Simulating Occupancy for Short-Term Hospital Planning

John Minty[1] and Paul Harper
School of Mathematics, University of Cardiff, UK, CA24 4AW

Christophe Sarran
Met Office, Fitzroy Road, Exeter, Devon, UK, EX1 3PB

Sujit Sahu
SSSRI, University of Southampton, UK, SO17 1BJ

Bernard Baffour
ISSR, University of Queensland, Australia, QLD 4072

Abstract. Improved short-term predictions of hospital occupancy offer potentially more effective use of resources. The MetSim project makes short-term simulated forecasts of bed needs, distinguishing patients by age-group, gender and broad class of medical speciality so as to reflect the logistics of the hospital. Models of admissions take weather into account. Patients' length of stay is modelled by hazard ratios. The project is a collaboration between Cardiff University, the University of Southampton and the Met Office.

Keywords: occupancy; admissions; length of stay; hazard ratios; simulation; speciality supergroups.

1. Introduction

The overall aim of the project is to create a tool which forecasts admissions and bed occupancy in hospitals in light of meteorological information. Forecasts are for twenty-one days including the day of forecast itself. The tool is of benefit to managers who can plan services or utilise resources better. For example, it is more useful to anticipate a surge in geriatric patients two weeks in advance than recognizing the event only two days ahead. As another example, if the tool predicts a trough in non-elective admissions a week ahead then managers may consider summoning additional elective patients, so earning greater revenue. The long-term ambition is to produce a tool

[1] Corresponding author MintyJH@cardiff.ac.uk

which is flexible enough to suit a variety of hospitals, but this exploratory study, entitled "The MetSim Project", has focused on two hospitals: Cardiff & Vale Hospital and Southampton University Hospital.

Throughout, our simulation is for non-electives only. Elective patients have been well-analyzed, mostly by queueing theory; see Utley *et al* (2003), or Worthington (1987). Moreover, we are concerned with the main hospital, not Accident and Emergency which is administratively quite separate. Klein & Reinhardt (2012) and Rasheed *et al* (2012) present studies of simulation in an emergency department. Wiler, Griffey & Olsen (2011) provide an overview of the literature on emergency departments and also compares methodologies, in particular regression models, time-series models, models based on queueing theory and discrete event simulation. Vasilakis and El-Darzi (2001) and Bowers (2009) give simulations of the main hospital; both papers observe that seasonal variation occurs, for which holidays are an explanatory factor.

After consultation with hospital managers, we refined the tool so as to reflect certain logistic fundamentals of hospitals. It is necessary to have forecasts by age-group, by gender and by broad group of speciality. In particular, forecasts for paediatrics must be separate from non-paediatrics since the two groups of patients are legally required to be separate; the MetSim tool does allow the user to subdivide each of paediatric and non-paediatric into finer age bands. A further legal requirement is that hospitals segregate patients by gender, so our tool does this too. Hospitals also classify patients into large blocks according to medical speciality, but the classification is not completely standard across hospitals. We term the large speciality blocks *supergroups*. The code is flexible here and allows the hospital to specify their own partition, subject to a maximum of two paediatric supergroups, five non-paediatric supergroups, together with one further supergroup "outliers" of patient-spells which do not conform to the specified groupings and which are disregarded from all subsequent modelling and forecasting. We constrain the number of supergroups since further refinement no longer matches logistics; moreover, it risks over-granularity of data with consequent low cell counts and uninformatively broad forecasts. The precise choice of categorizations by age-group and speciality is termed the *design* of the hospital. In practice, the two participating hospitals, Cardiff and Southampton, requested very similar designs: that paediatrics remain a single block, undivided by age or speciality. Non-paediatrics should be partitioned by age-groups into adult and elderly, and partitioned into supergroups Medicine, Surgery, Trauma and Cardiothoracic, although there were minor differences in the demarcations. Outliers should include numerous miscellanea such as audiology, obstetrics and dental.

The MetSim tool to simulate occupancy consists of two steps: admissions and length of stay. The method of simulating admissions rests on a forecasting model in Sahu *et al* (2014): for any age-group and gender of patient, we forecast admissions using Bayesian models to allow for uncertainty in weather forecasts. Our tool then combines the forecasted admissions with current (prevalent) patients. For each cohort of patients, by age, gender, supergroup and time already spent, we simulate flow out of hospital, so producing distributions of occupancy. To simulate departure from hospital, the tool requires information on how long patients stay.

Two families of distributions are commonly used to model length of stay in hospital: mixture distributions, especially hyperexponential, and Coxian phase distributions. Mixture models are appropriate in the following situation. Patients are all of a limited number of types, and within any type the lengths of stay are independent and identically distributed. Then the length of stay of any admitted patient is found by conditioning on the type of patient. McClean and Millard (1993) employ mixture models for geriatric patients. By contrast, Coxian distributions are used for an homogeneous group of patients undergoing various phases, equivalently steps or stages, within hospital. McClean and Millard (1998) employ a 3-stage model again for geriatric patients; a patient progresses through acute care, rehabilitation and long care, in that order. The model implicitly contains a fourth state, namely discharge or death, and it is possible to progress there directly from any of the other states. Many further examples appear in Marshall, Vasilakis & El-Darzi (2005), an account which also discusses the merits and drawbacks of discrete event simulation.

Fackrell (2009) elegantly describes how hyperexponential models and Coxian distributions are themselves special instances of a very broad family of distributions: the phase-type distributions. Fackrell lists various algorithms to fit phase-type distributions, while acknowledging that one must guard against overparameterization. It may be added that, even if one attains a good, parsimonious fit, such models are not always easy to interpret. In the special instance of fitting a Coxian distribution, McClean and Millard (1998) use the Levenburg-Marquart algorithm. McGrory, Pettit & Faddy (2009) take a Bayesian approach.

