Predicting and forecasting the impact of local resurgence and outbreaks of COVID-19: Use of SEIR-D quantitative epidemiological modelling for healthcare demand and capacity

Eduard Campillo-Funollet[1*], James Van Yperen[2], Phil Allman[3], Michael Bell[4], Warren Beresford[5], Jacqueline Clay[6], Graham Evans[7], Matthew Dorey[6], Kate Gilchrist[4], Anjum Memon[8], Gurprit Pannu[9], Ryan Walkley[6], Mark Watson[10], Anotida Madzvamuse[2*]

[1] School of Life Sciences, Centre for Genome Damage and Stability, University of Sussex, Brighton, BN1 9QH, UK
[2] School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Brighton, BN1 9QH, UK
[3] NHS Sussex Commissioners, Wicker House, Worthing, BN11 1DJ, UK
[4] Public Health Intelligence and Adult Social Care, Brighton and Hove City Council, 2nd Floor, Hove Town Hall, Norton Road, Hove, BN3 3BQ, UK
[5] Planning and Intelligence, Brighton and Hove, Sussex Commissioners, East Sussex, UK
[6] Public Health and Social Research Unit, West Sussex County Council, 1st Floor, The Grange, Tower Street, Chichester, West Sussex, PO19 1RG, UK
[7] Public Health Intelligence, East Sussex County Council, St Anne's Crescent, Lewes, BN7 1UE, UK
[8] Brighton and Sussex Medical School, Brighton, BN1 9QH, UK
[9] Sussex Health and Care Partnership, First Floor, Millview Hospital, Nevill Avenue, Hove, East Sussex, BN3 7HY, UK
[10] Sussex Health and Care Partnership, 36-38 Friars Walk, Lewes, BN7 2PB, UK

*Corresponding author: E.Campillo-Funollet@sussex.ac.uk
*Corresponding author: A.Madzvamuse@sussex.ac.uk

Summary

Background The world is at the cusp of experiencing local/regional hot-spots and spikes of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes COVID-19 disease. We aimed to formulate an applicable epidemiological model to accurately predict and forecast the impact of local resurgence and outbreaks to guide the local healthcare demand and capacity, policy making, and public health decisions.

Methods The model utilised the aggregated daily COVID-19 situation reports (including counts of daily admissions, discharges, and occupancy) from the local NHS hospitals and Covid-19 related weekly deaths in hospitals and other settings in Sussex (population 1·7M), Southeast England. These datasets corresponded to the first wave of COVID-19 infections from 24 March-15 June 2020. The counts of death registrations and regional population estimates were obtained from the Office of National Statistics. A novel epidemiological predictive and forecasting model was then derived based on the local/regional surveillance data. Through a rigorous inverse parameter inference approach, the model parameters were estimated by fitting the model to the data in an optimal sense and then subsequently validated to make predictions subject to 95% confidence.

Findings The inferred parameters were physically reasonable and matched up to the widely used parameter values derived from the national datasets. Unlike other predictive models, which are restricted to a couple of days, our model can predict local hospital admissions, discharges (including deaths) and occupancy for the next 10, 20, and 30 days at the local level.

Interpretation We have demonstrated that by using local/regional data, our predictive and forecasting model can be utilised to guide the local healthcare demand and capacity, policy making, and public health decisions to mitigate the impact of COVID-19 on the local population. Understanding how future COVID-19 spikes/waves could possibly affect the regional populations empowers us to ensure the timely commissioning and organisation of services. Primary care and community services can be guided by the projected number of infectious and recovered patients and hospital admissions/discharges to project discharge pathways to bedded and community settings, thus allowing services to understand their likely load in future spikes/waves. The flexibility of timings in the model, in combination with other early warning systems, produces a timeframe for these services to prepare and isolate capacity for likely and potential demand within regional hospitals. The model also allows local authorities to plan potential mortuary capacity and understand the burden on crematoria and burial services. The model algorithms have been integrated into a web-based multi-institutional toolkit, which can be used by NHS hospitals, local authorities, and public health departments in other regions of the UK and elsewhere. The parameters, which are locally informed, form the basis of predicting and forecasting exercises accounting for different scenarios and impact of COVID-19 transmission.

Funding This study was supported by the Higher Education Innovation Fund through the University of Sussex (ECF, JVY, AMa). This work was partly supported by the Global Challenges Research Fund through the Engineering and Physical Sciences Research Council grant number EP/T00410X/1: UK-Africa Postgraduate Advanced Study Institute in Mathematical Sciences (AMa, ECF). ECF is supported by the Wellcome Trust grant number 204833/Z/16/Z.

Research in context

Evidence before this study

Since the beginning of the COVID-19 pandemic, healthcare managers and policy makers relied on epidemiological models based on national datasets to predict and mitigate the spread of the disease. The performance of these models has not always been validated against the available data, and they depend strongly on the values for the model parameters. Statistical models, e.g. those arising from time-series analysis, lack the temporal dynamics of the compartmentalised epidemiological model for the evolution of the disease and thus fail to capture the evolution far into the future with great accuracy. Compartmental models, on the other hand, capture the underlying dynamics of an infectious disease but typically use parameters estimated using datasets from other regions or countries, thus lacking the ability to capture local demographics and policy and therefore lack predicting local dynamics with accuracy.

