Hospital Length of Stay For COVID-19 Patients: Data-Driven Methods for Forward Planning

DOI:
10.21203/rs.3.rs-56855/v1
https://doi.org/10.1186/s12879-021-06371-6

Document Version
Accepted author manuscript

Published in:
BMC Infectious Diseases

RESEARCH

Hospital length of stay for COVID-19 patients: Data-driven methods for forward planning

Bindu Vekaria[2,3,6*], Christopher Overton[2,7^], Arkadiusz Wiśniowski[1†], Shazaad Ahmad[8], Andrea Aparicio-Castro[1], Jacob Curran-Sebastian[2,6], Jane Eddleston[6], Neil A Hanley[5,6], Thomas House[2,4,6], Jihye Kim[1], Wendy Olsen[1], Maria Pampaka[1], Lorenzo Pellis[2], Diego Perez Ruiz[1], John Schofield[6], Nick Shryane[1] and Mark J Elliot[1]

[*]Correspondence: bindu.vekaria@manchester.ac.uk
[2] Department of Mathematics, University of Manchester, Oxford Road, M13 9PL Manchester, UK
Full list of author information is available at the end of the article
[†]Co-corresponding author: a.wisniowski@manchester.ac.uk
[^]Co-corresponding author: christopher.overton@manchester.ac.uk

Abstract

Background: Predicting hospital length of stay (LoS) for patients with COVID-19 infection is essential to ensure that adequate bed capacity can be provided without unnecessarily restricting care for patients with other conditions. Here, we demonstrate the utility of three complementary methods for predicting LoS using UK national- and hospital-level data.

Method: On a national scale, relevant patients were identified from the COVID-19 Hospitalisation in England Surveillance System (CHESS) reports. An Accelerated Failure Time (AFT) survival model and a truncation corrected method (TC), both with underlying Weibull distributions, were fitted to the data to estimate LoS from hospital admission date to an outcome (death or discharge) and from hospital admission date to Intensive Care Unit (ICU) admission date. In a second approach we fit a multi-state (MS) survival model to data directly from the Manchester University NHS Foundation Trust (MFT). We develop a planning tool that uses LoS estimates from these models to predict bed occupancy.

Results: All methods produced similar overall estimates of LoS for overall hospital stay, given a patient is not admitted to ICU (8.4, 9.1 and 8.0 days for AFT, TC and MS, respectively). Estimates differ more significantly between the local and national level when considering ICU. National estimates for ICU LoS from AFT and TC were 12.4 and 13.4 days, whereas in local data the MS method produced estimates of 18.9 days.

Conclusions: Given the complexity and partiality of different data sources and the rapidly evolving nature of the COVID-19 pandemic, it is most appropriate to use multiple analysis methods on multiple datasets. The AFT method accounts for censored cases, but does not allow for simultaneous consideration of different outcomes. The TC method does not include censored cases, instead correcting for truncation in the data, but does consider these different outcomes. The MS method can model complex pathways to different outcomes whilst accounting for censoring, but cannot handle non-random case missingness. Overall, we conclude that data-driven modelling approaches of LoS using these methods is useful in epidemic planning and management, and should be considered for widespread adoption throughout healthcare systems internationally where similar data resources exist.

Keywords: COVID-19; Length of stay; Survival Analysis; England

Background

Since its emergence in December 2019 and classification in January 2020, SARS-CoV-2, the coronavirus that causes COVID-19, has spread rapidly, with 270 thousand confirmed infections in the UK by the end of May 2020 [1]. The exponential growth in the early days of each nation's outbreak has led to a doubling time of around three days [2]. Coupled with potentially high estimates of R_0 (the average number of new infections generated by an infected individual, in the absence of control measures and population acquired immunity) [3, 4, 5], this has continued to have substantial impacts on healthcare systems across the world. Large growth rates and a delay between new infections and their detection can lead to unexpected surges in bed demand. In order to restrict the spread of the pathogen, many countries have implemented mass quarantine (also known as lockdown) strategies, including England where the mass quarantine began on 23 March 2020 [6]. However, the effects of such interventions are not seen for at least a week [7], emphasising the need for careful, evidence-based, planning; particularly as the easing of mass quarantine measures is considered. In this context, the use of clinical care data to predict the demand for hospital and Intensive Care Unit (ICU) beds by patients presenting with COVID-19 is invaluable in optimising the effectiveness of planning by hospitals and, therefore, patient outcomes.

Understanding the impact of COVID-19 on hospital capacity breaks down into two core measurement tasks: first, to predict incidence (and thereby hospital admissions rates); and second, to estimate total length of stay (LoS) accurately allowing for variation in severity of disease and healthcare needs. The combination of these two measures can then be used to predict bed demand. This challenging task requires a careful modelling approach, particularly when high-quality data is limited within often fragmented healthcare systems. National datasets are crucial in understanding demand in hospitals across the country, but are flawed by amounts of record-level (or whole case) missingness that can bias the estimates. Routinely collected data generated by individual hospitals are, by definition, smaller and non-general but tend to be less prone to missingness and these can complement national data by providing insights for planning on a local level.

Estimating LoS has not been the primary focus of previous modelling; and studies that calculate LoS tend to use *ad-hoc* approaches [8]. There is currently a lack of statistically principled modelling that accounts for both delays in patient outcomes and complex hospitalisation pathways. This problem is particularly important during the COVID-19 pandemic, since some groups of patients spend extended periods in hospital, and, for the most severe cases, in critical care. Furthermore, estimates of LoS that use deterministic models or observations drawn directly from data fail to take missingness into account [9, 10, 11]. Accurately calculating LoS therefore requires mathematical and statistical techniques that specifically address these issues.

In this paper, we present three methods for estimating LoS for patients with COVID-19 infection using both a nationally collected dataset and local data from a large inner city hospital Trust in the UK. The truncation corrected (TC) method corrects for the fact that observations are truncated at the day of reporting; accelerated failure time models (AFT) explicitly account for all observed LoS including

those censored by not having seen the outcome; and the multi-state (MS) approach analyses LoS and takes into account dependence between outcomes such as discharge or death. Finally, we include measures of uncertainty in each of our model results, which should be incorporated into hospital planning strategies. With this principled approach, past data can be appropriately used to better prepare for the next phase of the COVID-19 pandemic.

The results presented in this article use data that were available as of 26 May 2020. At this stage of the pandemic, many patients were still in hospitals, leading to right-censoring in their lengths of stay. To evaluate the performance of the methods at correcting for this right-censoring, we compare the estimated distribution to the full LoS distributions, using data available as of 21 January 2021. We do not re-analyse the LoS for the second and third waves, since this manuscript focuses on comparing methods for estimating LoS whilst correcting for right-censoring. However, the methods are readily applicable to these more recent data.

Methods

Data

Outcome variables

We define two outcome events: death or discharge. All patients admitted to hospital will eventually experience one of these two outcomes. Then, we model LoS from hospital admission to either death or discharge. For the analysis shown in the Results section, we focus on LoS until any outcome, to facilitate comparison of the three methods. We account for whether the patient was in ICU or not and also estimate the LoS from hospital admission to ICU and LoS on ICU. In Appendix F, we further examine different outcomes using the TC and MS methods.

CHESS

The COVID-19 Hospitalisation in England Surveillance System (CHESS) [1] collects reports from all NHS acute care hospital trusts to provide daily patient-level and aggregate data on COVID-19 hospitalisations. In the patient-level data, patients are followed through their hospitalisation pathway; the dates of various events are recorded, such as date of admission to hospital, date of admission to ICU and final outcome date.

CHESS predictors

We used four variables as predictors. First, *sex*, for which we removed patients with unknown values. Second, *age*, which we grouped into four categories (< 50, $50-64$, $65-74$, $75+$), and removed negative values and patients with a recorded age equal to zero (which did not seem genuine, based on the number of such cases and other factors such as comorbidities). Third, *week of admission to hospital*, which, in the TC model, we categorised in two groups: weeks 12 to 14 (i.e. from 16 March to 5 April 2020), and weeks 15 to 20 (from 6 April to 17 May 2020). In the AFT model, we used single week as a fixed effect predictor but present results for the two groups of admissions. Fourth, we used a binary indicator on whether a patient was

[1] Since October 2020 this has been replaced with the Severe Acute Respiratory Infection (SARI) data.

admitted to ICU or not, and omitted the patients for whom this information was unknown. The resulting analytical sample is $n = 6208$. Details of the data processing procedure, and inclusion/exclusion criteria, are presented in Appendix C.

Whilst we can identify predictors such as sex, age, and week-of-admission from these data, we cannot identify other potential predictors such as which variant contributed to the infection or treatment strategies. This would be of interest with the emergence of new variants of concern. Instead, the effect of new variants has to be approximated using week-of-admission, but this may be confounded with other factors, such as treatment changes and hospital burden.

Routinely collected hospital data (MFT)

Routine data on the hospitalisation of patients were provided by Manchester University NHS Foundation Trust (MFT). MFT is the largest NHS Trust in England, comprising nine hospitals and accounting for approximately 2.5% of the National Health Service. For COVID-19 admission, there were three geographically distinct acute hospitals across South and Central Manchester: Manchester Royal Infirmary; Wythenshawe Hospital; and Trafford General Hospital. MFT serves the population of Greater Manchester, a large, ethnically diverse conurbation of approximately 2.8 million people. The data follow all patients through their clinical pathway for the duration of a single hospitalisation, and provide timings and lengths of stay in all critical care episodes. Patient data are complete unless patients are still in hospital, in which case they are censored.

MFT data preparation

Data were drawn from the Patient Administration System (PAS) and WardWatcher to join information on a patient's hospitalisation pathway and critical care episodes. Patients were selected from the MFT database if a swab was taken either on the day of their hospitalisation, or within two days of their hospital admission, and tested positive for COVID-19. This was to discount any hospital-acquired cases since COVID-19 positive cases who required hospitalisation due to non-COVID related health conditions may bias LoS estimates. We also excluded patients admitted for elective procedures requiring treatment for chronic illnesses such as dialysis. As a result of having multiple admissions close together, it was difficult to determine whether these cases were hospital-acquired or genuine COVID-19 admissions. The resulting sample included $n = 786$ patients. The models based on the MFT data did not use information on predictors due to the smaller sample size, although from a methodological point of view these could be easily added to the models. Details of the data generating process are presented in Appendix A.

Data quality issues in length of stay data

There are several types of data quality issues that tend to be present in length of stay data and all are present in one or both of the two datasets. Some of these are a consequence of the reporting and data collection methods. Others are inherent to the nature of outbreaks, and will be present regardless of the data collection. Here, we present some key issues that need to be adjusted for, and discuss the implications of ignoring them. Accounting for these biases for COVID-19 can enable robust estimates that provide timely insight for policy and planning.

Missing cases

One issue with the CHESS dataset is missing cases. For example, the number of deaths recorded in CHESS is considerably less than the official figures. These also suffer from reporting lag issues but some indication about the level of missingness in CHESS can be obtained by comparing to the COVID-19 patient notification system (CPNS), which records all deaths attributed to COVID-19 in England. On 26 May, there were 23504 deaths in hospital as attributable to COVID-19 in the CPNS data. This compares to an equivalent figure of 4071 in the raw CHESS data for the same day. This is indicative of case level missingness within CHESS of over 80%. We discuss this issue in more detail in Section "Discussion".