Any model of length of stay faces difficulties inherent in the distribution. Typically, there is a mode near zero, measured in hours, and a protracted tail of patients who remain months or in extreme cases years after being admitted. An instance is illustrated in Figure 1 which shows the length of stay of non-elective patients in Southampton University Hospital, comparing the figures for 2009 against earlier years. The plot was part of our early investigations, to see whether length of stay itself evolves in time. At the broad level of supergroups, no such pattern was found: improved medical techniques reduce length of stay while aging populations increase it.

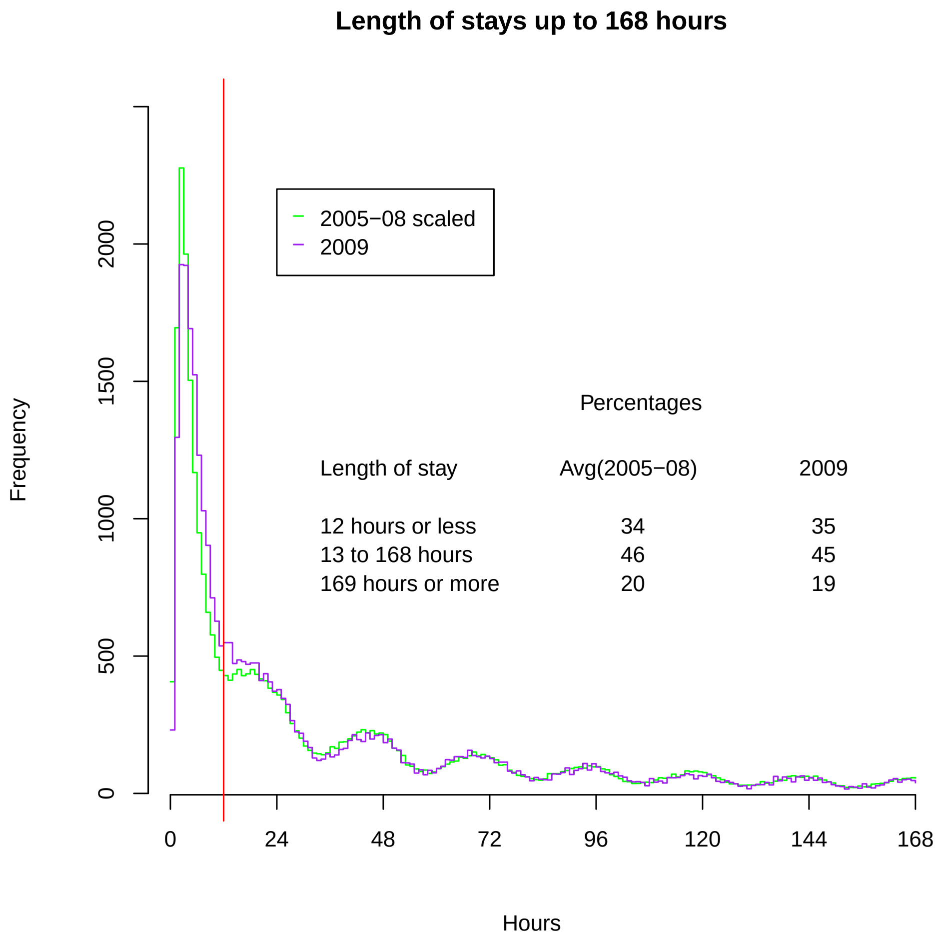

Figure 1. Southampton University Hospital.

Further early investigation showed that, within any supergroup, lengths of stay are often irregular. Attempts to model length of stay using mixed models or phase-types have proved unsatisfactory, rarely explaining more than 30% of the variation, and so instead we have modelled the hazard rates: a patient's probability of discharge at any given time, given that the patient has survived thus far. In this context, "survival" means not leaving hospital by whatever route, whether by discharge, transfer or death. Empirical observations of hospital data suggest that the hazard rate depends on age and on factors such as supergroup and current day of week. Figure 6 illustrates hazard rates for Cardiff and Southampton Hospitals.

2. Datasets, timeline and machinery

A participating hospital regularly provides datasets on flows of patients: one, labelled *historic*, covers an interval which is at least one year long and which strictly precedes the date of execution. The second dataset, labelled *current*, is simply a census of prevalent patients, namely patients recorded as present shortly before the start of the forecast. The hospital also states the design of how patients are to be classified into supergroups. The Met Office provides datasets on weather, executes the machinery of the code and returns summary forecasts back to the hospital. Figures 2, 3 and 4 show the machinery in some detail. A description follows.

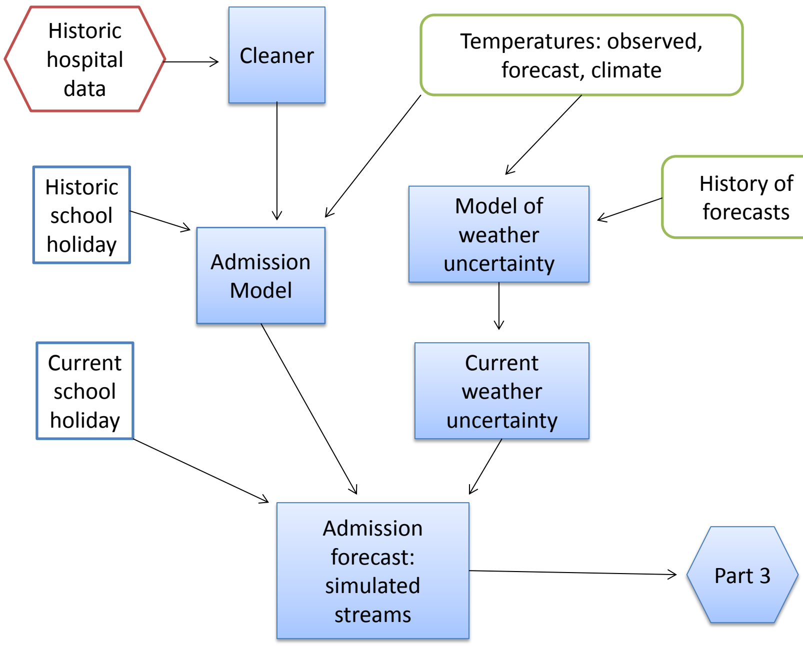

Figure 2. Modelling and simulating admissions.