Added value of this study

Although our compartmental model follows standard SEIR-D model structure, the inference algorithm described and applied in this report is novel, along with the prediction technique used to validate the model. We checked bioRxiv, medRxiv, and arXiv up to the end of August 2020 using the terms “mathematical inference”, “COVID-19”, and “SIR” and found that there is a substantial use of Bayesian approaches to fit parameters but none that use the combination of statistical approaches with compartmental models, hence the originality of our work. We designed a compartmentalised epidemiological model that captures the basic dynamics of the COVID-19 pandemic and revolves around the data that are available at the local/regional level. We estimated all the parameters in the model using the local surveillance data, and in consequence, our parameters reflect the characteristics of the local population. Furthermore, we validated the predictive power of the model by using only a subset of the available data to fit the parameters. To the best of our knowledge, this is the first study which combines statistical approaches with a

compartmental model and as such benefits greatly from the ability to predict and forecast much further into the future using the dynamical structure of the compartmental model with a relatively much higher accuracy than previously presented in the literature. This research sets the gold-standard benchmark by laying the framework for future adaptations to the model when more precise (and comprehensive) datasets are made available.

Implications of all the available evidence

The predictive power of our model outperforms previously available models for local forecasting of the impact of COVID-19. Using local models, rather than trying to use national models at a local scale, ensures that the model reflects the local demographics and provides reliable local-data-driven predictions to guide the local healthcare demand and capacity, policy making, and public health decisions to mitigate the impact of COVID-19 on the local population. Local authorities can use these results for the planning of local hospital demand as well as death management services by developing scenario-based analysis to which different values of the reproduction number R exiting a COVID-19 lockdown are assumed and results, such as maximum hospital occupancy, are compared to the first wave to establish a potential strain on resources. This can work as an early warning detection system to see what value of R that is currently followed, which in turn informs the relevant capacity and resources needed to mitigate the impact of COVID-19. The Web toolkit developed by us as a result of this study (https://alpha.halogen-health.org) demonstrates the predictive power of our model as well as its flexibility with the scenario-based analysis. Although our model is based on the data from Sussex, using similar variables/data from other regions in our model would derive respective COVID-19 model parameters, and thus enable similar scenario-based investigations to predict and forecast the impact of local resurgence to guide the local healthcare demand and capacity, policy making, and public health decisions.

Introduction

The world is at the mercy of the severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2), the virus causing COVID-19, whose roots originated from Wuhan, China, where it was identified in December 2019.[1] Since then, COVID-19 has swiftly and rapidly spread to nearly all countries in the world, becoming an ongoing global world pandemic that has required unprecedented international, national, and regional interventions to try and contain its spread.[1,2] Unlike the 1918-19 H1N1 pandemic, which is considered one of the greatest medical disasters of the 20th century,[1] the spread of COVID-19 has unfolded live on multimedia platforms with real-time updates, statistics and with remarkable reporting accuracy,[3] and yet reliable, accurate, and data-validated epidemiological modelling with forecasting and prediction capabilities remains largely out of reach.[1,4–7] Given the lack of pharmaceutical interventions, such as vaccination and antiviral drugs, epidemiological modelling has been thrust to the forefront of world organisations' and governments' responses, rapid decision-making, and public health interventions and policy.[1,4,8] Until these pharmaceutical interventions become widely available, the only measures for infection prevention and control are self or group-isolation (quarantine), testing and contact tracing, physical distancing, decontamination, use of personal protective equipment, and hygiene measures. A lot of these unprecedented actions/decisions have resulted in complete lockdowns of countries, economies, and yet these decisions are based on qualitative/quantitative predictions/models using national datasets alien to the countries imposing the lockdowns on the basis of these models. A fair criticism of the underlying approach has been the lack of rigorous model validation and applicability given the datasets available at the time of the study, the lack of risk assessment associated with the decisions and their impact on the healthcare demand, capacity, and delivery and subsequently the lack of precision forecasting that is driven by data.[9] Unfortunately, early epidemiological models needed to make assumptions out of necessity about

parameters and disease progression. However, the predictions of these models were near impossible to validate.[1,4,6–8] At the forefront of these epidemiological models that have played a pivotal role in guiding national public health policy and healthcare responses that include the current social distancing, contact tracing, and quarantine measures, is the well documented Imperial College London model.[1] The societal and economic impact of the aforementioned decisions have hardly been quantified, only estimates in the range of trillions of dollars loss to the world economy are reported.[10,11] A few models dealing with decision-making within the COVID-19 crisis have been reported;[8,12] however, these lack the power of model prediction and forecasting based on appropriate variables and datasets.