Missing values on important variables

Many rows in the data are incomplete. This is particularly problematic for data pertaining to outcome events: for example in some cases it is unclear whether a patient has not been discharged yet, or whether they have but the data have not been recorded. The amount of, and patterns of, missing patient information in the CHESS data is associated with the trust that reports the cases, with varying levels of missingness across different trusts (see Appendix B).

Censoring

In time-to-event studies, we observe a collection of individuals who are infected or have been exposed to infectious material. If these individuals could be followed indefinitely, the outcomes of all individuals would be observed. Therefore, these data can be used to determine the length of stay in the various compartments (states) of the disease progression pathway, as well as the probabilities of transitions into other states. However, during an outbreak we only observe individuals up until the most recent reporting date. This leads to right-censoring (e.g. [12]), when we only know the lower bound of duration until the next event in the pathway, and cannot accurately determine the length of time until their next transition nor to which state this will be. Thus, censoring may lead to the underestimation of the LoS.

Truncation bias

To remove the uncertainty around censored cases, we can instead condition our sample to only look at cases for whom the outcome has been observed. However, such a sample includes only cases with outcomes that occurred before the most recent reporting date, causing the sample to be truncated by the reporting date. This truncation leads to an over-expression of short LoS, since the recently infected individuals are only included if their LoS is short. Failing to account for this bias will underestimate the LoS of interest. [2]

Truncation is exacerbated by exponential growth in the early stages of an outbreak, since a higher proportion of cases will have been infected recently. By the final phase of an outbreak, truncation has a smaller effect since the majority of cases occurred sufficiently long ago to be unaffected by the truncation date. However, it will always be present as long as the epidemic is ongoing. Even in these late stages, whilst it may have a negligible impact across the whole outbreak, its effect might

[2] This can be seen in Figure 2 by comparing the TC results to the LoS observed in the data.

be of concern in certain scenarios, such as when using time as a predictor variable. In such a case, for events early in the epidemic, truncation will have very little effect, but for more recent events many cases may still be truncated. Such biases are often considered in the HIV literature [13, 14], due to the long infectious periods involved, but are often ignored for acute outbreaks. As alluded to in [15], this is potentially due to high quality data being available only after an explosive outbreak has finished, by which point these biases have little or no effect. However, when attempting to control ongoing epidemics, we require estimates of LoS distributions that are robust in the face of censoring and truncation.

Survival analysis

Survival analysis describes a collection of statistical procedures for which the outcome of interest is time until an event, often as a function of predictor variables [16, 17, 18]. A central assumption of most survival analytic methods is that the time to event will have been censored for some observations, as discussed in Section "Data quality issues in length of stay data".

Survival analysis may assume an underlying distribution for LoS in each state. Generally, LoS are observed to be right-skewed, so a distribution with this property should be used. In this paper, LoS is assumed to follow a Weibull distribution, which is a popular choice in survival analysis as it is robust in terms of violation of its assumptions. Therefore, the choice allows us to focus on the comparison between the different methods rather than the issues of model fit.

Figure 1 outlines the model used to represent the hospital pathways we consider in our analysis. Allowed transitions are indicated by directed arrows between any two states. Below, we outline the survival methods we selected for our analyses. Code for all methods is available at `https://github.com/thomasallanhouse/covid19-los`.

Accelerated Failure Time (AFT) model

In the AFT model, rather than considering all of the hospitalisation pathways shown in Figure 1, we focus on predicting LoS in a given state, until another pre-specified event occurs. That is, we are interested in estimating the time between subsequent events in the pathway, such as from hospital admission to being admitted to ICU. We aggregate the final outcomes of death and discharge into a single outcome. This is necessary since it is not clear what the outcome will be for the censored cases in the CHESS data.

The response variable is the natural logarithm of the LoS, denoted by $\ln(t)$, which is explained by a vector of predictors x, with associated parameter vector β, and error term ξ:

$$\ln(t) = x \cdot \beta + \xi. \tag{1}$$

The assumed probability distribution of ξ defines the hazard function, i.e. the probability that a case will experience an event at time t, given that they have not already experienced it until time t [19, 20]. For ξ we assumed a Weibull distribution, giving the hazard function $h(t) = p\lambda t^{p-1}$, where $\lambda = \exp(-px \cdot \beta)$ and p is the

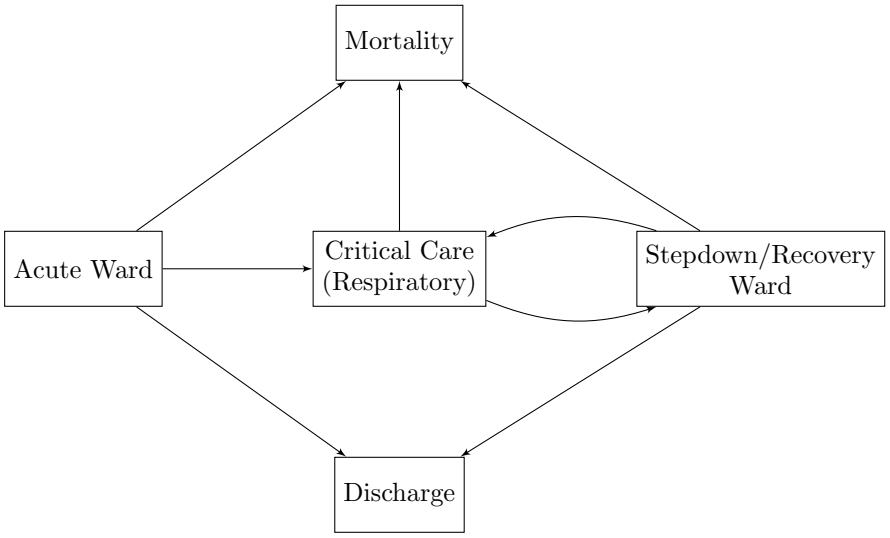


Figure 1 A schematic representation of the possible hospital pathways considered by our methods; at any given time, patients are considered to be in one of the five following states: Acute Ward, Critical Care, Stepdown Ward, Discharge or Mortality.

shape parameter defining the Weibull distribution. If $p > 1$ the hazard is increasing over time, if $p < 1$ the hazard is decreasing over time, and for $p = 1$ the hazard is constant over time (which is equivalent to an exponential error term distribution). The predictors x therefore increase or decrease the hazard and so accelerate (shorten) or decelerate (lengthen) the time to event, t.

The AFT model explicitly takes into account cases with right-censoring [20]. Thus, the model corrects for the potential underestimation of the LoS when only a portion of patients in the sample have observed the event.

A limitation of this simple model is when there is more than one potential event of interest [18]. In this study there were two events of interest: death and discharge. These are 'competing hazards', i.e. if a patient experienced one they were censored for experiencing the other. We could have run the model twice, once for each event, and treated patients who experienced the other event as being censored. This would have given unbiased results if the competing hazards were independent, but, for a given patient, as the hazard of death increases, it decreases for discharge, and vice versa. For this reason we considered a model of the joint event: death *or* discharge.[3]

We fitted separate models for patients who never entered ICU versus patients who did enter ICU at some point, as these groups were expected to have different baseline hazard functions. In all models, the predictors in x were sex, age group and week of hospital admission (see Section "CHESS predictors").

All models were estimated using JAGS software implemented in the `rjags` R package [21]. For the shape parameter, we used a uniform prior, $p \sim \mathrm{U}(0, 10)$, which represents our lack of information on this parameter. There is not a conjugate prior

[3] This is not as counter-intuitive as it might sound since, although death is certainly not an equivalent outcome for the patient, our primary concern here is in length of stay regardless of outcome.

simultaneously for both the shape and scale parameters in the Weibull distribution [22]. An alternative specification for this prior is a Gamma distribution [23]. However, in our tests the results were virtually the same with both priors for p. The scale parameter λ is specified via a prior for the predictors' coefficients β, which is multivariate normal with mean zero and variance equal to 10, i.e. each element of β is distributed as $\mathcal{N}(0, 10)$.[4]

Truncation corrected method

In this method, we again focus on estimating the single LoS in a given state. We assume that LoS is given by a random variable X, drawn from a distribution with density function $f_\theta(\cdot)$, parameterised by a set of parameters θ. In this analysis, we assume that X is drawn from a Weibull distribution. We aim to determine the underlying parameters for this distribution by fitting the observed data using maximum likelihood estimation.

To use maximum likelihood estimation, we need to construct a likelihood function for the observed data. For each data point, the LoS is not directly observed. Instead, the arrival and departure dates and/or times that bracket the period of stay are observed. These correspond to two random variables, E_1 and E_2, linked by the LoS random variable, i.e. $E_2 = E_1 + X$. Instead of treating incomplete entries as censored, here we condition the data on observing both events. For example, if interested in the time from hospital admission to ICU admission, we condition on cases that have been admitted to hospital and to ICU. This introduces a truncation bias (See Section "Truncation bias"), which needs to be corrected in the likelihood function. This approach does not take into account competing hazards, since we condition the data on observing the outcome of interest. However, this method enables LoS for different patient outcomes to be estimated, since censored cases are not included.

Our likelihood function is defined as the probability that the second event occurs on the observed date, given the time of the first event and that the second event must have occurred before the truncation date [14]. This removes censored observations since we condition on observing the second event. Therefore, we need to find

$$f(E_2 = e_2 \mid \{E_1 = e_1\} \cap \{E_2 \leq T\}) = \frac{g_{E_1,E_2}(e_1, e_2)}{\int_{e_1}^{T} g_{E_1,E_2}(e_1, x)\mathrm{d}x}, \tag{2}$$

where g_{E_1,E_2} is the joint distribution of E_1 and E_2. The time of the second event is the time of the first event plus the delay, $E_2 = E_1 + X$. Therefore $g_{E_1,E_2} = g_{E_2|E_1}(e_2 \mid e_1)g_{E_1}(e_1) = f_\theta(e_2 - e_1)g_{E_1}(e_1)$, which gives

$$f(E_2 = e_2 \mid \{E_1 = e_1\} \cap \{E_2 \leq T\}) = \frac{f_\theta(e_2 - e_1)g_{E_1}(e_1)}{\int_0^{T-e_1} f_\theta(x)g_{E_1}(e_1)\mathrm{d}x} = \frac{f_\theta(e_2 - e_1)}{\int_0^{T-e_1} f_\theta(x)\mathrm{d}x}. \tag{3}$$

This can be maximised across all data points to find the maximum likelihood estimator for θ. [5]

[4] The model can also be estimated using maximum likelihood implemented in Stata 14 using the command `streg` (`https://www.stata.com/manuals/ststreg.pdf`)

[5] We maximise this using command fminsearch in MATLAB, but it is relatively simple to implement in any language. We provide both MATLAB and Python code in the Github repository.