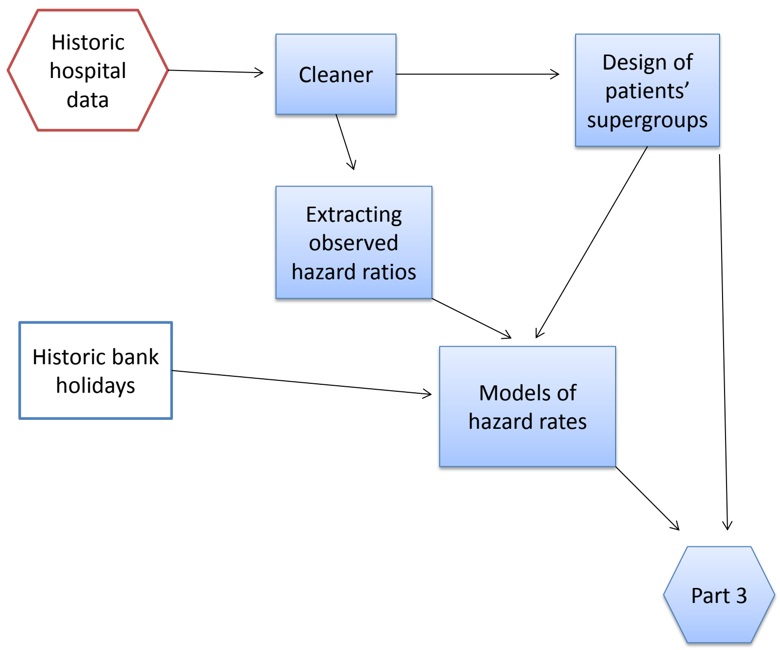

Figure 3. Modelling hazard ratios.

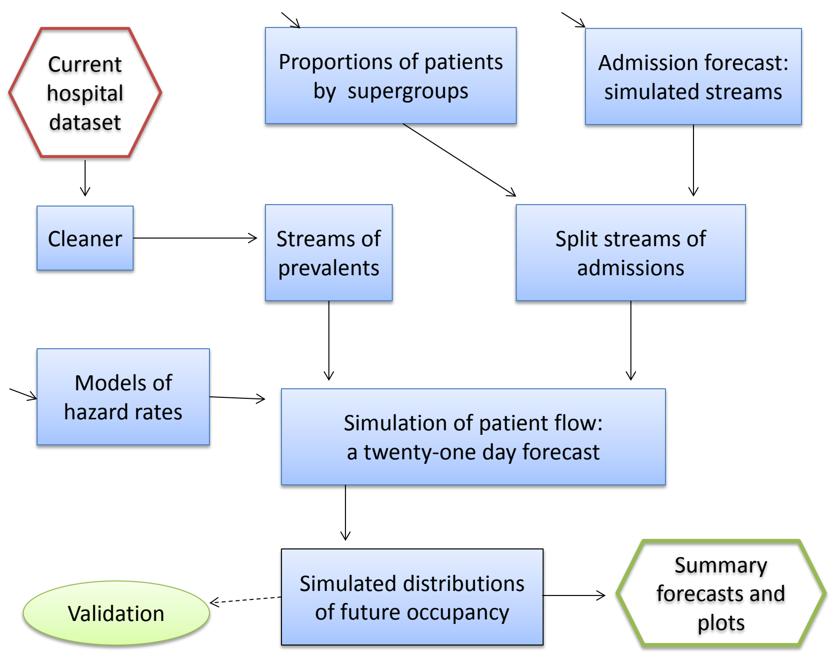

Figure 4. Simulating patient flow.

The historic dataset which the hospital supplies consists of anonymized patient spells over the course of at least one year. The interval of observation is open at both ends: we include all spells of patients who were either admitted or discharged or both. The latest endpoint should ideally be about six to eight weeks earlier than the date of execution. The current dataset, or census, should be taken within the last week. For every patient-spell the following fields should be included. Fields marked * are desirable but not mandatory.

Table 1. Hospital datasets

Historic	**Current**
Date of query	Date of query
	Hour of query*
Spell start date	Spell start date
Spell start hour*	Spell start hour*
Age at admission	Age at admission
Gender	Gender
Speciality code	Speciality code
Spell end date	
Spell end hour*	

The Met Office provides mean and minimum temperatures over a long time interval, starting from two-and-a-half years before the date d of execution. In our early analysis we considered several other variables such as rainfall and humidity and so on, but found that once temperature is included the others are too highly correlated to be informative. The mean temperatures up to d are observed; for approximately the next seven days, including the day of execution itself, we rely on weather forecasts; for the remainder of the future interval we use climate averages as predictions. The minima also change qualitatively but are lagged by one week.