In order to understand the temporal dynamics of COVID-19, a lot of modelling work has been undertaken, focusing primarily on national datasets from China, Italy, Spain, UK, and the USA.[1,4,6–8,13–15] Given the inhomogeneous nature of such datasets, accurate predictions and forecasting of the spread of COVID-19 have so far not been possible. Where such predictions were made, caveats accompanied these predictions simply because of the lack of rigorous mathematical and statistical validation of the models and the lack of robust data on which mathematical assumptions are based.[1,4,6–8,14,15] Forecasting requires ample historical information/datasets, which were lacking during the first wave of COVID-19. Current state-of-the-art forecasting models are based, on one hand, on time series analysis without an underlying dynamic epidemiological model.[6,16] On the other hand, where forecasting is based on epidemiological models,[6] these lack rigorous validation, sensitivity analysis, and analysis with respect to identifiability of parameters and therefore have limited forecasting power. An interesting approach is proposed in Bertozzi et al. (2020)[4] where three models were presented depending on the forecasting timescales; an exponential growth model, a self-exciting branching process, and the classical susceptible-infected-recovered (SIR) compartmental model. The exponential growth model is assumed valid at the early stages of the pandemic, the self-exciting branching process models the individual count data going into the development of the pandemic, and the SIR is a macroscopic mean-field model that describes the pandemic dynamics as it approaches the peak of the infection and disease. Another interesting and alternative approach is to build machine learning and artificial intelligence techniques on top of epidemiological models to allow for model predictions and forecasting.[6] This approach, so far, has been applied to national datasets from the USA but no regional modelling of this type has been undertaken.

The use of local datasets is critical for managing and mitigating COVID-19 secondary spikes/waves and re-infection within local communities. Already there is ample evidence that local forecasting models could help local/regional authorities to plan lockdowns, restrictions, opening of schools/universities, as well as planning for healthcare demand and capacity. For example, all the 50 states in the USA recently started to relax lockdown restrictions, however, several states are now either putting on hold their efforts to open fully or started to backtrack due to the resurgence in COVID-19 infections and the start of secondary waves. Here in the UK, cities such as Leicester, Bradford, and Oldham are in the midst of experiencing secondary COVID-19 waves and re-infection. In Australia, the city of Melbourne in the state of Victoria is in stage 4 lockdown while the remainder of the state is in stage 3 lockdown. During the first wave, Australia was hailed as a global success story in suppressing the spread of COVID-19 and even at the height of the initial outbreak, it only reported a little over 600 infections a day. A similar story is emerging in Spain with regions in Catalonia undergoing secondary lockdowns. The usefulness of national models, in all these countries, is not clear in terms of being able to predict and forecast the emergence of such spikes, waves or new incidences locally until they have already taken place, which is too late. We propose therefore an alternative quantitative predictive approach which gives local (and national) authorities the ability to predict and forecast COVID-19 scenarios based on their current

historical datasets to visualise future dynamic temporal trends of the infection/disease progression for healthcare planning purposes.

In this study, we want to demonstrate the usefulness and utility of a locally data-driven epidemiological model, based on recent datasets from the three adjoining regions in Sussex, Southeast England (i.e. Brighton and Hove City Council, East and West Sussex County Councils), to make predictions and forecast to guide local/regional decision-making and healthcare delivery. The approach is based on a modified SIR-type model, (Figure 1), that has been formulated to reflect the dynamics of the combined Sussex populations of approximately 1·7 million and the mathematical interpretation of the data available.

The aim of our study is to propose a systematic modelling approach that addresses healthcare demand and capacity at a local level – using the Sussex datasets – by conducting healthcare demand modelling that naturally leads to a standardised framework to quantify demand suppressed and generated as a result of COVID-19. This will facilitate long-term forecasting and predictions, allow for investigations into the impact of COVID-19 on healthcare provision and planning within the local area and to mitigate long-term changes in local hospital demand as a result of further COVID-19 secondary waves. We used the local datasets collected throughout the first wave, which included local daily hospital data and weekly deaths data. Our approach differs substantially from current state-of-the-art modelling-forecasting approaches where unknown parameters driving epidemiological models have been based on various assumptions which vary substantially from one model to the next as well as variations between the domain-expertise of the researchers involved in making those assumptions. Instead, we propose that we know nothing about the values of the model parameters and these are instead inferred through an inverse modelling approach by requiring the model to fit to local data in an optimal sense.[4,6,16] In this way, by fitting our SEIR-D model to data, we obtain the best values of the unknown model parameters (all the parameters shown in Figure 1), accurate to some degree of confidence.[17,18]

Methods

Data Collection

As part of the national COVID response, all the National Health Service (NHS) hospitals in England, treating COVID-19 patients, submitted a Daily Situation Report to NHS England (NHSE). The regional data for Sussex hospitals were then sent to the Sussex Clinical Commissioning Group (CCG) who aggregated the data and combined it with the death registrations (with COVID-19 as the underlying cause of death) from the Office for National Statistics (ONS). The subsets of the hospital datasets included daily admissions, discharges, and occupancy. The death dataset consisted of weekly COVID-19 related deaths in hospitals and community settings (e.g. nursing homes). These datasets corresponded to the first wave of infections from 24th March until 15th June 2020. For the regional population count, the ONS Mid-Year Estimates (MYE) for 2018 were used. By identifying these compartments where data were available, a mathematical model was generated with the objective of forecasting local hospital occupancy and mortuary requirements. Although death is an absolute count, the policy regarding what constituted a COVID-19 related death changed frequently throughout the lead up to and during the lockdown period. Similarly, testing was not optimal when the hospital data collection started and so admissions and occupancy counts were retroactively edited to incorporate newly tested patients, thus the balance of total patients between days may not match up. To mitigate this issue, we account for significant levels of noise in the observations.