This method can be used to examine LoS to individual outcomes by specifying the events, e.g. specifying that the second event is a death. Additionally, the effect of predictor variables can be analysed by sub-setting the data and then modelling the LoS of each subset.

Multi-state model

Multi-state survival analysis extends the above two methods by permitting us to model the time to multiple outcome events in the presence of competing hazards [24, 25]. Thus, we can model complex patient pathways upon admission to hospital.

Each permitted transition in Figure 1 is a survival model, where the instantaneous rate of transition from one state, r, to another state, s, otherwise known as the transition intensity, can be modelled similarly to hazard functions. For all transitions, we assume a Weibull AFT model, but this method can easily accommodate the use of any parametric or flexible parametric models used in standard survival analysis [19]. When there are n_r competing events for state r, a patient entering state r at time t_j has their next event at t_{j+1}, which is given by the minimum of the survival times for the competing events, $s_1, \ldots, s_{n_r}$.

The data are formatted in such a way that we have a series of event times and LoS, each corresponding to a change in state. The last of these may be observed so that the patient has entered an absorbing state, i.e. they are discharged or dead, or right-censored if the patient is still in the hospital. Therefore, the data to inform the n_r models consist of an indicator corresponding to whether or not the transition is observed or censored at t_{j+1}. In this format, we can separate the data by transition and fit a transition-specific Weibull model to each subset. [6]

We calculate time to each transition, and the confidence and prediction intervals for these, using forward simulation together with bootstrapping [26]. Individual survival times are simulated for patients using estimates from each fitted Weibull model, and iterating through all possible transitions until all patients have reached an absorbing state or are censored at a specified maximum follow-up time. More detail on the method, including equations, is provided in Appendix D.

Results

Overall LoS

Table 1 and Figure 2 show the overall estimated LoS for all three methods. Here, we present results for LoS aggregated across the outcomes of death and discharge, since this can be estimated by all three methods. In Appendix F, we consider the lengths of stay to specific outcomes. The AFT and TC estimates were all based on models adjusted for the week of admission, sex, and age group. The effect of sex was found to be small and non-significant in all of the models[7], thus, we do not present breakdowns by it. Overall, the explanatory power of the predictor variables was only modest. They accounted for a maximum of 10 per cent of the variance in observed LoS in any of the AFT models. MS models were run without adjusting for any predictors.

[6] We estimate the parameter values by using maximum likelihood estimation in Python.

[7] This is an interesting finding; although the severity of COVID is associated with the sex of the patient [27], the length of stay conditioning on severity is not. This has also been found with the CHESS data when controlling for other predictors, see, e.g., [28].

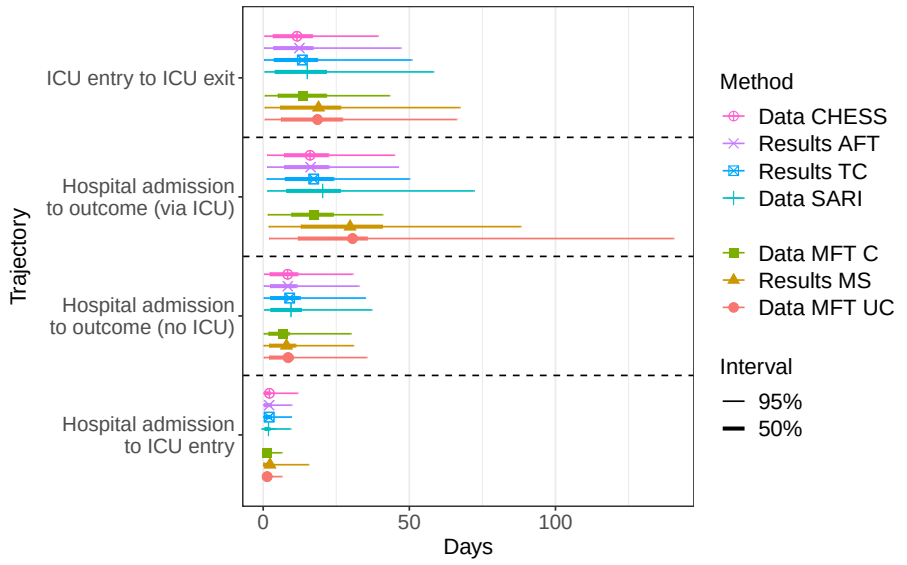


Figure 2 Overall Length of Stay mean estimates with 50% and 95% Predictive Intervals (PI). For CHESS and SARI data, the intervals are based on empirical percentiles. Notes: CHESS denotes data used for predictions as of 26 May 2020; SARI are the data after all patients have had seen their outcomes and missing cases have been added; MFT C denotes data with censoring; MFT UC - without censoring (after all patients have seen the outcome). Source: own elaboration using CHESS and MFT data.

The lack of power in the predictor variables reflects the high individual-level stochasticity of LoS. Trying to predict LoS at an individual level for COVID-19 has been shown to be inaccurate [29]. The highly stochastic dynamics of infectious diseases within host, from magnitude of the initial dose to where the pathogen colonises within host, could drive differences in LoS. Therefore, the majority of variance in observed LoS are driven by the underlying stochastic process rather than explanatory variables. Although the predictors may not explain a large portion of the variance in LoS, they do have a substantial influence on the LoS distributions, with age in particular having a large influence on the expected distribution.

CHESS data for England

For the ICU patients (Hospital to Outcome via ICU), the shape parameters in AFT and TC methods were larger than one, implying the baseline hazard increased over time. For the non-ICU patients and LoS within the ICU, the baseline hazard remains constant in the AFT model and is slowly decreasing in TC, whereas for the Hospital to ICU admission it is decreasing in both models.

Overall, for hospital admission to final outcome, the mean LoS for patients not admitted to ICU was shorter, with an AFT mean of 8.4 (TC mean: 9.1) days, than that of patients who were admitted to ICU at some point, with an AFT mean of 16.2 (TC mean: 17.3) days. ICU admission was estimated to take 2.0 (2.0) days from hospital admission, and ICU patients were estimated to spend an average of 12.4 (13.4) days in ICU.

Standard Deviations (SD) of the estimated LoS are presented in Table 1 whereas Predictive Intervals (PIs) for the LoS in AFT and TC methods are shown in Fig-

Table 1 Overall length of stay estimates for England using the AFT and TC method, and for Manchester trusts using the MS method. Source: own elaboration using CHESS and MFT data. For the multi-state model, the sample size in brackets indicates the observed and censored data (including competing risks), with the first number indicating observed transitions. For TC, for sample size indicates the number of observed transitions, and for AFT the sample size is the number of observed and censored transitions.

Method	Hospital trajectory	Mean	SD	N
TC	Hospital admission to outcome (no ICU)	9.1	9.5	2794
TC	Hospital admission to outcome (via ICU)	17.3	13.1	2517
TC	ICU entry to ICU exit	13.4	13.8	1809
TC	Hospital admission to ICU entry	2.0	2.7	2983
AFT	Hospital admission to outcome (no ICU)	8.4	8.9	2805
AFT	Hospital admission to outcome (via ICU)	16.2	12.0	2555
AFT	ICU entry to ICU exit	12.4	12.8	1809
AFT	Hospital admission to ICU entry	2.0	2.7	2983
Multistate	Hospital admission to outcome (no ICU)	8.0	8.4	620 (786)
Multistate	Hospital admission to outcome (via ICU)	29.7	22.9	73 (101)
Multistate	ICU entry to ICU exit	18.9	18.0	92 (101)
Multistate	Hospital admission to ICU entry	2.3	4.5	101 (786)

ure 2. The standard deviations (SD) for both the AFT and TC models are remarkably similar in depicting the large variability in the observed LoS. With the exception of the LoS from the hospital admission to outcome via ICU, all SD suggest that the waiting times till outcome are approximately exponentially distributed.

MFT data

Similarly to AFT and TC methods, in the MS approach, we used a Weibull distribution for each of the transition times between states in Figure 1. Then, using fitted parameters, we used 1000 bootstraps and 10^3 forward simulations in order to obtain estimates of the mean lengths of stay in each state, given each transition. The MFT data-based results (comparable with trajectories obtained using CHESS dataset with AFT and TC models) are presented in Figure 2 and Table 1, along the summaries of the data.

As with the AFT and TC methods, LoS for patients admitted to ICU is longer, with a mean of 29.7 days, than that of patients not admitted to ICU, with a mean of 8.0 days. ICU admission was estimated to take 2.3 days from hospital admission and ICU patients were estimated to spend an average of 18.9 days in critical care. Taking into consideration competing hazards between stepdown and death, our mean LoS estimate for a patient in ICU is between 15.8 and 20.1 days (Table A1 in Appendix F), though in the data we observe people that have much longer critical care periods (20% of patients have over 40 days in an ICU).

Planning with LoS

Figure 3 predicts bed occupancy in acute ward and ICU after running our simulator with the parameter estimates of all three methods. The red and blue lines represent the implementation of, and relaxation of mass quarantine (or "lockdown"), respectively. These are considered to change the shape of the admissions trajectory to reflect that observed. We simulate hospital admissions from 23 February, first

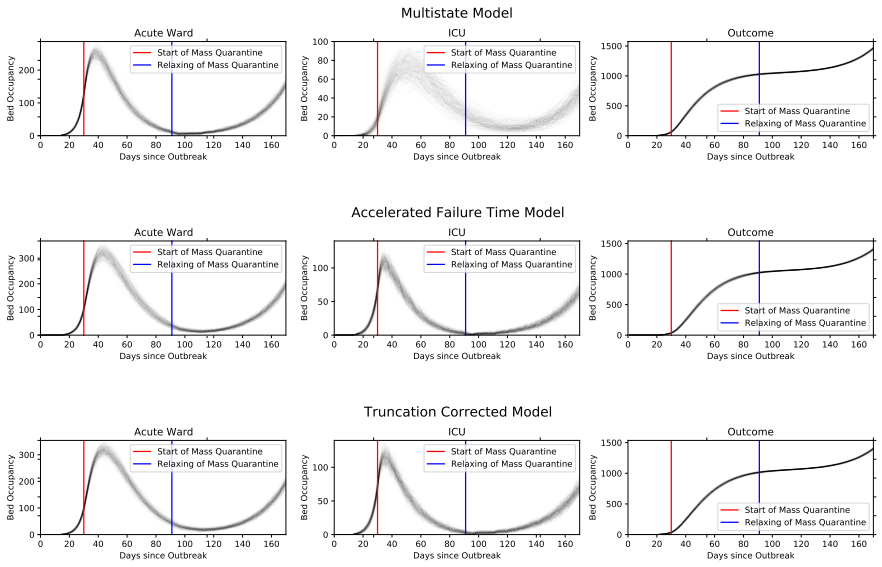


Figure 3 Output of our simulation for transition parameters estimated using each of our three methods, starting from 23 February, which we take to be the start of the outbreak in the UK. Source: own elaboration using CHESS and MFT data.

assuming exponential growth with a doubling time of 3 days, followed by exponential decay shortly after the implementation of mass quarantine. Following the blue line, we plan for a reasonable worst case scenario, and so assume a slower growth in cases with a doubling time of 15 days. Changing the assumptions used to generate hospital admissions allows us to predict and plan for any scenario of interest.