The same dataset of temperatures brings another, major benefit. It serves as a convenient timeline against which other datasets can be aligned for modelling and simulation. The code checks that the historic hospital dataset is not too recent; if it is, the code first attempts to trim the historic dataset or, if that reduces the dataset to less than one year's duration, halts the programme. The code likewise checks the date c_o of observation of the current dataset. If that is more than one week old then the programme halts. The code further edits the current dataset since under-reporting is both a frequent and an extensive problem: patients who have been admitted immediately before the census was taken are at high risk of not being counted. To correct for this, the code truncates the current dataset a couple of days before the date of the census itself, on a day c_t we call the "day of last trusted admissions" and which is derived from the date of census according to some policy or rule set by the hospital. Details are in Minty *et al* [9]. Instead of employing the under-reported numbers, the code *simulates* admissions over the untrusted period. Incidentally, the current dataset also carries risk of over-reporting but that occurs comparitively rarely.

By way of illustration, suppose hypothetically that the day d on which we execute the simulation is Wednesday 13th March 2013. The principal weather dataset consists of observed temperatures extending back several years, together with weather forecasts which span the interval from 13th March to 19th March inclusively, followed by climate averages until 2nd April. An historic dataset covers the interval from 1st December 2011 to 22nd January 2013 and so presents no difficulty in this instance. The most recent current dataset, submitted to the Met Office on 13th March, is for an observation on 12th. We choose Monday 11th March as our last trusted day. Then part of the timeline in this example looks like the following.

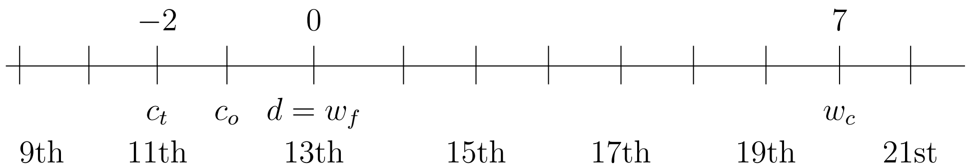

Figure 5. Example of a timeline in March 2013.

The first day of forecast is w_f and climate average is w_c. The current dataset is observed on day c_o and our last trusted day is c_t. The code starts simulating admissions from day $c_t + 1$ even though 12th March is a day of observed weather in this example.

Quite separately from the principal dataset of temperatures, the Met Office supplies an auxiliary dataset for a rather subtle reason. When we model admissions, we use weather as an explanatory variable, using observed temperatures. But in simulating future admissions we rely on weather forecasts or even climate average. To allow for error in weather forecasts we employ a dataset which shows a history of forecasts at the given hospital postcode: for every observed value, there are predictions made $0 \leq t \leq 20$ days ahead. Other datasets needed are bank holidays and school holidays in the neighbourhood of the hospital.

3. Modelling

The code, which is held at the Met Office, depends on models of admissions and length of stay developed at Southampton and Cardiff Universities respectively. For a comprehensive account of modelling admissions, see Sahu *et al* (2014). We construct a model of historic admissions in order to forecast future admissions; the dataset of current occupancy plays no part in this model. We choose to fit a generalized linear model to the historic hospital admissions in preference to a time-series model: it seems that any day-to-day correlation of admissions is explained wholly through weather and day of the week, both of which are given to us, and a generalized linear model is computationally more tractable. The data are positively skewed and we select a square-root transformation for the response variable number of admissions. We denote the rescaled response variable by Y_{ijt}, where i is gender, j is age-group and t time. The model is for $0 \leq t \leq 20$ days ahead. We assume a regression model

$$Y_{ijt} \sim \mathrm{N}(\mu_{ijt}, \sigma^2)\,, \tag{1}$$

where

$$\begin{aligned} \mu_{ijt} \;=\; & \beta_0 \;+\; \alpha_i \;+\; \gamma_j \;+\; h(t) \;+\; w(t) \;+\; \lambda_m\, m(t) \;+\; \lambda_n\, n(t) \\ & +\; (\alpha : \gamma)_{ij} \;+\; \big(\gamma : n(t)\big)_j\,. \end{aligned} \tag{2}$$

The initial term β_0 is an intercept. The factor α_i denotes gender; we set female $\alpha_1 = 0$. The factor γ_j is the contribution of the jth age-group. The binary-valued function $h(t)$ is a school holiday effect and is zero for non-holidays. The effect $w(t)$ is day of the week; it is zero for Sundays.

On day t, the mean temperature is $m(t)$ and the minimum a week ago is $n(t)$; the coefficients are λ_m and λ_n respectively. We also have interactive terms, between gender and age-group, and between age-group and lagged minima. Such a model is very parsimonious and yet is best according to both R^2 and AIC. We do not use supergroups of speciality in modelling historic admissions. If included, supergroups lead to low cell counts and, out of all explanatory variables, they contribute least information. The tricky factor is the age-group γ_j, for the number of levels occurring depends on the design of the hospital. Choice of design is important; in a good design, age alone explains a remarkable 74% of the variation in numbers of admissions. We allow the number of paediatric age-groups to be 1 or 2 and of non-paediatric to be 1, 2 or 3. Thus the number of age-groups ranges from 2 to 5, and the number of contrasts in age-groups from 1 to 4. We constrain γ_j to be zero for the youngest age-group: paediatrics if they are unsplit or young paediatrics if they are split. As for the interaction terms, $(\alpha : \gamma)_{ij}$ and $\big(\gamma : n(t)\big)_j$, the parameter is set to be zero whenever one of the factors is at its base level. The factors β_0, α_i, γ_j, $h(t)$, $w(t)$ and coefficients λ_m, λ_n together with their interactive terms have the general appellation "parameters". The number of parameters, which depends on the number of age-groups, we call the dimension of the design.