Data-driven SEIR-D modelling

The temporal dynamics of the compartmentalised epidemiological model are depicted in Figure 1, following classical approaches for formulating SIR models.[2,20] The mathematical interpretation of

the schematic diagram in Figure 1 leads to a temporal epidemiological dynamical system modelled by a system of ordinary differential equations supported by non-negative initial conditions. The full model is summarised in equations (1) – (9) in the Supplementary Material (SM).

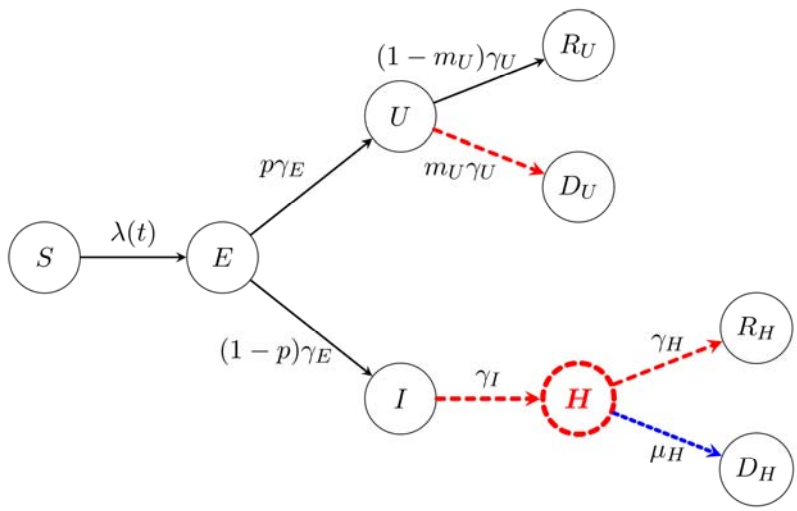

Figure 1: Schematic representation of the compartmental model. The susceptible population becomes infected through contacts with infectious individuals, and . Infected individuals incubate the disease first,[6] and are not infectious in this state, denoted by , and after an incubation time, they become infectious. The compartment accounts for individuals that are not detected in our data, i.e. they are not hospitalised. Many models have split the compartment into two separate compartments (see for example Blyuss et al. (2020)[2]), one to describe individuals who are asymptomatic and the other to describe individuals who have symptoms but do not require hospitalisation. However, this approach is constrained by the lack of reliable datasets, and therefore, models of this nature rely purely on the merits of the simulations with no forecasting capabilities. For such models, it is challenging to obtain reliable data on those who are asymptomatic, especially on the scale of multiple regions/counties. Individuals in the compartment are eventually hospitalised and move to compartment . We added the hospital compartment into the model as a transition compartment due to the data we have access to. There are two possible outcomes for COVID cases, recovery or death, denoted by or respectively, each subscripted with the severity of the infected, being undetected and being hospitalised. We note that there is also a difficulty gaining reliable data that considers those who are undetected and recovered. However, we have reliable datasets for those who die outside of hospital and thus, in the spirit of this model, are related to the undetected pathway. Coloured and dashed arrows or compartments indicate that data are available: admissions to hospital, red dashed arrow from to , discharges from hospital, red dashed arrow from to , daily counts of cases in hospital, red dashed compartment, and independent weekly data on deaths, both in hospital, red dashed arrow from to , and out of the hospital settings, red dashed arrow from to . A novel feature of our model and inference method is that even though information about is hard to come by, we can still obtain information by using the red dashed arrow between and . The same thing can be said for obtaining the information on by using the red dashed arrow between and . The parameters in the model regulate the rates from one compartment to the next and are described in Table 1. All parameters are inferred from the data using a minimisation process under constraints on the total population and the effective reproduction number.

Our model follows the general principles of SIR modelling approaches with one clear difference in that this model system is data-driven formulated where we have highlighted in dashed colour those compartments or pathways in Figure 1 where data is available within our local area. The physical justification of the SEIR-D model above is well-grounded in the modelling literature for COVID-19 and the general theory of epidemiology.[2,20]

Inferring model parameters given hospital datasets

From the schematic diagram shown in Figure 1, we are interested in finding the best or optimal set of eight model parameters: , , , , , , , and , such that the SEIR-D model best-fits the observed data. We estimate the parameters in the model in two steps. First, we exploit the

linear relationship, arising from the mathematical model, between mortality in hospitals and discharged patients, depicted by the blue double dashed line and the red dashed line between $H(t)$ and $R_H(t)$ in Figure 1, respectively, to fit the parameter $\eta = \mu_H \gamma_H^{-1}$. The second step is to find the rest of the parameters by reducing the model to a system involving the terms of the populations $U(t)$, $I(t)$, and $H(t)$ only, since it is for these compartments that data is available. Once the reduced system is found by means of the minimisation algorithm L-BFGS-B,[21,22] we find the maximum likelihood estimation (MLE) corresponding to the negative log-likelihood described in the SM. In both cases, we explore the relationship between model parameters where we have access to reliable datasets to mitigate parameter identifiability issues.[23–25] Details of the linear relationship of discharges and deaths in hospital and the reduced model are given in the SM.