In the MS model, the hazard functions account for the competing risks of different pathways and outcomes. Therefore, hospital occupancy can be obtained by simulating the hazard functions and following the shortest transitions. In the TC and AFT models, the hazard functions are conditional on pathways and outcomes. Therefore, to simulate hospital occupancy these hazard functions need to be coupled with probabilities of each pathway. With the aggregated outcomes considered in this article, the only competing risk is whether a patient goes to ICU or not. From the MS model, the ICU admission probability is approximately 13%, so we assume the transition probability of 13% for going from the acute ward to ICU. Hospital occupancy can be then obtained by simulating the ICU probability combined with the conditional hazard functions. See Appendix E for more details.

The estimates from the AFT and TC methods yield similar predictions of bed occupancy and total observed outcomes. The MS model also gives similar predictions for acute ward and outcome but differs for ICU. The peak in bed occupancy in ICU in the MS output occurs roughly two weeks later than in the AFT and TC model outputs, and there is a slower decline after the peak. This is caused by the larger LoS estimates for the MS models as seen in Table 1 and Figure 2.

The effect of predictors – England

In Figure 4 and Table 2, we present the estimates of LoS broken down by two main predictors: age and week of admission. The mean waiting time from hospital admission to ICU entry (first column of Figure 4) is around two days irrespective of age. For hospital admission to outcome without ICU stay (second column of Figure 4), increasing age raises the length of stay, with length of stay around five days for the youngest age group and twelve days for the oldest, irrespective of the AFT or TC model. For individuals who go via ICU (third column of Figure 4), the pattern with age is less clear [8]. For the first three age groups, the length of stay is roughly similar (especially AFT model), with a slight decrease in the oldest age group with respect to the first two. The 75+ age group, however, has a much shorter length of stay. A similar pattern is observed for mean LoS from ICU admission to ICU exit (fourth column of Figure 4).

Considering the week of admission as a predictor variable, there is less variability in LoS than in LoS disaggregated by age. For a great majority of hospital trajectories, the mean LoS seems to have decreased by, on average, 16 per cent, depending on the age group and the method used. This could be explained by potential behavioural changes in the later admission weeks. Firstly, after mass quarantine progressed individuals may have waited longer before presenting at hospital. Secondly, treatment policy has changed over the course of the outbreak, with the criteria for discharge being relaxed to ensure hospitals had capacity. Nonetheless, we also note large variability in predicted LoS both in earlier and later weeks under study.

Model validation

Analysis of the Cox-Snell and deviance residuals for individual patients for the AFT models, in which these are well-defined [30], showed good model fit and little evidence of bias for three of the models (although there was less precision for the right-hand tails of the LoS distributions, where the effective sample size was smaller because of earlier deaths, discharges, and censoring). The exception was the model from hospital admission to ICU entry. This observed distribution was very skewed (median LoS was 0.7 days, with 20 percent over 3 days). The choice of Weibull error distribution did not represent this well, and the model showed bias in predicted LoS.

To evaluate how well the models compensated for censoring we compared the model estimated mean LoS with the data that were available during the original data collection window (i.e. including censored cases) and also the fully observed, uncensored data, which was eventually available in 2021 when all patients had left hospital. The LoS summaries based on fully observed data are presented in Figure 2 and denoted as 'Data SARI' (the updated CHESS dataset), and 'MFT UC' (the uncensored MFT dataset). These data correct on the data used in the original analysis in three ways. Firstly, the right-censored data available at the time have been uncensored (except for a negligible proportion of patients, who we remove from the final sample). Secondly, data have been retrospectively corrected. In the

[8] This is likely to be caused by the LoS to death (via ICU) following the opposite pattern to the LoS to discharge (via ICU). Younger age groups appear to have a longer duration until death on ICU and a shorter duration before discharge. These patterns seemingly cancel out when looking at the LoS until any outcome. This analysis is not shown here since we are focusing on length of stay in hospital rather than different outcomes.

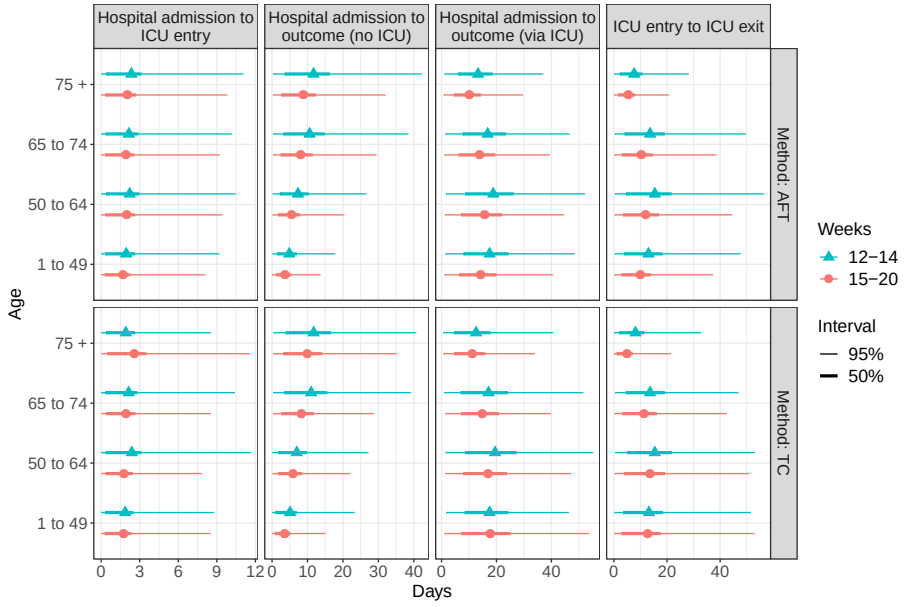


Figure 4 Mean Length of Stay by age and week of admission with 50% and 95% Predictive Intervals (PI). Source: own elaboration using CHESS data for England.

original analysis, we removed the last week of data to reduce the effect of data corrections, but there could still be potential revisions. Thirdly, new patients have been added to the CHESS/SARI data. In this comparison, we only used patients admitted before 17 May 2020, to be consistent with the original data. However, in the original CHESS data, after processing this left 6208 patients, whereas in the uncensored SARI data we have 13800 patients admitted before this date. Therefore, in this validation, we investigated how the models simultaneously deal with the right-censoring, errors in the data, and case missingness of patient records.

Table A2 (see Appendix G) shows that the mean LoS from hospital admission to final outcome for patients who went into ICU at some point was on average underestimated by over five days in the original data compared to the fully observed data, and mean LoS in ICU was underestimated by over 2 days. The TC model was able to compensate for about a quarter of the underestimate in LoS for the former, and over 70 percent of the underestimate for the latter. The AFT model made smaller adjustments to the observed LoS and so captured less of the underestimate. In the original CHESS data set, we had data from 16 March 2020 to 17 May 2020, so the maximum LoS included could be 62 days. In the uncensored data, the maximum observed LoS was 245. Therefore, although the models attempted to adjust for the truncated/censored tail observations, there was insufficient data on the true extent of the tail to make the full adjustment. This illustrates how challenging it can be to estimate LoS during an emerging epidemic, even with large volumes of data.

Both TC and AFT models performed poorly for the LoS from hospital admission to ICU entry, underestimating LoS even more than the original, censored data. This is perhaps due to the Weibull distribution being inappropriate for this length of stay, and therefore struggling to capture the long tail.

Table 2 Length of stay estimates with predictor variables for AFT and TC methods. Sample sizes differ due to the inclusion of censored observations in the AFT method. Source: own elaboration using CHESS data for England.

			AFT model			TC model		
Trajectory	Age	Weeks	Mean	SD	N	Mean	SD	N
Hospital admission to outcome (no ICU)	1 to 49	12 to 14	4.9	4.8	146	5.1	6.5	146
		15 to 20	3.7	3.6	210	3.6	4.1	210
	50 to 64	12 to 14	7.3	7.2	223	7.0	7.4	223
		15 to 20	5.6	5.4	304	5.9	5.9	304
	65 to 74	12 to 14	10.6	10.4	204	11.0	10.5	204
		15 to 20	8.1	7.9	270	8.3	7.7	266
	75 +	12 to 14	11.7	11.4	609	11.7	10.9	607
		15 to 20	8.8	8.6	839	10.0	9.4	834
Hospital admission to outcome (via ICU)	1 to 49	12 to 14	17.5	12.5	312	17.5	11.8	312
		15 to 20	14.3	10.5	267	17.8	14.0	262
	50 to 64	12 to 14	18.8	13.4	641	19.5	14.3	626
		15 to 20	15.7	11.5	467	17.0	12.1	455
	65 to 74	12 to 14	16.8	12.0	391	17.1	13.5	388
		15 to 20	13.9	10.2	225	14.9	10.2	223
	75 +	12 to 14	13.3	9.5	161	12.6	10.8	161
		15 to 20	10.2	7.6	91	11.3	8.9	90
ICU entry to ICU exit	1 to 49	12 to 14	13.0	12.8	239	13.2	14.0	239
		15 to 20	10.0	10.0	210	12.7	14.5	210
	50 to 64	12 to 14	15.4	15.2	468	15.4	14.2	468
		15 to 20	12.0	12.0	337	13.6	13.8	337
	65 to 74	12 to 14	13.6	13.4	237	13.6	12.5	237
		15 to 20	10.4	10.4	152	11.4	11.4	152
	75 +	12 to 14	7.6	7.5	109	8.1	8.9	109
		15 to 20	5.5	5.6	57	5.0	5.9	57
Hospital admission to ICU entry	1 to 49	12 to 14	2.0	2.6	340	1.9	2.5	340
		15 to 20	1.7	2.3	336	1.8	2.4	336
	50 to 64	12 to 14	2.2	2.9	732	2.4	3.3	732
		15 to 20	2.0	2.7	610	1.8	2.2	610
	65 to 74	12 to 14	2.2	2.9	421	2.1	2.9	421
		15 to 20	1.9	2.6	276	1.9	2.4	276
	75 +	12 to 14	2.4	3.1	168	1.9	2.4	168
		15 to 20	2.1	2.8	100	2.6	3.2	100

The Multi-state model, on the other hand, performed well at estimating LoS for each transition (Table A2). This is in part due to local data from MFT exhibiting fewer biases than the national CHESS data so that a trust-specific LoS can be estimated with greater accuracy. The performance of the MS model can also perhaps be explained by the fact that it fully takes into account the competing risks at each transition. Of all of the LoS considered in Table 1, the maximum absolute difference between the final LoS observed in the uncensored MFT data and our estimate from the MS model is 0.98 days (from hospital admission to ICU entry), so that all of

our estimates are within 1 day of the true, observed values. Again, this transition is potentially not well-captured using a Weibull hazard function.

Discussion

Analysis of results

Comparison of the three different models

In this study, we have presented three methods for estimating the LoS of patients with COVID-19 infection. Overall, the AFT and TC methods produced similar estimates for LoS for all four hospital trajectories. This is reassuring and forms an effective cross-validation of both methods and results.