A classical approach gives point estimates of the parameters together with confidence intervals. But anticipating difficulties with temperature, we prefer a different approach, one which uses Bayesian MCMC regression to simulate, say, $10K = 10\,000$ randomly generated values of the parameters which we store in a matrix $\mathbf{P}$.

$$\mathbf{P} \;=\; \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\beta_0} & \boldsymbol{\alpha_2} & \boldsymbol{\gamma_j} & \cdots & \boldsymbol{\lambda_m} & \boldsymbol{\lambda_n} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \qquad \begin{array}{c} \uparrow \\ 10K \\ \downarrow \end{array}$$

The matrix $\mathbf{P}$ has $10K$ rows; the number of columns equals the dimension of the design. The same MCMC regression simultaneously generates $10K$ random values of the variance σ^2 in Equation 1. For a suitable choice of design matrix $\mathbf{X}$, consisting of entries 0's, 1's and temperatures $m(t)$ and $n(t)$, the product

$$\mathbf{P\,X} \tag{3}$$

has entries exactly as in the right-hand side of Equation 2. The columns of $\mathbf{X}$ may be arranged, for convenience, into blocks by age-group and gender, and within any block by timeline of forecast interval. The product $\mathbf{PX}$ has $10K$ rows or realizations; the columns occur by age-group j and gender i, and within any such block are ordered by time. For each of the $10K$ realizations of (μ_{ijt}) and of the variance σ^2, it is possible to generate values of (Y_{ijt}) using Expression 1. However, we choose *not* to use design matrix $\mathbf{X}$ directly. When simulating coefficients in $\mathbf{P}$, we use observed mean temperatures $m(t)$. During the interval of patient forecasts, we rely on weather forecasts and climate averages $m(d+t)$, for $0 \leq t \leq 20$. If we plug $\lambda_m \, m(d+t)$ into our realization we fail to capture uncertainty in weather predictions. Accordingly, we distinguish true future mean temperatures $\breve{m}(d+t)$ from forecast future mean temperatures $m(d+t)$.

To simulate $\breve{m}$, we use the auxiliary weather dataset showing a history of forecasts against actual observation; this is its sole use in the whole canon of programmes. For $0 \leq k \leq 20$ let k denote how far ahead a forecast is. For a fixed but arbitrary value of k, we assume true means are some linear function of forecasts

$$\breve{m}(t+k) \;=\; a_k + b_k \, m(t+k) + \varepsilon_k \qquad \text{for all } t,$$

where ε_k is noise. The values of t span a year. We run a sequence of further MCMC regressions, one for each value of k, to simulate values of a_k and b_k which we stash in a matrix.

$$\begin{pmatrix} a_0 & b_0 \\ a_1 & b_1 \\ \vdots & \vdots \\ a_{20} & b_{20} \end{pmatrix}$$

We duly compute values of $\breve{m}(d+t)$ over our interval of forecast and substitute these into $\mathbf{X}$. Only now do we carry out our realization $\mathbf{PX}$ and so generate the means in Equation 2 and so simulate streams of admissions. That completes modelling and simulating admissions as shown in Figure 2.

We regard length of stay as a non-negative integer number of days $n \geq 0$. To model length of stay we use survival analysis, where "survival" is not leaving the hospital by whatever route. Let $h(n)$ denote the hazard rate. Figure 6 shows the hazard rates for the hospitals, segregating patients according to age-group. Broadly, the hazard rates are decreasing: the longer one spends in hospital, the smaller the probability of leaving. But there are short-term increases about 7 days, especially for paediatrics, suggesting that patients are sometimes intentionally held for a week and then discharged. Furthermore, a length of stay marked 0 has a peculiarity. We round length of stay to the nearest integer number of days, labelling a stay of 0 to 11 hours as 0 days. All remaining days cover an interval of 24 hours. Values from 15 to 21 days should be read with caution since the number of observations declines over time.

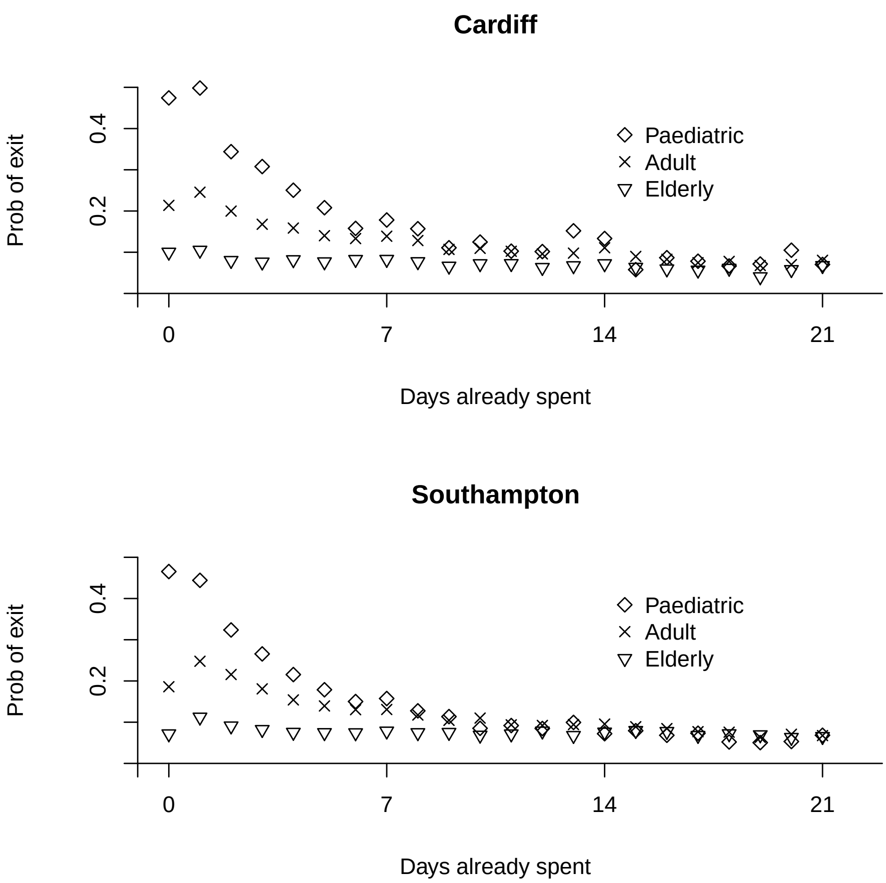

Figure 6. Hazard rates by age.