Forecasting and validation

As outlined above, models of this variety often lack parameter validation and thus lack the ability to predict relatively far into the future.[9] To validate the predictive power of our modified SEIR-D model (Figure 1), we used the previously outlined inference algorithm to obtain new estimates for the model parameters using only a limited number of data points, and focused on predicting the hospital admissions, discharges, and occupancy due to having a large dataset, using a minimum of 12 and a maximum of 51 data points. We evaluated the predictive power of a parameter set by performing a prediction for the next 10, 20 and 30 days, starting the day after the last data point used for the parameter estimation. By comparing the prediction with the available data, we computed the percentage of days that are correctly predicted. It was considered that a day is correctly predicted if it lies within a given tolerance of standard deviations from the available data. This approach quantifies the risk associated with the decision of selecting a certain number of days into the future, e.g. 10, 20 and 30 days.

Role of the funding source

The funders of this research played no role in the derivation of the model, the novel inference algorithm, the application of the methods nor the generation of this manuscript. All authors had full access to all the data and had final responsibility for the decision to submit for publication.

Results

Parameter values

Using the compartmental model along with the novel inference algorithms, we derived the parameter values summarised in Table 1. Figure 2 shows the daily number of patients admitted to hospital, those in hospital and those who were discharged, respectively. To demonstrate the accuracy of the fitting procedure, we super-impose the observed datasets and their continuum mathematical counterparts as well as their 95% confidence intervals (95% CI) for these curves. It can be easily verified that the fitting captures the trends of the data and fits the majority of the data within the 95% CIs. Moreover, small perturbations in most parameters result in small changes to the overall fit of the data, whilst others result in quite a large change in the overall fit (See SM for details). In this setting, the perturbation of the parameters producing large changes in the overall fit is due to the actual parameters themselves being well characterised from the data rather than sensitivity of the model. This can be seen since the standard deviation of the parameters when smaller datasets were used to calculate the parameters are low, as well as the forecasting pattern not changing significantly. Accounting for the noise in the log-likelihood and the prediction technique, as well as the other sensitivity tests, demonstrates the robustness of the model when the actual data are perturbed and fitted. It is noteworthy that our set of optimal inferred parameters give a value of effective reproduction number $\mathcal{R}_t$ = 0·85 throughout the lockdown.

Parameter	Value	Epidemiological meaning
β	0·151 days $^{-1}$	Transmission rate
γ_E^{-1}	6·51 days	Average incubation period
p	0·929	Fraction of undetected cases
γ_U^{-1}	6·10 days	Average infectious period (undetected cases)
γ_I^{-1}	6·28 days	Average infectious period (hospital cases)
γ_H^{-1}	19·7 days	Average hospitalisation period (recovered)
m_U	0·0220	Infected fatality ratio (undetected cases)
μ_H^{-1}	12·9 days	Average hospitalisation period (deaths)

Table 1: Description of the parameters of the compartmental model, and their values when the model is fitted to the data. The values are inferred using only the data from the Sussex region, without taking any information from other regions or countries.

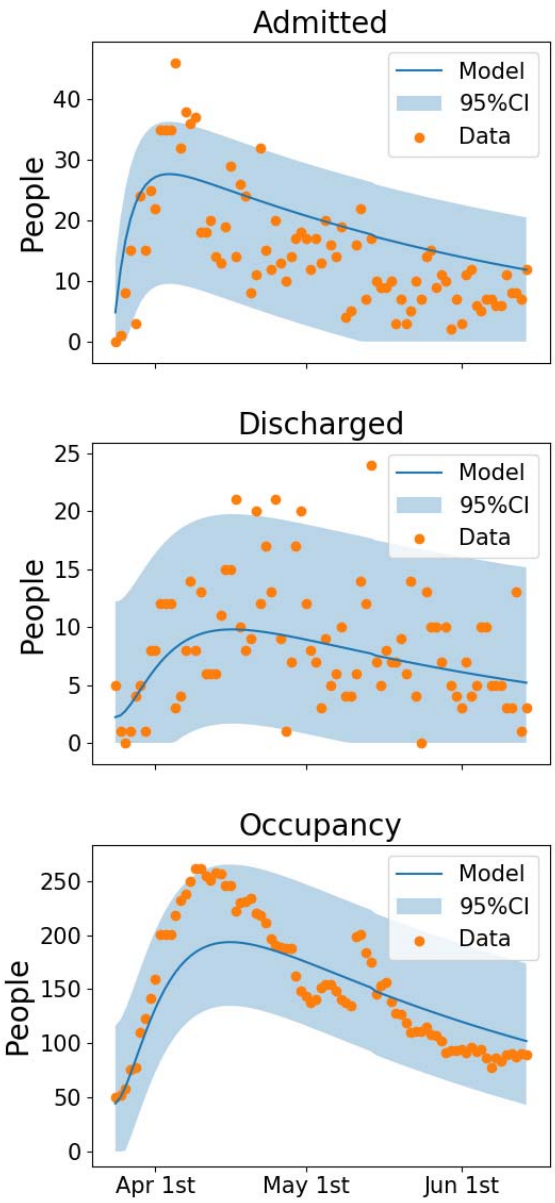

Figure 2: Output of the compartmental model and comparison with data. The solid line represents the output of the model with the parameters inferred from the data. The shaded region depicts the 95% confidence interval (95% CI) computed from the data, that is, attributing all the error to measurement error. The dots correspond to observed data. Since all data are collected by manual counting and recording, there is a significant amount of noise. Furthermore, we cannot verify that the counting protocol has not changed during the period. There are between 1 and 5 outliers in each data set, out of a total of 82 data points, but generally the model captures the dynamics of the data and the situation.

