warning/detection of the upcoming Delta-variant infection wave. At the NHS regional level (Fig. 10), where the larger population sizes likely lead to less stochasticity in results, these trends were more clearly visible. Seemingly no region anticipated the return to endemicity in March 2021, but most successfully indicated the non-stationary/increasing periods in reported cases in October-December 2020 and July 2021.

Our spatial analysis also highlights the importance of considering the spatial component of early warning signals. The spatial variance performed well as a signal at the NHS regional level, anticipating the majority of the waves in the reported case time series. The spatio-temporal maps (Figs. 8 and 11) also illustrate that considering the spread of detections across a country or region may provide extra information to modellers. If certain sub-regions have varied detection times/statuses, this may be a signal that the region as a whole may soon experience a non-stationary period. Future analysis should also consider the potential for spatial analysis of EWS, especially in the context of metapopulation disease models [19, 27]. The effects of spatial detrending in regions with differing sub-populations (such as a large rural/urban divide, or Manchester struggling with NPI efficacy compared to the rest of the North West) should also be considered.

The spatio-temporal analysis also highlights the possible real-world limitations of early warning signal analysis. The South East predicted the Christmas peak earlier and in more LTLAs than other areas. It is unclear whether this is a simple time delay from Alpha spreading throughout the country or other factors known to influence the predictive power of EWS. These factors (for each LTLA/region) include the population size [8], the sharpness of the peak in cases [17, 20], testing delays and rates [56], and how successfully/quickly interventions change the transmission rate [20]. Each of these present a limitation to successful EWS analysis in future epidemics, especially for regions that may have less granularity in infection data than the UK (such that even if not all LTLAs were signalling, EWS of possible infection waves could still be seen in other LTLAs).

Our analysis of hospitalisation-based signals indicates their potential usefulness for future epidemics (Fig. 10). They performed similarly to the spatial-statistics in prediction power, but also have their own inherent limitations. Hospitals may face social or operating pressures to reduce occupancy rates, potentially causing bias. Further, hospitalisation-based statistics may be skewed by operating capacity in hospitals. Such external factors can limit EWS efficacy.

The data-driven analysis nonetheless highlights the benefits of including EWS in future pandemic modelling efforts. However, to our knowledge, theoretical EWS analysis in the literature mainly considers SIS and SIR dynamics. In practice, more complex models (beyond just SIS or SIR dynamics) are usually used [10], and future work should further investigate analytical EWS for these systems. Furthermore, although EWS analysis of COVID-19 incidence data has overwhelmingly shown signs of CSD on approach to critical transitions [16, 17, 20, 21], previous theoretical work on the SIS model has shown that temporal-statistics for incidence decrease on approach to elimination [26]. Future work should investigate whether this holds for more complex disease models too. Recent theoretical analysis of interacting ecological systems indicates only certain sub-populations may show evidence of CSD upon approach to critical transitions [57]. Future studies should investigate whether this is also a property of metapopulation-style disease models. Composite metrics have been proposed as a method of combining different statistics into a more robust EWS indicator, and

combining statistics from different time series (such as hospitalisations and cases) into a composite should be considered further [7, 48]. Finally, the recent work into EWS in COVID-19 case data indicates how dependent robust predictions, especially of re-emergence, are on the disease time scales (how quickly a wave blows up/dies out). Further study is required into whether this would affect the applicability of EWS to future outbreaks of pathogens of concern.

In this paper, we have analysed the ability of EWS to predict peaks and troughs in England's COVID-19 reported case time series using case incidence and hospitalisation data. We also performed a theoretical analysis of EWS under various disease models and scenarios. These advances show the potential efficacy of EWS in aiding modellers and decision-makers in future epidemic scenarios to better inform the suitability of implementing intervention measures and predicting infection waves.

Author contributions

Joshua Looker: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualisation, Writing - Original Draft, Writing - Review & Editing.
Kat Rock: Conceptualisation, Supervision, Visualisation, Writing - Review & Editing.
Louise Dyson: Conceptualisation, Supervision, Visualisation, Writing - Review & Editing.

Financial disclosure

JL was supported by the Engineering and Physical Sciences Research Council through the MathSys CDT [grant number EP/S022244/1]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Data availability

All data utilised in this study are publicly available, with relevant references and data repositories provided.

Code availability

The code repository for the study is available at:
`https://github.com/joshlooks/COVIDews`
Archived code available at:
`https://doi.org/10.5281/zenodo.14852091`

Competing interests

All authors declare that they have no competing interests.

References

1. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. *Acta Bio Medica : Atenei Parmensis* **91**(1):157–160 (2020). doi:10.23750/abm.v91i1.9397.