In practice, the hazard rates are mostly quite small and so we take the log-odds transformation

$$K(n) \ := \ \ln \frac{h(n)}{1-h(n)} \,,$$

numbers which we call K-values, or $K(n)$ if we wish to emphasize the number of days spent thus far. The transformation is monotonic, so K increases if and only if h increases. From earlier comments, K is usually decreasing. Figure 7 shows K-values for the non-paediatric patients; we illustrate the different speciality supergroups.

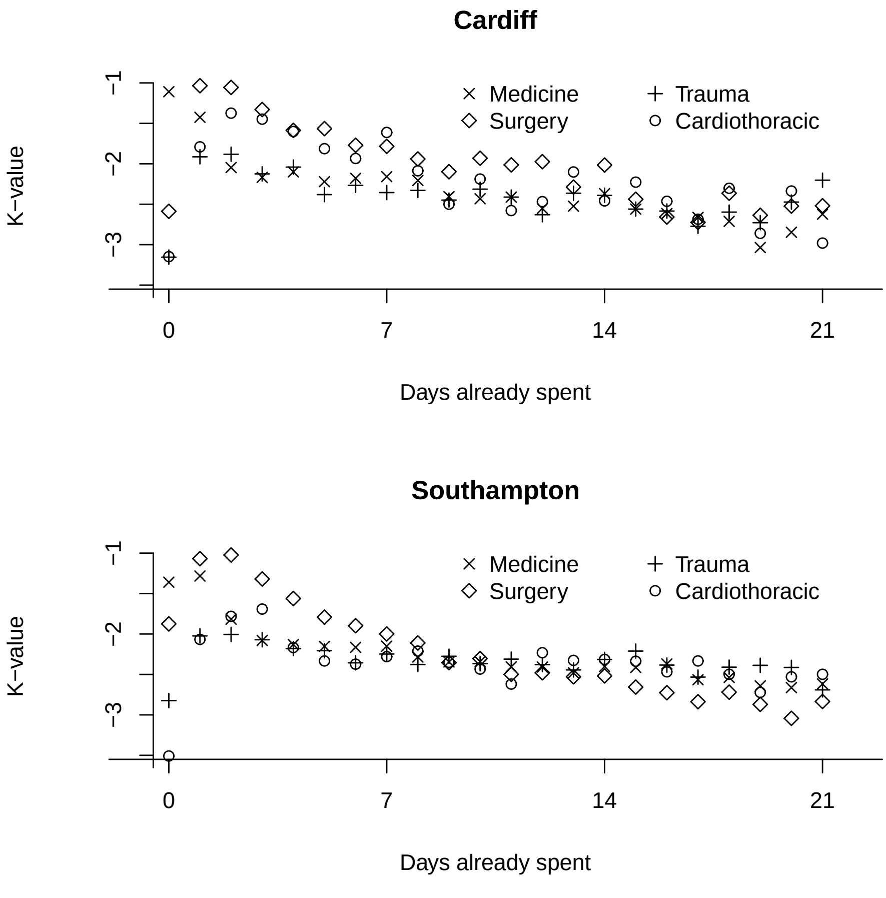

Figure 7. Non-paediatric K-values by speciality supergroups.

We adopt a hybrid strategy for fitting estimated K-values to the historic data.

The first few days of length of stay are inherently irregular but the historic dataset provides an abundance of observations. Accordingly, we model the K-values for each day separately. For day n the main effects model is

$$K(n) \;=\; \alpha_{in} + \gamma_{jn} + w(n)\,.$$

Each of the explanatory variables is factorial. The variable α_{in} denotes gender. Its contribution is marginal; after the first few days, men and women show little difference in rate of discharge. But we retain it for logistic reasons. The variable $w(n)$ denotes current day of week. Rates of discharge at weekend differ from during the week; we bin days of week, allowing w to represent one of three cases: weekend or bank holiday, Monday to Wednesday and Thursday-Friday. Note that we are looking not at the day of admission but the day of putative discharge. The variable γ_{jn} denotes categorization by age and speciality supergroup. Since paediatric specialities differ from non-paediatric specialities we nest supergroups into age. For the very first few days, $n = 0, 1$, we also include an interactive term $\big(\gamma, w(n)\big)_j$.

Up to a length of stay of about a week, such models perform well: even in the least favourable cases over 70% of the variation is explained and values over 90% occur about half the time. Thereafter predictive accuracy steadily deteriorates due to low cell counts, and so we offer an alternative approach. For each separate age-speciality category, we fit the K-values to log-time.

$$K(n) \;=\; a + b\,\ln(n) + w(n) + \alpha_i$$

The term a is an intercept. The factor w is binary and is zero for weekends. The factor α_i denotes gender and is zero for female. Such models usually explain between 50% to 80% of the variation. However, we cannot guarantee robustness, especially for elderly trauma patients. The system itself assesses model performance and automatically decides which of the two models to use. Incidentally, under the secondary model the hazard rate has the simple form

$$h(n) \;=\; \frac{1}{1 + Cn^{-b}}$$

for some $C > 0$, a form close to the hazard rate of a Weibull distribution. That completes the steps in Figure 3.