Comparisons between our parameters and those used widely in the literature are shown in Table 2.[1,6,15] It must be noted that the physical interpretations of some of the parameters differ from one model to another, however, the overall picture appears plausible. To our knowledge, there has not been a quantified estimate of the transmission rate as well as the infected fatality ratio outside of hospitals, only scaled estimates and these are based on pre-conceived values of the basic reproductive number and the recovery rates given by the Imperial College London model[1] and other similar reports.

Parameter	Campillo-Funollet et al. (2020)	Ferguson et al. (2020)[1]	Kissler et al. (2020)[6]	Lourenço et al. (2020)[15]
Average incubation period	6·51 days	5·1 days	4·6 days	N/A
Fraction undetected cases	0·929	0·956	0·956	N/A
Average infectious period	6·28 days	5 days	5 days	4·5 days
Average hospitalisation	19·74 days	8-16 days	6-8 days	N/A
Infected fatality ratio	0·0220	0·009*	N/A	0·14**

Table 2: Comparison of parameter values from different studies. There is evidence to support that one becomes infectious before presenting symptoms,[26,27] and also that one becomes infectious after presenting symptoms.[6] Different studies use different definitions for the incubation period, for instance time from exposure to onset of symptoms instead of time from exposure to transmissibility. This therefore has a knock-on effect on the understanding of the average infectious period. *The infected fatality ratio (IFR) in Ferguson et al. (2020)[1] includes all cases, whilst in our model, it is limited to non-hospital cases but is heavily influenced by mortality in care-homes. **In Lourenço et al. (2020)[15], the IFR is limited to severe cases.

Predictive power of the SEIR-D model

Using the predictive power method outlined above corresponds to a total of 1776 parameters sets. The resulting values of the parameters from the inference algorithm using the subsets of data are similar to the global fit using all the available data. Figure 3 shows the results for predictions 10, 20 and 30 days into the future. To our knowledge, this is the first result of its kind that validates the forecasting in this manner.

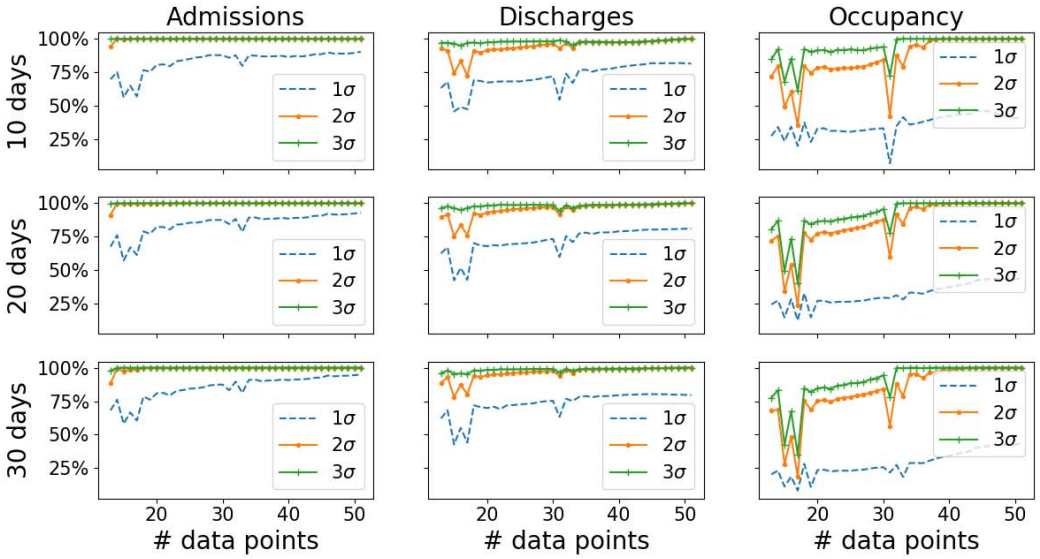

Figure 3: Validation of the predictive power of the model. We fitted the parameter models using all possible sequences of consecutive admissions, discharge and occupancy data points, from 12 to 51 points. Note that since there are only 82 data points available, we could not use more than 51 points to validate a prediction for 30 days, otherwise we will not have data to compare with. The predictive power is quantified as the number of days predicted within an accuracy of one, two, or three standard deviations of the data. There are significant differences in the predictive power for different variables. Admissions and discharges can be accurately predicted using as low as 15 data points to fit the model, whilst hospital occupancy requires about 30 data points to reach the same accuracy levels. Admissions/discharges and occupancy are different in nature: the former are rates (individuals per day), whilst the latter are absolute counts - this might explain the difference in the predictive power.