2. UKHSA. Chapter 8: non-pharmaceutical interventions (2023). URL `https://www.gov.uk/government/publications/technical-report-on-the-covid-19-pandemic-in-the-uk/chapter-8-non-pharmaceutical-interventions`. [Online] (Accessed: 2024-08-12).

3. Lison A, Banholzer N, Sharma M, Mindermann S, Unwin HJT, *et al.* Effectiveness assessment of non-pharmaceutical interventions: lessons learned from the COVID-19 pandemic. *The Lancet. Public Health* **8**(4):e311–e317 (2023). doi:10.1016/S2468-2667(23)00046-4.

4. Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Physics Reports* **913**:1–52 (2021). doi:10.1016/j.physrep.2021.02.001.

5. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, *et al.* A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour* **5**(4):529–538 (2021). doi:10.1038/s41562-021-01079-8. Publisher: Nature Publishing Group.

6. Fetzer T. Subsidising the spread of COVID-19: Evidence from the UK'S Eat-Out-to-Help-Out Scheme*. *The Economic Journal* **132**(643):1200–1217 (2022). doi:10.1093/ej/ueab074.

7. Drake JM, Griffen BD. Early warning signals of extinction in deteriorating environments. *Nature* **467**(7314):456–459 (2010). doi:10.1038/nature09389. Publisher: Nature Publishing Group.

8. Southall E, Brett TS, Tildesley MJ, Dyson L. Early warning signals of infectious disease transitions: a review. *Journal of The Royal Society Interface* **18**(182):20210555 (2021). doi:10.1098/rsif.2021.0555.

9. Delecroix C, Nes EHv, Leemput IAvd, Rotbarth R, Scheffer M, *et al.* The potential of resilience indicators to anticipate infectious disease outbreaks, a systematic review and guide. *PLOS Global Public Health* **3**(10):e0002253 (2023). doi:10.1371/journal.pgph.0002253. Publisher: Public Library of Science.

10. Keeling MJ, Rohani P. *Modeling infectious diseases in humans and animals.* Princeton Univ. Press, Princeton, NJ (2008).

11. Kuehn C. A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics. *Physica D: Nonlinear Phenomena* **240**(12):1020–1035 (2011). doi:10.1016/j.physd.2011.02.012.

12. O'Regan SM, Drake JM. Theory of early warning signals of disease emergenceand leading indicators of elimination. *Theoretical Ecology* **6**(3):333–357 (2013). doi:10.1007/s12080-013-0185-5.

13. Kéfi S, Dakos V, Scheffer M, Van Nes EH, Rietkerk M. Early warning signals also precede non-catastrophic transitions. *Oikos* **122**(5):641–648 (2013). doi:10.1111/j.1600-0706.2012.20838.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0706.2012.20838.x.

14. Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, *et al.* Anticipating Critical Transitions. *Science* **338**(6105):344–348 (2012). doi:10.1126/science.1225244.

15. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, *et al.* Early-warning signals for critical transitions. *Nature* **461**(7260):53–59 (2009). doi:10.1038/nature08227. Publisher: Nature Publishing Group.

16. Dablander F, Heesterbeek H, Borsboom D, Drake JM. Overlapping timescales obscure early warning signals of the second COVID-19 wave. *Proceedings of the Royal Society B: Biological Sciences* **289**(1968):20211809 (2022). doi:10.1098/rspb.2021.1809.

17. O'Brien DA, Clements CF. Early warning signal reliability varies with COVID-19 waves. *Biology Letters* **17**(12):20210487 (2021). doi:10.1098/rsbl.2021.0487.

18. Dutta PS, Sharma Y, Abbott KC. Robustness of early warning signals for catastrophic and non-catastrophic transitions. *Oikos* **127**(9):1251–1263 (2018). doi:10.1111/oik.05172. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/oik.05172.

19. Liu R, Zhong J, Hong R, Chen E, Aihara K, *et al.* Predicting local COVID-19 outbreaks and infectious disease epidemics based on landscape network entropy. *Science Bulletin* **66**(22):2265–2270 (2021). doi:10.1016/j.scib.2021.03.022.

20. Kaur T, Sarkar S, Chowdhury S, Sinha SK, Jolly MK, *et al.* Anticipating the Novel Coronavirus Disease (COVID-19) Pandemic. *Frontiers in Public Health* **8**:569669 (2020). doi:10.3389/fpubh.2020.569669.

21. Proverbio D, Kemp F, Magni S, Gonçalves J. Performance of early warning signals for disease re-emergence: A case study on COVID-19 data. *PLOS Computational Biology* **18**(3):e1009958 (2022). doi:10.1371/journal.pcbi.1009958.