The estimated mean LoS from the AFT model are shorter by around one day than the TC means, except for the Hospital-ICU entry. This might be due to the exclusion of potentially censored cases in the AFT method[9], since it was not clear these were genuinely censored or incomplete data entries. Both methods also yielded similar predictive uncertainty about the LoS, with TC producing slightly wider predictive intervals than the AFT method. This might be explained by the explicit inclusion of the predictors in the AFT model with a joint assessment of their effect on the LoS. The TC method assumes independence between predictors and is applied to the subsets of CHESS data disaggregated by the predictor categories.

There were large differences in predicted ICU LoS between the two CHESS based methods and the MS method. The mean estimates derived using AFT and TC methods (12-13 days) were 5-6 days less than those from the MS method. The predictive intervals overlap suggesting the variability in LoS is large. However, given the focus of the paper is on comparison, and bearing in mind the MFT data is an effective census of the MFT patients and therefore that estimates are reliable in terms of the mapping of the data to the population, it is valuable to consider possible explanations for the differences in the point estimates.

These differences may reflect several substantive factors. First, MFT is one of five adult centres in the UK to have an extracorporeal membrane oxygenation (ECMO) unit. Combined with expertise in specialist respiratory care, MFT takes referrals for severe COVID-19 cases requiring ECMO treatment from other hospital trusts in the UK's North West and Midlands regions. This higher proportion of severe cases could contribute to the longer, on average, lengths of stay observed at MFT. Unfortunately, referrals and ECMO cases cannot be separated from the MFT data, so we were unable to account for this in our analysis.

Second, the underlying data were different: the AFT and TC models used the country-wide but very incomplete CHESS data, whereas the multi-state model was based on data from just one NHS trust, but largely free of missing data. There is potentially large heterogeneity between LoS at different trusts, so data at a single trust may not reflect the national data.

Third, differences in excess bed demand from trust to trust potentially further explain discrepancies in our estimates. For trusts experiencing significant increases in demand, it is possible that they do not have the ability or resources to accurately

[9] For the ICU entry to ICU exit trajectory there were $n = 108$ censored cases included in the model. For the trajectory from hospital admission to outcome via ICU: $n = 43$, and without ICU: $n = 14$. There were no censored cases in hospital admission to ICU entry.

generate daily CHESS reports which are collected in addition to routinely collected data (see Appendix A). This partially explains the case-missingness in the CHESS data.

In order to check sensitivity of the findings for the differences in the data, we evaluated the AFT model and TC method using CHESS data for Manchester University NHS Foundation Trust only. MFT contributed 53 cases with recorded LoS in ICU to CHESS. Running the AFT model on these cases gave a predicted ICU LoS of 16.5 days (SD=17.3). For the TC method, the predicted mean was 16.1 (SD= 16.7).The estimated LoS were longer than the full-sample CHESS estimates but still shorter than the predicted LoS from the MS models (18.9 days). In the MFT data, 83 cases are included. This discrepancy between the data sets could be contributing to the difference between the MS model and the AFT and TC models. Additionally, when evaluating model performance, the MS model appears to better account for the right-censoring, which could be further contributing to this discrepancy.

All methods captured the variability in the data and reflected it in the predictive distributions. This uncertainty should be taken into account when planning for the number of beds during the pandemic. For example, upper bounds of the predictive intervals can be used to construct extreme-case scenarios for the beds occupancy. These can be fed into the multi-state model to predict the number of patients in hospital at various stages of the pandemic (Figure 3).

In the main LoS analysis above, we did not distinguish between different outcomes, such as death or discharge. Particularly in ICU, the baseline hazards for these competing hazards may be strongly diverging over time. In Appendix F, we analyse the length of stay for given outcomes using the TC and MS methods, finding that in general the length of stay to discharge is longer than to death.

Evaluation of model performance

When evaluating the performance of the three methods at accounting for the right-censoring, we observe different levels of performance across the methods. Using the CHESS data, the AFT model struggles to appropriately adjust for the right-censoring, resulting in an underestimate of the true distribution. The TC model does a better job at accounting for this, but still slightly underestimates the LoS. The TC model struggles to capture the true LoS because this method requires sufficient tail observations in order to adjust for the truncation bias. However, in the uncensored data there are some tail observations over twice the length of the maximum possible LoS included in the original analysis data. Therefore, the TC method does not have sufficient information to construct the true tail of the LoS distributions. The AFT model is also affected by this issue. A further complication with the AFT model is the challenge with censoring in the CHESS data. With high levels of data missingness and incompleteness at the time of the analysis, it was unclear whether cases were genuinely censored or had failed to be updated. This resulted in many censored cases being omitted from the analysis data set, leading to further underestimation of the LoS. Using the MFT data, the MS model captures the true LoS much more accurately. This model uses higher quality data, so can appropriately adjust for the censoring and the competing risks of different hospital pathways. Therefore, provided sufficiently high quality data are available,

the MS approach is superior for estimating LoS during an epidemic. However, such high quality data may not be available early in a pandemic, particularly in smaller trusts. The CHESS data are not well suited for such analysis, due to the unclear case inclusion biases. This may affect the proportion of cases entering each pathway, which can interfere with the competing risks aspect of the MS model.

Limitations of research

The CHESS dataset suffers from large amounts of case-missingness, which may lead to bias in the estimates. There appear to be three types of this. **Update delay** where a record has not been updated (with a transition) which may lead to incorrect censoring. This leads to the patient being removed from the data for some of the models. **Reporting delay** where a patient does not appear in the data at all until sometime after their admission. **Non-reporting** where no report is ever made on a patient. All three of these may cause bias in the models if they are correlated with either LoS or with extraneous variables (that are not controlled for within a given model). Another limitation of both datasets was that only cases of COVID-19 infection that led to hospital admission were included in the data. During March 2020, the hospitalised patients in England were considered to reflect the underlying population of patients with severe COVID-19 infection, but by 14 April, care-home deaths reported on death certificates caused a revision of views [31]. Those severe cases not attending hospital and COVID-19-related deaths outside of hospital may have different properties from hospitalised patients and deaths. So care should be taken in extrapolating the findings to general statements about disease progression outside of the hospital setting. Given that the goal here was to model length of stay in hospital this is less of a concern. However, change in hospitalisation practice could lead to changes in the estimates that the models produce.

Our models were also limited by the missing values in the CHESS data. A notable limitation was that around half of the cases did not have their final outcome or current status recorded. We did not know if this implied that the patient was still in hospital or whether it was an omission or whether this was a result of update delay. In either case, we had no reliable way to estimate the last time point at which the patient was observed to be in hospital, and thus these patients could not contribute to the LoS estimates. The fact that the CHESS-based LoS estimated by using the AFT models were not adjusted sufficiently to capture this suggests that many such patients were indeed still in hospital.

Compared with the AFT model, the TC method should, in theory, be less sensitive to this issue since it ignores censored cases. However, this method relies on sufficient tail observations being recorded. With the long duration of this study (over 60 days), one might expect sufficient tail observations to be included. However, with the very long lengths of stay observed in the uncensored SARI data (over 200 days), it is apparent that the original censored sample did not contain enough information on the tail of the distributions. Further complications are caused by non-random case missingness. For example, omitted cases might correspond disproportionately to tail observations, which would cause the truncation corrected method to underestimate LoS.

The strength of bias due to the truncation and censoring varies depending on the phase of the epidemic, with it having a large impact during exponential growth and

lessening impact during the decay phase. The data used in this analysis is from the decay phase, so the truncation bias does not have a huge impact, and ignoring this bias would underestimate LoS by up to two days (using TC method). However, for a sample earlier in the outbreak, this underestimation may be amplified, as well as the difference in model performance. This is also true for censoring biases, since early in the outbreak the majority of cases will have censored outcomes. A large number of right-censored cases would lead to relatively large values of LoS when using the AFT model. For the purposes of this paper, we have opted not to investigate the performance of each model at different sampling dates. This is to focus on the presentation of the different methods using a single illustrative example to improve clarity. Future research could extend this in several ways, including running iteratively through the data available on different dates, modelling the impact of truncation, censoring, reporting and updating lags as the epidemic progresses.

Another issue is that clustering of patients within the NHS trusts, which were at different stages of the epidemic at different times leading to variations in pressure on capacity, could mean that there are spatial-temporal interactions in the processes driving LoS which are not captured in the models. Further, these may in turn interact with the data generating processes for CHESS with more non-reporting and reporting delays likely during high demand times. These issues could have unpredictable effects on the estimates of LoS.

With respect to the MFT data, most limitations arise due to the small absolute sample size. The multi-state method requires seeing an adequate number of patients for each state transition before any reliable modelling can take place. Indeed, although it is clinically known that stepdown to mortality is a valid transition, after applying our exclusion criteria, there were no observations of this transition occurring for patients with COVID-19 infection within this Manchester Trust. The analysis conducted in this paper therefore excluded this transition, and it is not possible to see how this influences overall hospital LoS of those patients who have an ICU episode during their hospitalisation. Together with uncommonly long ICU periods, the relative delay in the Manchester epidemic compared to other parts of the country means that MFT patients with long critical care spells are either still in ICU or only just starting to move onto stepdown. Given more weeks of data, we might be able to include stepdown to mortality in our model.

The above suggests differences between the estimates of LoS for the two datasets may therefore be due more to differences in the available data than differences in the statistical methods *per se*. It is important to acknowledge these uncertainties in the data when interpreting length of stay estimates. We further note that not only would we obtain more power in predictions through a larger amount of complete data, but also a better understanding of how the complex interactions between the virus and background risk factors affect disease severity. Additionally, inclusion criteria are slightly different between the CHESS/SARI and MFT data sets. In the CHESS/SARI data set, there is a column which indicates whether the admission was due to COVID-19. However, there is no clear definition for this, so individual hospital trusts could use different cutoff criteria, such as positive on admission or showing clear signs of COVID-19 pneumonitus. For the MFT data, we defined our own inclusion criteria, including all patients with a positive test 2 days either side

of admission. At the time of the analysis (March 2020 to May 2020), there was some admissions screening at MFT, but not as widespread as the current (April 2021) requirements. Therefore, the majority of patients captured through this definition are likely to be symptomatic COVID-19 patients requiring acute care for COVID-19, rather than general admissions who return a positive swab. In both data sets we do not consider nosocomial COVID-19 cases.

Conclusions

In this paper and its supporting materials, we provide a freely accessible set of models and tools to estimate LoS with an application to patients with COVID-19 infection. Together with a prediction of hospital admissions, which depends on the severity of outbreaks in the local area, LoS predictions can be implemented to provide organisational support within hospitals to ensure the demand for hospital and, in particular, ventilated ICU beds does not exceed availability. The models use routinely collected hospital data which are available within many national healthcare systems. Thus we anticipate our approaches will have utility across diverse healthcare systems in many different countries.