Discussion

Predicting the local/regional resurgence of COVID-19 is the number one priority of governments and local authorities in the UK and around the world, to control and halt the local and national transmission of infection. The pandemic itself has thrown to the forefront of science the role and utility of epidemiological modelling at a time when questions of urgency, national importance, and uncertainty simultaneously come into play, thereby exposing its current limitations in terms of predictions and forecasting.[9] A comment by Saltelli et al. (2020)[9] outlines a manifesto highlighting five ways in which mathematical models should serve society. These include minding the assumptions (the minimal the better), being mindful of model complexities (hubris – balancing the usefulness of the model with the breath of its predictions), being mindful of the interests of the researchers (techniques and methodology can be limited in scope to the expertise of the researchers), being aware of the consequences (mitigate the uncertainty), and finally being mindful of the unknowns (communicating what is unknown is as important as communicating what is known). Our approach is based on these five pillars to ensure that our research outcomes are engrained and driven by reliable local surveillance data with minimal assumptions and an explicit simple data-formulated model. Predictive epidemiological modelling applied to local data has the unique ability to offer local authorities a framework for decision-making that is based on temporal trends of these local datasets. Modelling lessons learnt at the local level can possibly be transferred to the national arena to help guide data acquisition such that datasets are amenable to model-data prediction approaches as well as providing avenues for short-, medium-, and long-term forecasting.

To date, many models have failed to make meaningful quantitative predictions and forecast about the impact of COVID-19 despite employing vast amounts of resources and highly sophisticated tools.[1,4,6,16,28] During the early stages of COVID-19, parallels between COVID-19 and the Spanish flu (among other influenza diseases) that killed more than 50 million people with an average age of 28 years, were drawn.[1,4,6,16,28] As a result, to mitigate and prepare for COVID-19 hospitalisations and deaths, national governments and hospitals suspended or postponed important critical diagnostic procedures/treatments, such as cancer diagnosis and treatment. The mental health of the nation and access to health services deteriorated enormously as patients with debilitating conditions avoided visiting hospitals, and yet, locally, the number of COVID-19 deaths was nowhere near the expected numbers predicted by the national models driving the national government decision-making process.[1,4,6,16,28] Recent studies have highlighted how predictions need to be transparent and humble in order to instill confidence and invite insight and not blame.[9] For a disease such as COVID-19, espoused wrong predictions can have a devastating effect on billions of people around the world in terms of the economy, job-security, health, education, and societal turmoil, just to mention a few. In this report, we have demonstrated that our inference process and resulting parameters allow us to produce forecasts for up to 30 days into the future to a high accuracy, for quantities of interest such as hospital occupancy, where such a time period can ensure that decisions, and changes in decisions, can be enacted. The underlying temporal dynamics fit the pattern of an infectious disease outbreak and does not rely solely on statistically inferred parameters, in the absence of a dynamic model. Such statistical models lack the ability for long-term forecasting. A recent review by Jewell et al. (2020)[29] established the need for accurate forecasting in the timescales we have demonstrated to help ease public uncertainty and anxiety by aiding local policy planning in the exact manner we are using the presented results in our collaborations with the Sussex local authorities and public health departments.

It is clear from the literature that the lack of predictions and forecasting is closely correlated with the underlying theoretical assumptions and the use of pre-determined values of the parameters that are alien to the models under study.[9] This, in turn, is driven by the lack of reliable

datasets appropriate for model-data validation and sensitivity analysis. In this study, we have proposed a bottom-up approach where a model built on local datasets has the ability to guide local decision-making in terms of healthcare demand and capacity, in particular given the likelihood of COVID-19 secondary spikes/waves. Other widely used publications, such as Ferguson et al. (2020)[1], used datasets mainly from Wuhan and other national datasets for similar infectious diseases which means that overall policy and data collection will, in general, differ to the current situation. The important highlight and applicability of our work is that we used local datasets for our modelling, and so we fully understand how the data were collected and know exactly the physical interpretation of the parameters, something that cannot be claimed by using the parameters found in the other publications of predictive modelling.[29] The SEIR-D model itself is simple and transparent. Moreover, we have designed our approach in such a way that this method can be used by other regions/counties across the UK provided they have the required data, and as such, we have created a toolkit which makes our approach more accessible.[30] This allows for users who are not familiar with mathematical modelling to use our approach and generate their own parameters to inform local policy.

Our modelling framework is not only tailored to deal with COVID-19 but can be applied to other excess death situations in summer and winter months which are known to kill thousands of people every year. Epidemic forecasting and the development of early warning systems for healthcare demand and capacity has been thrown to the forefront of epidemiological modelling. By working in close collaboration, theoreticians, local authority Public health teams, and NHS planners have a unique opportunity to bring novel approaches to healthcare decision-making and planning within Public Health, Health and Social Care, with forecasting capabilities similar to those for weather forecasting.

Author Contributions

The Sussex Modelling Cell (PA, MB, WB, JC, GE, MD, KG, GP, RW, MW) conceived and presented the research questions. ECF, JVY, and AMa formulated the epidemiological model, conducted mathematical and statistical analysis and simulations of the model, and drafted the manuscript. PA and KG generated the datasets from the Sussex Clinical Commissioners and the ONS and did appropriate checks to assert the completeness and accuracy of the data. ECF and JVY assembled the figures with substantial contribution from AMa. PA, JC, MD, KG, AMe and GP critically reviewed and edited the manuscript for intellectual content. All authors contributed to the final manuscript.