4. Simulation

The above Bayesian model of admissions randomly generates streams over the timeline, starting from immediately after the last trusted day and ending at twenty days from the day of execution. It distinguishes patients by age-group and gender but not by the speciality supergroup. So for any age-group i and gender j the first step is to split streams of admissions, using multinomial probabilities (p_{ijk}) that a patient falls into the respective supergroup. For each i, j the sum $\sum_k p_{ijk} = 1$, where $1 \leq k \leq 2$ for paediatrics and $1 \leq k \leq 5$ for non-paediatrics. To estimate (p_{ijk}) we read the observed proportions in our historic dataset. Now we go through every row of admissions in the source file for i, j and randomly generate a substream for each supergroup. We repeat for the same row of our source file to give a second simulated split. According to the design of the hospital, some probabilities may be zero, leading to the construction of "empty" streams which consist of zeros, but that does no harm.

Throughout the main simulation we consider patients at the finest level of categorization. As indicated in Figure 4, for any age-group, gender and speciality supergroup we have three large sources of information: models of hazard rates of leaving hospital, simulated streams of admissions over the timeline, and a file which contains the distribution of prevalent patients by length of stay for some recent day. The timeline holds additional information such as bank holidays. The main aims of the simulation are, of course, to simulate streams of discharge and occupancy over the timeline. "Discharge" is a shorthand meaning all forms of exit. A secondary aim is to give a detailed profile of occupancy for a single future date during the interval of forecast: not just the total number of occupants forecasted but also how long these future occupants have themselves stayed.

We partition our chosen category of patient into cohorts by length of stay. Explicitly, for each day n, let

$$\mathbf{S}_n = (s_0, s_1, s_2, \dots)$$

denote the counts s_k of patients who have already stayed $k \geq 0$ days. In particular, s_0 is the number of patients who have been admitted only recently. The sum $||\mathbf{S}_n|| = \sum_k s_k$ gives the total occupancy on day n. The simulation constructs $\mathbf{S}_n$ recursively along the timeline. Initially, we set $\mathbf{S}_0$ to be $\mathbf{0}$; that is for the last trusted day, immediately before the start of our simulated admissions. The general method of constructing $\mathbf{S}_{n+1}$ from $\mathbf{S}_n$ is as follows.

1. Using binomial distributions $\mathrm{Bin}(s_k, p_k)$ where p_k is the appropriate hazard rate, generate a vector
$$\mathbf{D}_n \;=\; (d_0,\; d_1,\; d_2,\; \ldots)$$
showing the number of discharges on day n.

2. Take the difference to find the number of remaining patients
$$\mathbf{S}_n - \mathbf{D}_n \;=\; (s_0 - d_0,\; s_1 - d_1,\; s_2 - d_2,\; \ldots)\,.$$

3. Slide one place to the right and insert the admissions; that gives the occupancy for the next day.
$$\mathbf{S}_{n+1} \;=\; (a,\; s_0 - d_0,\; s_1 - d_1,\; s_2 - d_2,\; \ldots)\,.$$

There is a complication to the above. When the iteration reaches the day of observation, we include the prevalent patients. If the prevalents were observed during the morning then we add the prevalents before randomly simulating the number of discharges; otherwise, we add the prevalents after deducting discharges. Only now do we start to record total occupancy $||\mathbf{S}_n||$, and we continue recording until the end of the twenty-one days of forecast. For the day of profile, we record not just the total but the full distribution of $\mathbf{S}_n$.

In practice, to speed up computations in step 1 of the algorithm, we compute tables $(\mathbf{B}_k)_{k\geq 0}$ of binomial probabilities before we start the iterations. That is strongly preferable to recomputing the probabilities as the algorithm progresses. For $k \geq 0$, for $0 \leq n < N$, row n of the $N \times N$ array $\mathbf{B}_k$ contains the cumulative distribution function of the binomial distribution $\mathrm{Bin}(n, p_k)$. Note that the first row of the array is regarded as row 0, not 1, and is for the trivial binomial distribution which takes 0 with probability 1. The size N of the array, which depends on k, represents a bound on how many patients of the given category are likely to serve k days. From empirical testing, for $k = 0$ we set $N = 120$ but we reduce N steadily thereafter. To actually compute $\mathbf{B}_k$, we iteratively derive row $n+1$ from row n; an alternative method is given by Loader (2000).

That completes the simulation. It is possible to plot each of the generated streams of admissions, discharges and occupancy. We present a plot of forecasted adult admissions, taken from subsequent validation. The periodic effect of a week is stark; there is a gentle trend towards fewer admissions. The last Friday, 29th March 2013, happens to be Good Friday.

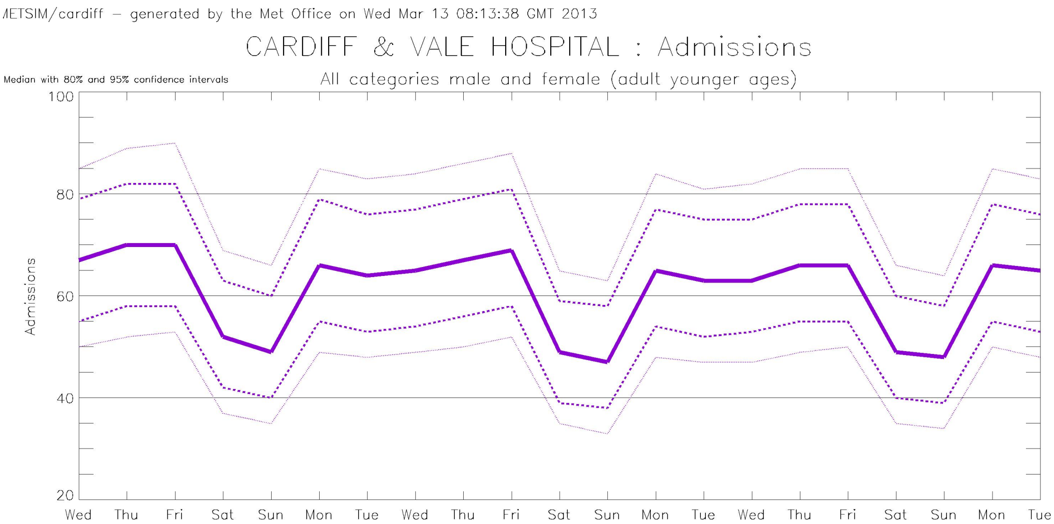

Figure 8. Forecast of admissions from Wed 13th March 2013.