22. Ghadami A, O'Dea EB, Drake JM, Rohani P, Epureanu BI. Anticipating epidemic transitions in metapopulations with multivariate spectral similarity. *Nonlinear Dynamics* **111**(18):17605–17615 (2023). doi:10.1007/s11071-023-08727-w.

23. Brett T, Ajelli M, Liu QH, Krauland MG, Grefenstette JJ, *et al.* Detecting critical slowing down in high-dimensional epidemiological systems. *PLoS Computational Biology* **16**(3):e1007679 (2020). doi:10.1371/journal.pcbi.1007679.

24. Schnoerr D, Sanguinetti G, Grima R. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *Journal of Physics A: Mathematical and Theoretical* **50**(9):093001 (2017). doi:10.1088/1751-8121/aa54d9. Publisher: IOP Publishing.

25. Kampen NGv. *Stochastic processes in physics and chemistry.* North-Holland personal library. Elsevier, Amsterdam ; Boston, 3rd ed edition (2007). OCLC: ocm81453662.

26. Southall E, Tildesley MJ, Dyson L. Prospects for detecting early warning signals in discrete event sequence data: Application to epidemiological incidence data. *PLOS Computational Biology* **16**(9):e1007836 (2020). doi:10.1371/journal.pcbi.1007836.

27. Gama Dessavre A, Southall E, Tildesley MJ, Dyson L. The problem of detrending when analysing potential indicators of disease elimination. *Journal of Theoretical Biology* **481**:183–193 (2019). doi:10.1016/j.jtbi.2019.04.011.

28. Fošnarič M, Kamenšek T, Žganec Gros J, Žibert J. Extended compartmental model for modeling COVID-19 epidemic in Slovenia. *Scientific Reports* **12**(1):16916 (2022). doi:10.1038/s41598-022-21612-7. Publisher: Nature Publishing Group.

29. Aronna MS, Guglielmi R, Moschen LM. A model for COVID-19 with isolation, quarantine and testing as control measures. *Epidemics* **34**:100437 (2021). doi:10.1016/j.epidem.2021.100437.

30. Brett TS, Drake JM, Rohani P. Anticipating the emergence of infectious diseases. *Journal of The Royal Society Interface* **14**(132):20170115 (2017). doi:10.1098/rsif.2017.0115.

31. Billah MA, Miah MM, Khan MN. Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. *PLOS ONE* **15**(11):e0242128 (2020). doi:10.1371/journal.pone.0242128. Publisher: Public Library of Science.

32. Wu Y, Kang L, Guo Z, Liu J, Liu M, *et al.* Incubation Period of COVID-19 Caused by Unique SARS-CoV-2 Strains: A Systematic Review and Meta-analysis. *JAMA Network Open* **5**(8):e2228008 (2022). doi:10.1001/jamanetworkopen.2022.28008.

33. Manica M, Litvinova M, De Bellis A, Guzzetta G, Mancuso P, *et al.* Estimation of the incubation period and generation time of SARS-CoV-2 Alpha and Delta variants from contact tracing data. *Epidemiology and Infection* **151**:e5 (2022). doi:10.1017/S0950268822001947.

34. Galmiche S, Cortier T, Charmet T, Schaeffer L, Chény O, *et al.* SARS-CoV-2 incubation period across variants of concern, individual factors, and circumstances of infection in France: a case series analysis from the ComCor study. *The Lancet Microbe* **4**(6):e409–e417 (2023). doi:10.1016/S2666-5247(23)00005-8. Publisher: Elsevier.

35. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**(25):2340–2361 (1977). doi:10.1021/j100540a008.

36. Liu Y, Rocklöv J. The effective reproductive number of the Omicron variant of SARS-CoV-2 is several times relative to Delta. *Journal of Travel Medicine* **29**(3):taac037 (2022). doi:10.1093/jtm/taac037.

37. Xia F, Yang X, Cheke RA, Xiao Y. Quantifying competitive advantages of mutant strains in a population involving importation and mass vaccination rollout. *Infectious Disease Modelling* **6**:988–996 (2021). doi:10.1016/j.idm.2021.08.001.

38. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, *et al.* Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. *New England Journal of Medicine* **386**(16):1532–1546 (2022). doi:10.1056/NEJMoa2119451. Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa2119451.

39. Gram MA, Emborg HD, Schelde AB, Friis NU, Nielsen KF, *et al.* Vaccine effectiveness against SARS-CoV-2 infection or COVID-19 hospitalization with the Alpha, Delta, or Omicron SARS-CoV-2 variant: A nationwide Danish cohort study. *PLOS Medicine* **19**(9):e1003992 (2022). doi:10.1371/journal.pmed.1003992. Publisher: Public Library of Science.