Declaration of Interest

All authors declare no conflict of interest.

Data management statement

All of the computational data output is included in the present manuscript. All images that are presented in this report are available upon request.

Acknowledgements

JVY acknowledges postdoctoral support from the Brighton and Hove City Council, East and West Sussex County Councils.

References

1 Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College, London

2020; published online March 16. DOI:10.25561/77482.
2 Blyuss KB, Kyrychko YN. Effects of latency and age structure on the dynamics and containment of COVID-19. medRxiv 2020; published online April 29. DOI:10.1101/2020.04.25.20079848 (preprint).
3 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 2020; 20: 533–4.
4 Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of COVID-19. arXiv 2020; published online April 9. arXiv:2004.04741 (preprint).
5 Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020; published online 8 May. DOI:10.1126/science.abb6936.
6 Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 2020; published online 22 May. DOI:10.1126/science.abb5793.
7 Viner R, Russell S, Croker H, et al. School Closure and Management Practices During Coronavirus Outbreaks Including COVID-19: A Rapid Narrative Systematic Review. SSRN 2020; published online March 27. DOI:10.2139/ssrn.3556648 (preprint).
8 Alvarez David Argente Francesco Lippi FE, Golosov M, Gozzi F, et al. A Simple Planning Problem for COVID-19 Lockdown. NBER 2020; published online April. DOI:10.3386/w26981.
9 Saltelli A, Bammer G, Bruno I, et al. Five ways to ensure that models serve society: a manifesto. Nature 2020; published online June 24. DOI:10.1038/d41586-020-01812-9.
10 Eichenbaum M, Rebelo S, Trabandt M. The Macroeconomics of Epidemics. NBER 2020; published online April. DOI:10.3386/w26882.
11 Jones CJ, Philippon T, Venkateswaran V. Optimal Mitigation Policies in a Pandemic: Social Distancing and Working from Home. NBER 2020; published online April. DOI:10.3386/w26984.
12 Acemoglu Victor Chernozhukov Iván Werning Michael Whinston DD, Anne Dix R, Garg T, et al. Optimal Targeted Lockdowns in a Multi-Group SIR Model. NBER 2020; published online May. DOI:10.3386/w27102.
13 Docherty AB, Harrison EM, Green CA, et al. Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol. medRxiv 2020; published online April 28. DOI:10.1101/2020.04.23.20076042 (preprint).
14 Friston KJ, Parr T, Zeidman P, et al. Dynamic casual modelling of COVID-19. arXiv 2020; published online April 9. arXiv:2004.04463 (preprint).
15 Lourenço J, Paton R, Ghafari M, et al. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. medRxiv 2020; published online March 26. DOI:10.1101/2020.03.24.20042291 (preprint).
16 Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLoS One 2020; 15: e0231236.
17 Campillo-Funollet E, Venkataraman C, Madzvamuse A. Bayesian Parameter Identification for Turing Systems on Stationary and Evolving Domains. Bull. Math. Biol. 2019; 81: 81–104.
18 Stuart AM. Inverse problems: A Bayesian perspective. Acta Numer. 2010; 19: 451–9.
19 Allen LJS, Burgin AM. Comparison of deterministic and stochastic SIS and SIR models in discrete time. Math. Biosci. 2000; 163: 1–33.
20 Satsuma J, Willox R, Ramani A, Grammaticos B, Carstea AS. Extending the SIR epidemic model. Physica A 2004; 336: 369–75.
21 Fletcher R. Practical Methods of Optimization. Chichester, UK: John Wiley & Sons, Inc, 1987.

22 Malouf R. A comparison of algorithms for maximum entropy parameter estimation. In: Roth D, van den Bosch A, eds. Natural Language Learning: Sixth International Conference, CoNLL 2002, Taipei, Taiwan, August, 2002. Proceedings. USA: Association for Computational Linguistics, 2002.
23 Boileau T, Leboeuf N, Nahid-Mobarakeh B, Meibody-Tabar F. Online identification of PMSM parameters: Parameter identifiability and estimator comparative study. IEEE Trans. Ind. Appl. 2011; 47: 1944–57.
24 Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J. Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics 2014; 30: 1440–8.
25 Wu H, Zhu H, Miao H, Perelson AS. Parameter identifiability and estimation of HIV/AIDS dynamic models. Bull. Math. Biol. 2008; 70: 785–99.
26 Ganyani T, Kremer C, Chen D, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Eurosurveillance 2020; 25: 2000257.
27 Xia W, Liao J, Li C, et al. Transmission of corona virus disease 2019 during the incubation period may lead to a quarantine loophole. medRxiv 2020; published online March 08. DOI:10.1101/2020.03.06.20031955 (preprint).
28 Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect. Dis. 2020; 20: 669–77.
29 Jewell NP, Lewnard JA, Jewell BL. Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections. JAMA 2020; 323: 1893–4.
30 Halogen Team. Halogen: forecasting the local impact of COVID-19. In Preparation, August 2020. https://alpha.halogen-health.org.