Modelling admission lengths within psychiatric intensive care units

Stephen Dye [ORCID],[1] Faisil Sethi [ORCID],[2] Thomas Kearney,[3] Elizabeth Rose,[4] Leia Penfold,[5] Malcolm Campbell [ORCID],[4] Koravangattu Valsraj [ORCID][6]

[1]Suffolk Access Team, Norfolk and Suffolk NHS Foundation Trust, Norwich, UK
[2]Dorset HealthCare University NHS Foundation Trust, Poole, UK
[3]NHS Improvement, London, UK
[4]Central and North West London NHS Foundation Trust, London, UK
[5]Suffolk Central IDT, Norfolk and Suffolk NHS Foundation Trust, Norwich, UK
[6]South London and Maudsley NHS Foundation Trust, London, UK

Correspondence to
Dr Stephen Dye;
stephen.dye@nsft.nhs.uk

ABSTRACT

Objectives To examine whether discharge destination is a useful predictor variable for the length of admission within psychiatric intensive care units (PICUs).

Methods A clinician-led process separated PICU admissions by discharge destination into three types and suggested other possible variables associated with length of stay. Subsequently, a retrospective study gathered proposed predictor variable data from a total of 368 admissions from four PICUs. Bayesian models were developed and analysed.

Results Clinical patient-type grouping by discharge destination displayed better intraclass correlation (0.37) than any other predictor variable (next highest was the specific PICU to which a patient was admitted (0.0585)). Patients who were transferred to further secure care had the longest PICU admission length. The best model included both patient type (discharge destination) and unit as well as an interaction between those variables.

Discussion Patient typing based on clinical pathways shows better predictive ability of admission length than clinical diagnosis or a specific tool that was developed to identify patient needs. Modelling admission lengths in a Bayesian fashion could be expanded and be useful within service planning and monitoring for groups of patients.

Conclusion Variables previously proposed to be associated with patient need did not predict PICU admission length. Of the proposed predictor variables, grouping patients by discharge destination contributed the most to length of stay in four different PICUs.

BACKGROUND

Psychiatric intensive care units (PICUs) admit patients in acutely disturbed phases of a severe mental disorder. Patients display increased risk to themselves or others that does not enable safe, therapeutic management within less secure conditions. Despite most admissions coming from acute psychiatric wards, over a third can come from the community or police custody.[1]

In England and Wales, PICU standards state admission length must be appropriate to clinical need and risk, but generally should not exceed 8 weeks.[2] Despite this, lengths of stay range from a day to over a year.[3] Some studies have shown association between diagnosis and admission length in acute psychiatric

WHAT IS ALREADY KNOWN ON THIS TOPIC
⇒ A specific mandatory tool has previously been developed to identify patient needs by clustering patient presentations.

WHAT THIS STUDY ADDS
⇒ Grouping of patient needs by clinical staff in psychiatric intensive care services is a better indicator of need (as measured by admission length) than the established tool.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY
⇒ Bayesian modeling can aid service monitoring and planning by considering different proposed predictor variables. By grouping patient need into categories, staff can develop more appropriate care pathway plans for patients.

services or PICU.[4 5] Although diagnosis can indicate what medical treatment is required, acute management and its duration also depend on severity, comorbidity, periodicity and other features.

A specific tool was developed to identify patient needs by clustering presentations according to the Mental Health Clustering Tool (MHCT) (NHS-England 2016).[6] This is based on the Health of the Nation Outcome Scale[7] and was put forward as a solution to assess local service needs.[8] MHCT usage has been questioned by some clinicians who described distinct clinical patient types within PICUs.[9–11] Patients with 'typical' needs are subsequently transferred to acute psychiatric wards or discharged. Those transferred to forensic psychiatry settings have different, specific needs, and those discharged to more specialised services have particular, complex requirements. Discharge destination is a measure of observed need, is crucial to care planning and something that clinical teams should be aware of early in a patient's admission (if not beforehand).

The primary objective of this study was to use statistical techniques to determine

Table 1 Predictor variables

Variable	Type of variable	Measurement values
Specific PICU	Discrete categorical	1, 2, 3, 4
Patient age	Discrete categorical	<26 years, 26–35 years, 36–45 years, 46–55 years, 56–65 years, >65 years
Patient gender	Discrete categorical	Male, Female
Service discharged to	Discrete categorical	Typical, Longer Low Secure, Other
Service admitted from	Discrete categorical	Typical, Longer Low Secure, Other
Primary diagnosis (ICD10 chapter)	Discrete categorical	F00-F09, F10-F19, F20-F29, F30-F39, F40-F48, F50-F59, F60-F69, F70-F79, F80-F89, F90-F98
MHCT cluster	Discrete categorical	1, 2, 3, 4, 5, 6, 7, 8, blank, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21

ICD10, International Classification of Diseases Tenth Edition; MHCT, Mental Health Clustering Tool ; PICU, psychiatric intensive care unit.

whether discharge destination is a useful predictor variable for length of stay (LoS) within PICUs.