5. Discussion

Prior to this tool, hospitals widely use one of two methods to forecast admissions and occupancy. The simplest model is the Persistence Model where the hospital looks at the figures for one year ago and adjusts for demographic trends. The other model is the Six-Week Moving Average. From preliminary validation and verification, our model seems at least comparable to the Six-Week Moving Average and both clearly outperform the Persistence Model. For example, Table 2 shows the relative root mean square error for Cardiff over a validation interval of 130+ days. We give the figures for Cardiff in preference to Southampton since we have a longer interval of validation.

Table 2 Cardiff Adult Admissions.

Days ahead	Relative RMSE
0	0.039
1	0.004
2	−0.014
3	−0.026
4	−0.004
5	−0.022
6	−0.013
7	−0.019
8	−0.015
9	−0.019
10	−0.033
11	−0.005
12	0.035
13	0.009
14	−0.009
15	0.007
16	0.010
17	−0.006
18	−0.000
19	−0.014
20	−0.003

For $0 \leq k \leq 20$ days ahead, the relative root mean square error is defined to be

$$\frac{\mathrm{RMSE}(y_k)}{\mathrm{RMSE}(z_k)} \;-\; 1\,.$$

where y_k and z_k refer respectively to the medians forecasted by our tool and the medians forecasted by the Six Week MA. On this measure, negative values are desirable, and that is what happens for 15 out of 21 days of forecasting. Other measures, such as percentage cover, confirm that the tool gives satisfactory forecasts for admissions. To date, the forecasts for discharges and occupancy are less impressive.

The tool seems to perform well under steadily evolving conditions such as seasonal changes in temperature. Although the tool cannot forecast abrupt changes, whether unforeseen such as a fire in a neighbouring hospital or planned such as the opening of a new ward, the tool does incorporate the changes swiftly once a fresh current dataset has been supplied.

We aim to improve the code at several points. We wish to create an interface for ease of use. We aim to simulate the admissions in a lower order language for speed. As for hazard rates, it is worth testing whether it is more robust to adopt a branching method which models the risks of leaving in week $0, 1, 2$ and, conditional upon leaving in a certain week, the risk of leaving on a particular day. Another point of investigation is the splitting probabilities (p_{ijk}). The code simply reads the proportions in the historical dataset, but recent current datasets may be more informative. We hope to pilot a revised version of the tool in several hospitals across the UK, especially diverse hospitals in a variety of geographical locations.

Acknowledgements
The authors are grateful to participating hospitals for discussion and data provided, especially to B Hendy from Southampton University Hospital Trust, to A Nelson, S Tarr and J Peters from Cardiff and Vale Health Board and to R Blackwell at Royal Devon and Exeter; also to supporting staff at the Met Office especially P Sachon and Y Clewlow, to G Pierce from Cardiff University and for grant EP/H010637/1 from the Engineering and Physical Sciences Research Council.

Competing interests

None.

References

Bowers J (2009) Simulating waiting list management *Health Care Management Science* **14**, 3, pp292-298.

Fackrell M (2009) Modelling healthcare systems with phase-type distributions. *Health Care Management Science* **12**, pp11-26.

Klein M, Reinhardt G (2012) Emergency department patient flow simulations using spreadsheets. *Simulation in Healthcare* **7**, 1, pp 40-47.

Loader C (2000) Fast and accurate computation of binomial probabilities. http://www.researchgate.net/publication/2640094_Fast_and_Accurate_Computation_of_Binomial_Probabilities

Marshall A, Vasilakis C, El-Darzi E (2005) Length of stay-based patient flow models: recent developments and future directions. *Health Care Management Science* **8**, pp 213-220.

McClean S and Millard P (1993) Patterns of length of stay after admission in geriatric medicine: an event history approach. *The Statistician* **42**, pp 263-274.

McClean S and Millard P (1998) A three compartment model of the patient flows in a geriatric department: a decision support approach. *Health Care Management Science* **1**, pp159-163.

McGrory CA, Pettitt AN, Faddy MJ (2009) A fully Bayesian approach to inference for Coxian phase-type distributions with covariate dependent mean. *Computational Statistics and Data Analysis* **53**, pp 4311-4321.

Minty R, Harper P, Sarran C, Sahu S, Baffour B (2013) Annotation of hospital simulation. *Unpublished.*

Rasheed F, Lee Y, Kim S, Park I (2012) Development of emergency department load relief area-gauging benefits in empirical terms. *Simulation in Healthcare, Journal of the Society for Simulation in Healthcare* **7**, 6, pp 343-352.

Sahu SK, Baffour B, Harper PR, Minty JH, Sarran C (2014) A hierarchical Bayesian model for improving short-term forecasting of hospital demand by including meteorological information. *J R Statist Soc A*.

Utley M, Gallivan S, Treasure T, Valencia O (2003) Analytical methods for calculating the capacity required to operate an effective booked admissions policy for elective inpatient services. *Health Care Management Science* **6**, pp97-104.

Vasilakis C and El-Darzi E (2001) A simulation study of the winter bed crisis. *Health Care Management Science* **4**, pp31-36.

Wiler J, Griffey R, Olsen T (2011) Review of modeling approaches for emergency department patient flow and crowding research. *Academic Emergency Medicine.* **18**, 12, pp 1371-1379.

Worthington D (1987) Queueing models for hospital waiting lists. *Journal of the Operational Research Society.* **38**, 5, pp 413-422.