Abbreviations
AFT: Accelerated Failure Time; CHESS: COVID-19 Hospitalisation in England Surveillance System; ECMO: Extracorporeal Membrane Oxygenation; ICU: Intensive Care Unit; LoS: Length of Stay; MFT: Manchester University NHS Foundation Trust; MS: Multi-State; PAS: Patient Administration System; PI: Predictive Intervals; SD: Standard Deviation; TC: Truncation Corrected

Declarations
Ethics approval and consent to participate
The University of Manchester Ethics Decision Tool ruled that no formal ethics approval was required for this particular study, and it would be unethical to use the time of an ethics committee to confirm this, given that the project involved secondary use of data subject that were functionally anonymised, and that the data were being used for the purposes that were compatible with the original consent of the data subjects with full permission of the data owner.

Consent for publication
Not applicable.

Availability of data and material
The data involved in this work is sensitive and therefore not publicly available. Code and parameter estimates are publicly available at `https://github.com/thomasallanhouse/covid19-los`.

Competing interests
The authors declare that they have no competing interests.

Funding
TH receives support from the Royal Society (Grant Number INF/R2/180067) and the Alan Turing Institute for Data Science and Artificial Intelligence. LP and CO are funded by the Wellcome Trust and the Royal Society (Grant Number 202562/Z/16/Z). BV is funded by an MRC Doctoral Partnership grant (Grant Number MR/R502236/1). AW receives support from the EPSRC (Grant Number EP/V027468/1). AAC is funded by the ESRC through the North West Social Science Doctoral Training Partnership (Grant Number ES/P000665/1). The funding bodies mentioned in this section had no role in the study design, data analysis, data interpretation, and writing of the manuscript.

Authors' contributions
BV, CO, AW, TH, ME, JCS and NS had roles in the study design, data analysis, data interpretation, literature search, and writing of the manuscript. AAC, JK, WO, DPR, MP, LP, and had roles in data analysis, data interpretation and study design. SA and JE provided clinical expertise and information. JS, JE, NH assisted with data collection and study design. All authors reviewed and commented on at least one draft of the paper and approved the final manuscript.

Acknowledgements
Not applicable.

Author details
[1] Department of Social Statistics, School of Social Sciences, University of Manchester, Oxford Road, M13 9PL Manchester, UK. [2] Department of Mathematics, University of Manchester, Oxford Road, M13 9PL Manchester, UK. [3] Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Oxford Road, M13 9PL Manchester, UK. [4] IBM Research, Hartree Centre, Daresbury, UK. [5] Division of Diabetes, Endocrinology & Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine & Health, University of Manchester, Oxford Road, M13 9PT Manchester, UK. [6] Clinical Data Science Unit, Manchester University NHS Foundation Trust, Oxford Road, M13 9WU Manchester, UK. [7] Department of Mathematics, University of Liverpool, Peach Street, L69 7ZL Liverpool, UK. [8] Department of Virology, Manchester Medical Microbiology Partnership, Manchester Foundation Trust, Manchester Academic Health Sciences Centre, Oxford Road, M13 9WU Manchester, UK.

References
1. UK Government. Slides and datasets to accompany coronavirus press conferences [Webpage]; 2020. Available from: `https://www.gov.uk/government/publications/slides-and-datasets-to-accompany-coronavirus-press-conference-29-may-2020`. Accessed 04/06/2020.
2. Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LH, Lythgoe KA, et al. Challenges in control of Covid-19: short doubling time and long delay to effect of interventions. arXiv preprint arXiv:200400117. 2020;.
3. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of travel medicine. 2020;.
4. Mahase E. China coronavirus: what do we know so far? BMJ. 2020;368. Available from: `https://www.bmj.com/content/368/bmj.m308`.
5. Sridhar D, Majumder MS. Modelling the pandemic. BMJ. 2020;369. Available from: `https://www.bmj.com/content/369/bmj.m1567`.

6. The Health Foundation. COVID-19 Policy Tracker [Webpage]; 2020. Available from: https://www.health.org.uk/news-and-comment/charts-and-infographics/covid-19-policy-tracker. Accessed 04/06/2020.
7. Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LHK, Lythgoe KA, et al. Challenges in control of Covid-19: short doubling time and long delay to effect of interventions. medRxiv. 2020;.
8. Rees EM, Nightingale ES, Jafari Y, Waterlow NR, Clifford S, Pearson CA, et al. COVID-19 length of hospital stay: a systematic review and data synthesis. BMC medicine. 2020;18(1):1–22.
9. Ferstad JO, Gu AJ, Lee RY, Thapa I, Shin AY, Salomon JA, et al. A model to forecast regional demand for COVID-19 related hospital beds. medRxiv. 2020;.
10. Weissman GE, Crane-Droesch A, Chivers C, Luong T, Hanish A, Levy MZ, et al. Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic. Annals of internal medicine. 2020;.
11. Massonnaud C, Roux J, Crépey P. COVID-19: Forecasting short term hospital needs in France. medRxiv. 2020;.
12. Sá C, Dismuke CE, Guimarães P. Survival analysis and competing risk models of hospital length of stay and discharge destination: the effect of distributional assumptions. Health Services and Outcomes Research Methodology. 2007;7(3-4):109–124.
13. Kalbfleisch J, Lawless J. Regression models for right truncated data with applications to AIDS incubation times and reporting lags. Statistica Sinica. 1991;p. 19–32.
14. Sun J. Empirical estimation of a distribution function with truncated and doubly interval-censored data and its application to AIDS studies. Biometrics. 1995;p. 1096–1104.
15. Becker NG, Britton T. Statistical studies of infectious disease incidence. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1999;61(2):287–307.
16. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological). 1972;34(2):187–202.
17. Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. Statistics in medicine. 1992;11(14-15):1871–1879.
18. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. vol. 360. John Wiley & Sons; 2011.
19. Crowther MJ, Lambert PC. Parametric multistate survival models: flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences. Statistics in medicine. 2017;36(29):4719–4742.
20. Carroll KJ. On the use and utility of the Weibull model in the analysis of survival data. Controlled clinical trials. 2003;24(6):682–701.
21. Plummer M, et al. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing. vol. 124. Vienna, Austria; 2003. p. 125.
22. Kaminskiy MP, Krivtsov VV. A simple procedure for Bayesian estimation of the Weibull distribution. IEEE Transactions on Reliability. 2005;54(4):612–616.
23. Bogaerts K, Komarek A, Lesaffre E. Survival analysis with interval-censored data: A practical approach with examples in R, SAS, and BUGS. Boca Raton, FL: CRC Press; 2017.
24. Andersen PK, Abildstrom SZ, Rosthøj S. Competing risks as a multi-state model. Statistical methods in medical research. 2002;11(2):203–215.
25. Andersen PK, Keiding N. Multi-state models for event history analysis. Statistical methods in medical research. 2002;11(2):91–115.
26. Davison AC, Hinkley DV. Bootstrap Methods and their Application. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press; 1997.
27. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. Nature. 2020;584:430–436.
28. Shryane N, Pampaka M, Aparicio Castro AL, Ahmad S, Elliot M, Kim JH, et al. Length of Stay in ICU of Covid-19 Patients in England, March-May 2020. International Journal of Population Data Science. 2020;5(4).
29. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. bmj. 2020;369.
30. Collett D. Modelling survival data in medical research. Springer; 1994.
31. Office for National Statistics. Deaths registered weekly in England and Wales, provisional: week ending 3 April 2020, Release date 14 April 2020. [Webpage]; 2020. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending3april2020. Accessed 01/06/2020.
32. Putter H. Tutorial in biostatistics: Competing risks and multi-state models Analyses using the mstate package. Companion file for the mstate package. 2011;.

Appendix A: Data generating process

The data that we analyse from MFT is routinely collected by the hospital administration teams. Details from doctors' notes and patient admissions are entered into the Trust's Patient Administration System (PAS). Patient data from PAS and WardWatcher are then aggregated together to determine an individual's entire hospitalisation pathway. We further make use of data collected by MFT on testing for COVID-19; this is, again, collected and entered manually into a database called Telepath, which is subsequently joined to the main Trust database by the Trust's data warehouse.

In addition to this routinely collected information, trusts have also been required by Public Health England (PHE) to report individual-level data on patients receiving care for acute respiratory infection and aggregate data on all COVID-19 admissions for CHESS. This information is submitted by 09:00 each day with data corresponding to the previous day. This data is compiled manually, requiring additional input from administrative staff to ensure that the data is sent on time and with the correct information. Information sent to PHE by individual trusts is then compiled into a dataset that is disseminated weekly to trusts and reported on weekly to NHS England.

Appendix B: Value-missingness in CHESS data

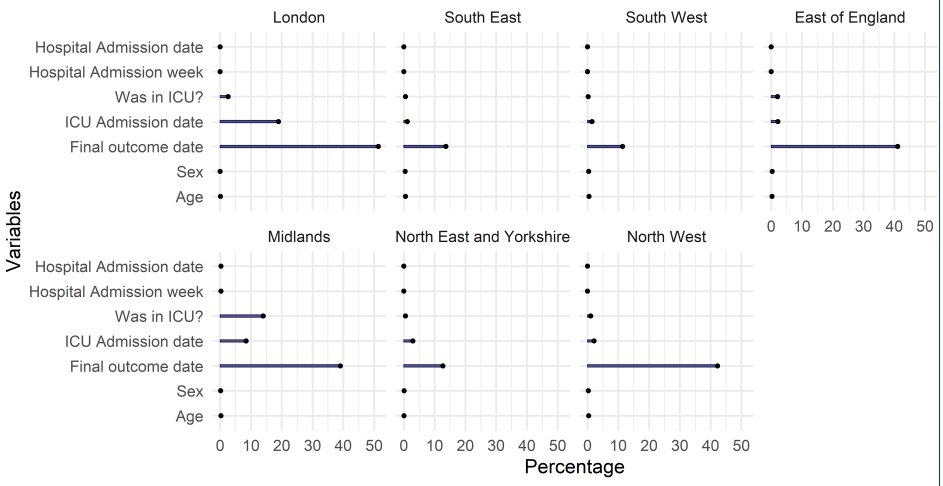


Figure A1 Percentage of values that are missing or recorded as unknown for the variables of the raw CHESS data that are used in this paper (Date and week of Hospital Admission, Was in ICU?, ICU Admission date - conditional on the patient being admitted to ICU, Final outcome date, Sex, and Age). Source: own elaboration using CHESS data for England.

Figure A1 presents the percentage of missing values in the raw CHESS data reported by NHS trusts grouped by region. London followed by Midlands have the highest percentages of missing values, while South West obtains the smallest one.

In terms of the variables, final outcome date variable has 33.6% of missing values. These will be split into cases where the missingness is because the final outcome has not yet happened and those where it has happened but has not been captured.

ICU admission date conditioned on whether the patient was admitted to ICU has 5.36% missing values. Age has 0.29% missing values, and date and week of hospital admission only have 0.06%.

By contrast, sex and the variable regarding whether a patient was in ICU or not do not have missing values. However, these do have some recorded values of "unknown" which we interpret as missing. Sex has 0.23% of unknown values, whereas the item regarding being admitted to ICU has 4.41%.