40. Laydon DJ, Mishra S, Hinsley WR, Samartsidis P, Flaxman S, *et al.* Modelling the impact of the tier system on SARS-CoV-2 transmission in the UK between the first and second national lockdowns. *BMJ Open* **11**(4):e050346 (2021). doi:10.1136/bmjopen-2021-050346. Publisher: British Medical Journal Publishing Group Section: Epidemiology.

41. Kirby T. New variant of SARS-CoV-2 in UK causes surge of COVID-19. *The Lancet Respiratory Medicine* **9**(2):e20–e21 (2021). doi:10.1016/S2213-2600(21)00005-9. Publisher: Elsevier.

42. Gurdasani D, Alwan NA, Greenhalgh T, Hyde Z, Johnson L, *et al.* School reopening without robust COVID-19 mitigation risks accelerating the pandemic. *The Lancet* **397**(10280):1177–1178 (2021). doi:10.1016/S0140-6736(21)00622-X. Publisher: Elsevier.

43. Solomon-Moore E, Lambert J, Grey E, Gillison F, Townsend N, *et al.* Life in lockdown: a longitudinal study investigating the impact of the UK COVID-19 lockdown measures on lifestyle behaviours and mental health. *BMC Public Health* **22**(1):1495 (2022). doi:10.1186/s12889-022-13888-1.

44. Emma B. Hodcroft. CoVariants (2021). URL `https://covariants.org/`. [Online] (Accessed: 2025-02-03).

45. UKHSA. UKHSA data dashboard (2024). URL `https://ukhsa-dashboard.data.gov.uk`. [Online] (Accessed: 2024-08-18).

46. Hill V, Du Plessis L, Peacock TP, Aggarwal D, Colquhoun R, *et al.* The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evolution* **8**(2):veac080 (2022). doi:10.1093/ve/veac080.

47. ONS. Census - Office for National Statistics (2022). URL `https://www.ons.gov.uk/census`. [Online] (Accessed: 2024-08-18).

48. Southall E, Tildesley MJ, Dyson L. How early can an upcoming critical transition be detected? (2022). doi:10.1101/2022.05.27.22275693. URL `https://www.medrxiv.org/content/10.1101/2022.05.27.22275693v1`. [Online] (Accessed: 2024-08-12). Pages: 2022.05.27.22275693.

49. Huisman JS, Scire J, Angst DC, Li J, Neher RA, *et al.* Estimation and worldwide monitoring of the effective reproductive number of SARS-CoV-2. *eLife* **11**:e71345 (2022). doi:10.7554/eLife.71345.

50. Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology* **178**(9):1505–1512 (2013). doi:10.1093/aje/kwt133.

51. Abbott S, Hellewell J, Thompson R, Sherratt K, Gibbs H, *et al.* Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts [version 2; peer review: 1 approved, 1 approved with reservations]. *Wellcome Open Research* **5**(112) (2020). doi:10.12688/wellcomeopenres.16006.2.

52. Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, *et al.* Practical considerations for measuring the effective reproductive number, Rt. *PLOS Computational Biology* **16**(12):e1008409 (2020). doi:10.1371/journal.pcbi.1008409. Publisher: Public Library of Science.

53. Eales O, Haw D, Wang H, Atchison C, Ashby D, *et al.* Dynamics of SARS-CoV-2 infection hospitalisation and infection fatality ratios over 23 months in England. *PLOS Biology* **21**(5):e3002118 (2023). doi:10.1371/journal.pbio.3002118.

54. Dibble CJ, O'Dea EB, Park AW, Drake JM. Waiting time to infectious disease emergence. *Journal of The Royal Society Interface* **13**(123):20160540 (2016). doi:10.1098/rsif.2016.0540. Publisher: Royal Society.

55. Scientific Advisory Group for Emergencies. SPI-M-O: Consensus statement on COVID-19, 6 January 2021 (2021). URL `https://www.gov.uk/government/publications/spi-m-o-consensus-statement-on-covid-19-6-january-2021`. [Online] (Accessed: 2024-09-02).

56. Brett TS, O'Dea EB, Marty E, Miller PB, Park AW, *et al.* Anticipating epidemic transitions with imperfect data. *PLoS Computational Biology* **14**(6):e1006204 (2018). doi:10.1371/journal.pcbi.1006204.

57. Dablander F, Pichler A, Cika A, Bacilieri A. Anticipating critical transitions in psychological systems using early warning signals: Theoretical and practical considerations. *Psychological Methods* **28**(4):765–790 (2023). doi:10.1037/met0000450. Place: US Publisher: American Psychological Association.