Appendix C: Data processing

We use CHESS data released on 26 May 2020 ($N = 16,138$). We first filter out 493 duplicated cases. The de-duplication rule set as follows.

Rule	0
number of records	1
IDS are	singular
Date/time of admission to ICU	Any
Other variables	Any values
Action	leave unchanged

Rule	1
number of records	2
IDS are	Identical
Date/time of admission to ICU	Identical values for dateadmittedicu
Other variables	different values for hospitaladmissiondate
Action	Include record with the earliest hospitaladmissiondate, delete the others

Rule	2
number of records	2
IDS are	Identical
Date/time of admission to ICU	Identical values for dateadmittedicu
Other variables	identical values for hospitaladmissiondate
Action	Include record with the earliest sbdate, delete the others

Rule	3
number of records	2
IDS are	Identical
Date/time of admission to ICU	Different values for dateadmittedicu
Other variables	ICU periods are non contiguous
Action	Leave all records in the file but record a different obsid for each record

Rule	4
number of records	2
IDS are	Identical
Date/time of admission to ICU	Different values for dateadmittedicu
Other variables	ICU periods are contiguous
Action	Merge the records to a single record which has the earliest hospitaladmissiondate and dateadmittedicu and the latest dateleavingicu

Rule	5
number of records	2
IDS are	Identical
Date/time of admission to ICU	One of them doesn't have dateadmittedicu
Other variables	
Action	Include record with dateadmittedicu, delete the others

Rule	6
number of records	2
IDS are	Identical
Date/time of admission to ICU	Neither has dateadmittedicu
Other variables	different values for hospitaladmissiondate
Action	Include record with the earliest hospitaladmissiondate, delete the others

Rule	7
number of records	2
IDS are	Identical
Date/time of admission to ICU	Neither has dateadmittedicu
Other variables	identical values for hospitaladmissiondate
Action	Include record with the earliest sbdate, delete the others

Rule	8
number of records	2
IDS are	Different
Date/time of admission to ICU	Identical values for dateadmittedicu, hoursadmittedicu and minutesadmittedicu (and those values are not missing).
Other variables	identical vales for: sex, ageyear, agemonth, hospitaladmissiondate, trustcode, postcode
Action	Use the record with the earliest estimateddateonset or infectionswabdate, delete the others

Rule	9
number of records	4
IDS are	Identical
Date/time of admission to ICU	identical values for dateadmittedicu for a and b, identical values for c and d and a $<$c
Other variables	a and c have identical values for hospitaladmissiondate as do b and d and a $<$b
Action	delete b and d and run the rule set over a and c

Rule	10
number of records	>1
IDS are	Identical
Date/time of admission to ICU	Any not meeting the above conditions
Other variables	
Action	mark for clerical review

The rules are applied in numerical order. In the small number of cases where rule 10 applies then this will always be where two or more of the rules need to be applied in combination.
We only analyse those patients whose records make explicit that the admission to the hospital unit was due to COVID-19. We make this assumption to exclude nosocomial cases, for whom the LoS begins when they were admitted to hospital for non-COVID-19 reasons. It does not make sense to compare these cases with LoS from COVID-19 hospitalisations. Thus, from $15,645$ deduplicated cases, $8,938$ entries were excluded.
Furthermore, we only analysed cases who were admitted to hospital from 16 March 2020 to 17 May 2020 (i.e. from week 12 to 20). Data before week 12 was omitted as this sample size was small and the treatment policy was different from that in more recent data, with patients having very long lengths of stay early in the outbreak. Data after week 20 was omitted as there are often corrections to historical data from the last week or so, so we could not treat the most recent data as reliable. This removed 317 additional cases.

Finally the omission of cases due to unknown *sex*, and negative values in *age* or recorded age of zero, and unknown (effectively missing) information regarding whether a patient was admitted to ICU or not led to a dataset of $6,208$ records. The removal of the records that have an age of 0 recorded needs further explanation. The number of records with age 0 was 725 (which compares to 15 age 1's). Many such records had characteristics that one would not attribute to newborns (e.g. obesity, smoking, diabetes, ulcers etc.). We also note that some cases with ages recorded as 0 in early versions of the CHESS dataset had been updated with non-zero ages by 26 May. It seems likely that the data entry system for CHESS has a default setting of "today" for the DOB and therefore in effect the vast majority of Age 0's were in fact cases where the age/DOB was not available when the data were entered. Hence removing these cases seems prudent.

Some LoS have zero length, where patients enter and leave ICU on the same day and only have to the day of arrival and departure recorded not the time. For such cases, we assumed the outcome occurred half a day after admission, since instantaneous durations are unrealistic. Half a day was chosen so that these cases were not biased to either side of their recorded data. Some cases recorded hourly data for some events but not all, causing some LoS to be in $(-1, 0)$. For them, we also adjust the outcome date to half a day after admission. All patients with LoS in $(-\infty, -1]$ were discarded. In total 849 cases had their ICU admission date changed to half a day after hospital admission, 41 cases had the ICU discharge date changed to half a day after ICU admission and 199 cases had final outcome date changed to half a day after hospital admission. Therefore, the choice of how to adjust this data has the largest impact on the hospital admission to ICU length of stay, where this constitutes 28% of cases. For this length of stay, moving the adjustment to different extremes (either adding 0.1 or 0.9) changes the length of stay estimates by no more than one tenth of a day. Therefore, this data processing method does not have a substantial impact on the LoS estimates.

Appendix D: Multi-state survival analysis

Here we present the details of the multi-state survival model used in our analysis. Suppose an individual is in state u at time t, then the move that an individual makes to their next state v is governed by a set of transition intensities $q_{uv}(t) = \lim_{\delta t \to 0} \Pr(S(t + \delta t) = v \mid S(t) = u)/\delta t$. The intensity represents the the instantaneous rate of transition from state u to state v.

Data structure and transition-specific parametric models

Given the granularity of routinely collected data in hospitals, all transition times between states are observed exactly, with no additional transitions between observation times. Such data allows us to efficiently model the transition intensities parametrically, which we show here with the use of a Weibull accelerated failure time (AFT) model.

It is important to note that the data must first be structured in a specific way. In contrast to standard survival analysis, in the multi-state case, we now have a series of event times $t_1, \ldots, t_n$ for each individual, corresponding to each change in state. The last of these, may be observed so that the patient has entered an absorbing state i.e. they are discharged or dead, or right censored if the patient is still in the hospital. When there are n_u competing events for state u, a patient entering state u at time t_j, has their next event at t_{j+1} which is defined as the minimum time amongst the survival times of the competing events $v_1, \ldots, v_{n_u}$. A row is created for each transition that is possible for the patient, with an additional column consisting of an indicator corresponding to whether or not the transition is observed or censored at t_{j+1}. In this format, we can separate the data by transition and fit a transition-specific parametric model to each subset [19]. Our required data format is described in detail in [32].

Weibull AFT model

In the survival framework, for a random variable T, denoting the time until an event of interest occurs, the survival function is given by $S(t) = 1 - F(t)$, where $F(t)$ is the cumulative density function of T. The hazard function, $\lambda(t)$, is defined as the instantaneous the instantaneous rate of occurrence of an event and is given by

$$\lambda(t) = \lim_{\delta t \to 0} \frac{P(t \leq T < t + \delta t \mid T \geq t)}{\delta t}.$$

If we assume that $T \sim \text{Weibull}(k, \alpha)$, for shape parameter k and scale α, then the baseline survival and hazard functions are given by $\bar{S}(t) = \exp(-\alpha t^k)$ and $\bar{\lambda}(t) = \alpha k t^{k-1}$, respectively.

In an AFT model, predictors, $\boldsymbol{x}$, act multiplicatively on time. This in contrast to the proportional hazards model where the predictors act multiplicatively on the hazard. If we let $\phi_i = e^{\boldsymbol{\gamma} \cdot \boldsymbol{x_i}}$, where $\boldsymbol{\gamma}$ are the regression coefficients, then we get that

$$\begin{aligned} S_i(t) &= \exp(-\alpha t_i^k \phi_i) \\ \lambda_i(t) &= \phi_i \bar{\lambda}(\phi_i t) = \phi_i k (\phi_i t)^{k-1} = \phi_i^k k t^{k-1}. \end{aligned}$$

The model is fit using the maximum likelihood estimation (MLE) method. Formulating the likelihood for a survival model requires the consideration of both the contribution of censored and uncensored individuals. For a potentially right-censored observation, let c_i be the event indicator for the ith individual with $c_i = 1$ if an event occurred and $c_i = 0$ otherwise. Then the likelihood is given by

$$L(\gamma, k; t_i) = \prod_{i=1}^{n} \lambda_i(t_i)^{c_i} S_i(t_i).$$

Therefore it follows that the log-likelihood for such a model is given by

$$\begin{aligned}\mathcal{L} &= \sum_{i=1}^{n} \left[c_i \log(\lambda_i(t_i)) + \log(S_i(t_i))\right] \\ &= \sum_{i} \left[c_i(\log(kt_i^{k-1}) + k\log(\phi_i)) - (\phi_i t_i)^k\right] \\ &= \sum_{i} \left[c_i(\log(kt_i^{k-1}) + k\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{x_i}) - t_i^k e^{k\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{x_i}}\right].\end{aligned}$$

Simulation/Bootstrap

In order to predict time to each transition from all states, we use a Monte Carlo simulation approach. This provides greater flexibility than computing length of stay via an integration, allowing us to predict patient pathways during an outbreak in addition to estimating length of stay in each state. As such, it creates a more powerful planning tool for hospital management. The number of simulated individuals, N, is based on COVID-19 positive hospital admissions from MFT since 23 February 2020. Individual survival times are simulated using estimates from each fitted transition-specific model, and iterating through the transition matrix until all patients have reached an absorbing state or are censored at a specified maximum follow-up time. The structure of the simulation treats the simulation as a sequence of competing hazards in the following way.

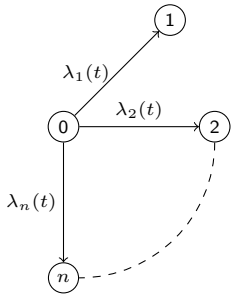
Let u be the patient's starting state, entered at time $t_u = s$ and $t_{\max}$ the maximum follow-up time of interest. For each day of interest, repeat the following to simulate paths for every new admission:

1 Let V be the set of states with an allowed transition from u and $Q_u = |V|$ be the number of possible transitions from u. While $v \in V$, let $\lambda_{uv}(t)$ be the transition intensity for the transition from u to v. Note, if $Q_u = 0$, we are in an absorbing state and stop.
2 For each possible transition, use the fitted parameter estimates of $\lambda_{uv}(t)$ to simulate a survival time, $\bar{t}_{uv}$.
3 Set $\bar{t} = \min\{\bar{t}_{u1}, \ldots, \bar{t}_{uQ_u}, t_{\max}\}$. If $\bar{t} = t_{\max}$, stop for this individual.
4 Let $u = z$ for $z \in V$ such that $\bar{t} = t_{uz}$ and set $t_u = \bar{t}$

Appendix E: Competing Hazards vs. Conditional Hazards

In this Appendix, we compare using conditional versus competing hazards within a multi-state framework. The MS model in the main text describes competing hazards, whereas the AFT and TC methods use conditional hazards. Competing hazards are perhaps more useful, but require very high quality data. If such data are not available, it may only be possible to estimate conditional hazards (where we condition on observing a given transition). However, here we demonstrate that these can be combined with estimates for the transition probabilities to obtain competing hazards. Thus, we conclude that coupling conditional hazards with transition probabilities can capture the same phenomena as a model based on competing hazards.

Consider the situation where a system has a state $X(t)$ and starts in state $X(0) = 0$ and moves to one of n states indexed by $i, j, \ldots \in [n]$, where $[n]$ is the set of integers from 1 to n inclusive.



Letting $\pi_i(t) = \Pr(X(t) = i)$, we get Chapman-Kolmogorov equations

$$\frac{\mathrm{d}\pi_0}{\mathrm{d}t} = -(1-\pi_0)\sum_{j=1}^{n}\lambda_j \ , \qquad \frac{\mathrm{d}\pi_i}{\mathrm{d}t} = (1-\pi_0)\lambda_i \ , \qquad i \in [n] \ , \tag{4}$$

for initial conditions $\pi_0(0) = 1$, $\pi_i(0) = 1$. We now consider two models. In a **competing hazards** approach, each rate is a general integrable function, and we write these integrals as

$$\Lambda_i(t) = \int_0^t \lambda_i(u)\mathrm{d}u \ , \quad i \in [n] \ . \tag{5}$$

In a **conditional** approach, we add an additional random variable, I, which is the state that the system will move to, i.e. $\lim_{t\to\infty} X(t)$. We then let

$$\lambda_i(t) = \mathbf{1}_{\{I=i\}} r_i(t) \ , \tag{6}$$

where $r_i(t)$ is the rate of going from 0 to i conditional on that being the event that happens. Integrals of these rates are defined as

$$R_i(t) = \int_0^t r_i(u)\mathrm{d}u \ , \quad i \in [n] \ . \tag{7}$$

Our aim is to show that the two approaches (competing and conditional) can be calibrated to give consistent results for quantities of interest.
One result needed for consistency is on the final outcome probabilities:

$$\Pr(I = i) = \lim_{t\to\infty} \Pr(X(t) = i) =: \pi_i^\infty \ , \quad i \in \{0\} \cup [n] \ , \tag{8}$$

where we have allowed for the possibility that the system may never leave the state 0, although for most of the parametric models we consider that will not be the case. Imposing consistency of the probability of being in state 0 over time, by integrating (4) and using the law of total probability for the conditional model,

$$\pi_0(t) = \exp\left(-\textstyle\sum_{i\in[n]}\Lambda_i(t)\right) = \pi_0^\infty + \sum_{i\in[n]}\pi_i^\infty \exp\left(-R_i(t)\right) \ . \tag{9}$$

Substituting (9) back into (4), we can then obtain (also using the law of total probability)

$$\frac{\mathrm{d}\pi_i}{\mathrm{d}t} = \lambda_i(t)\left(1 - \exp\left(-\textstyle\sum_{i\in[n]}\Lambda_i(t)\right)\right) = \pi_i^\infty r_i(t)\left(1 - \exp\left(-R_i(t)\right)\right) \ , \qquad i \in [n] \ . \tag{10}$$

We can then solve both (9) and (10) simultaneously using the Ansatz

$$\lambda_i = \mathcal{C}\pi_i^\infty r_i(t) \ . \tag{11}$$

This ensures that (10) is satisfied, then substituting into (9) we obtain

$$\exp\left(-\mathcal{C}\textstyle\sum_{i\in[n]}\pi_i^\infty R_i(t)\right) = \pi_0^\infty + \sum_{i\in[n]}\pi_i^\infty \exp\left(-R_i(t)\right) \ , \tag{12}$$

and hence

$$\mathcal{C} = \frac{\log\left(\pi_0^\infty + \sum_{i\in[n]}\pi_i^\infty \exp\left(-R_i(t)\right)\right)}{-\sum_{i\in[n]}\pi_i^\infty R_i(t)} \ . \tag{13}$$

This demonstrates that it is possible to capture the same phenomena of interest in the two models given appropriate calibration.

Appendix F: Additional results

In the main results section, we provided estimates for the LoS until any outcome. This was chosen to facilitate comparison between the different methods. In addition to this LoS, the TC method and the MS method can be used to estimate the length of stay until given outcomes such as discharge or death.[10] In this section, we compare estimates for these LoS. Again, pathways are disaggregated by whether the individual went via ICU. Here we choose to omit further predictor variables such as age or week of admission from the TC method to aid comparison.
Table A1 shows the comparison. For the LoS without ICU, the two methods give similar estimates. Using the TC method on the CHESS data predicts that LoS to mortality without ICU is slightly shorter than that to discharge, whereas the MS model on the MFT data predicts vice versa. This might be explained by the small sample size for the MFT data and the different demographic profile of the wider population captured by CHESS. The two methods predict very different LoS on ICU, with ICU to mortality being more than 5 days longer in MS (15.8 days) to that predicted by TC (10.2 days). Similarly, the LoS from ICU to stepdown, where individuals are discharged from ICU back to the general ward, is also longer with the MS model. As with the main results, this could most likely be explained by the presence of a much higher proportion of ECMO patients in the MFT data than in the CHESS data. The LoS from stepdown to discharge is similar for the two methods, with 7.9 from TC and 6.2 from MS. We are unable to estimate stepdown to ICU from CHESS due to the small number of such cases present in the data.
Figure A2 shows the output of our simulator on bootstrapped MFT data using the complete multi-state model in Figure 1. The red line represents true data that is plotted against 200 bootstrap simulations using fitted estimates for the transitions. Day 0 is taken to be 23 February, which we have taken to be the start of the national outbreak for the UK.

Table A1 Length of stay estimates to given outcome for England using the truncation corrected (TC) method, and for the Manchester Trust using the multi-state (MS) method. Source: own elaboration using CHESS and MFT data.

Method	Hospital trajectory	Mean	95% Confidence Interval
TC	Acute Ward to ICU	2.0	(1.9, 2.1)
TC	Acute Ward to Discharge	9.4	(8.9, 9.9)
TC	Acute Ward to Mortality	8.3	(7.8, 8.9)
TC	ICU to Stepdown	16.6	(15.6, 17.6)
TC	ICU to Mortality	10.2	(9.7, 10.7)
TC	Stepdown to ICU	NA	NA
TC	Stepdown to Discharge	7.9	(7.5, 8.3)
Multistate	Acute Ward to ICU	2.2	(1.9, 2.9)
Multistate	Acute Ward to Discharge	7.8	(7.0, 8.6)
Multistate	Acute Ward to Mortality	8.7	(7.5, 9.8)
Multistate	ICU to Stepdown	20.1	(15.9, 25.1)
Multistate	ICU to Mortality	15.8	(12.0, 21.5)
Multistate	Stepdown to ICU	2.2	(1.1, 7.6)
Multistate	Stepdown to Discharge	9.9	(7.2, 14.0)

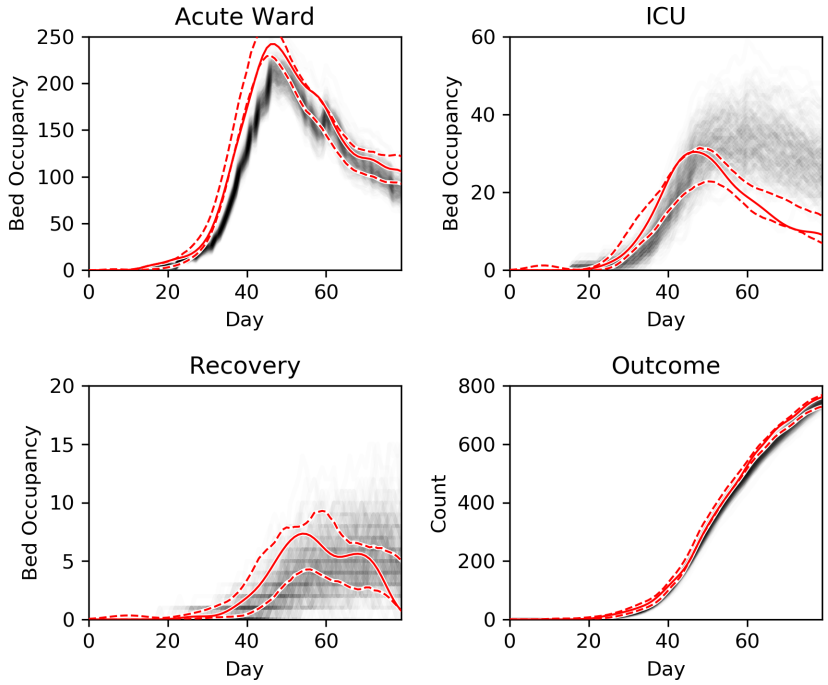


Figure A2 Output of our simulation using the MS method (black cloud of points) on MFT data (red lines). Source: own elaboration using MFT data.

Appendix G: Model Validation Results

Table A2 compares our mean LoS estimates from each model with the observed LoS in the data. For both the CHESS and MFT datasets we show the LoS from the censored data that was available prior to the 17th of May,

[10] A univariate AFT model is not fit for estimating LoS for competing hazards, such as for death and discharge.

which was the cutoff date for our original analysis, and from the uncensored data obtained once all patient outcomes had been observed. We also show the difference between the two, thereby showing by how much taking the observed mean LoS from the data underestimates the true LoS. For each model we then record the estimated LoS for each pathway and calculate both the absolute error when compared to the true uncensored LoS from the data as well as the error as a percentage of the LoS underestimate

Table A2 Validation of our models against the fully observed, uncensored (UC) data. For the accelerated failure time (AFT) and truncation corrected (TC) models we record both the absolute difference between our model mean estimates and the uncensored data from CHESS and as a percentage of the difference between the censored (C) and uncensored data. For the Multistate (MS) model we perform the same analysis, but against the data from MFT.

		Admission to ICU entry	ICU entry to ICU exit	Admission to Outcome (via ICU)	Admission to Outcome (no ICU)
CHESS Data	LoS (C)	2.12	11.58	16.03	8.37
	LoS (UC)	2.18	14.16	21.19	9.41
	LoS Underestimate	0.06	2.58	5.16	1.04
AFT model	Mean	2.00	12.40	16.20	8.40
	Difference	0.18	1.76	4.99	1.01
	% of adjustment	-200.00	31.78	3.29	2.88
TC model	Mean	2.00	13.40	17.30	9.10
	Difference	0.18	0.76	3.89	0.31
	% of adjustment	-200.00	70.54	24.61	70.19
MFT Data	LoS (C)	1.30	13.72	17.21	6.86
	LoS (UC)	1.35	18.58	30.57	8.52
	LoS Underestimate	0.05	4.86	13.36	1.66
MS model	Mean	2.33	18.93	29.73	7.97
	Difference	-0.98	-0.35	0.84	0.55
	% of adjustment	2060.00	107.20	93.71	66.87