METHODS
Patient types
A clinical reference group developed mutually exclusive PICU patient types based on where patients are discharged to:
1. *Typical*: adult acute psychiatric inpatient services or community settings.
2. *Longer secure care*: longer term secure environments (eg, Medium or Low Secure Units, 'Locked Rehabilitation' Units).
3. *Other*: other psychiatric settings (eg, older adult wards, mother and baby units, another PICU or other specialist services).

Predictor variables
The aim was to choose routinely recorded, easy to collect and clinically relevant data. Number of variables was minimised to ensure clinical usefulness (table 1).

Diagnosis was measured categorically using International Classification of Diseases (ICD) criteria.[12] Clusters and diagnoses may have some relation to length of PICU admission, but effects could interact with each other and/or with discharge destinations. It could be that *gender* influences time spent in a PICU and there may be different causes for agitated behaviour in the elderly compared with the younger population. *Age* was transformed into a categorical variable (following discussion within the reference group).

Outcome variable
LoS was measured in days. Clinical experience suggests that the distribution will be skewed by a minority of patients remaining for extended periods. To remedy this, admission lengths were transformed using natural logarithms. The number of days+1 was used when performing transformations to avoid potentially taking a logarithm of zero.

Method
When deciding sample size, previous research provided approximations for power analysis.[11] From that research, estimations were obtained using similar classifications described by that paper. The mean log_e (admission length$_{(days)}$ + 1) for a *typical* (type 1) admission was approximated to be 3.00, for a *longer secure care* (type 2) admission 4.25, and for the *other* (type 3) patient-type group 3.65. Overall variance was estimated at 1.2. Power analysis using these figures for a one-way analysis of variance (ANOVA) suggested that each group should contain 25 patients for a power of 0.95 at a 5% significance level.

In this study, attempts were made to collect data retrospectively from each PICU on 40 consecutive discharges of patients in each patient-type group. Due to circumstances surrounding COVID-19, the date of last discharge was taken to be 29 February 2020 (before the pandemic impacted on UK healthcare). Data collection was limited to a retrospective period of 4 years.

Bayesian models were developed that examined LoS distribution. Using standards and clinical experience, a prior for the mean admission, length was deemed to be 6 weeks (42 days). So log_e (*admission length*+1) = log_e (43) = 3.76. This was decreased to 3.75 for a prior.

In a normal distribution, heuristically nearly all observations lie within three SDs of the mean. If the minimum value is zero and assuming normality with a mean 3.75, $3\sigma = 3.75$, thus $\sigma = 1.25$. The 0.95th quantile of admission length = $e^{3.75+(1.96\times1.25)} = e^{6.2}$ = 492 days. Patients rarely spend more than a year in a PICU (but it is not unheard of), so 1.25 seemed an acceptable value to use for the log_e (*admission length*+1) SD weakly informative prior.

The **first model** allowed separate intercepts for each patient type (based on discharge destination):

$$patient\ type = i = 1,\ 2,\ 3$$

$$mean\left[log_e\left(length\ of\ stay + 1\right)_i\right] = \beta_i$$

Priors were:

$$log_e\left(length\ of\ stay + 1\right)_i \sim N\left(\beta_i,\ 1.25\right);$$

$$\beta_i \sim N\left(3.75,\ \sigma_e\right);$$

$$\sigma_e = half - Cauchy\left(2\right)$$

Similar models were developed using *diagnosis* (ICD-10 chapters) and *cluster* instead of patient type.

When estimating prior *distributions* for variance in hierarchical models, if the number of groups is less than 5, the *half-t* family of prior distributions is recommended.[13] Half-Cauchy priors were used for spread of variable means in this project. Based on initial data inspection, an appropriate prior (that allowed for flexibility) was decided to be a *half-Cauchy* with parameter 2 (a value slightly higher than expected for the SD of the underlying mean). Given uncertainty of what variables contributed most and in order to make priors extremely weak, it was decided to use this prior for an SD of all means in all proposed hierarchical models.

The first model can be improved on by considering the dependency between LoS for each unit. This was handled using mixed-linear modelling by specifying a by-unit intercept ($\alpha_{unit=u}$), thus allowing each unit to have a general level of variability. This strategy corresponds to the **second model**:

$$patient\ type = i = 1,\ 2,\ 3$$

$$unit = u = 1,\ 2,\ 3,\ 4$$

$$log_e\left(length\ of\ stay + 1\right)_i \sim N\left(\mu_i,\ 1.25\right)$$

Here, priors were:

$$\mu_i = \alpha_u + \beta_i;$$

$$\alpha_u \sim N\left(0,\ \sigma_u\right);$$

$$\sigma_u \sim half - Cauchy\left(2\right);$$

$$\beta_i \sim N\left(3.75,\ \sigma_e\right);$$

$$\sigma_e = half - Cauchy\left(2\right)$$

A **third model** adds a varying intercept for the interaction between *unit* and *patient type*:

$$log_e\left(length\ of\ stay + 1\right)_i \sim N\left(\mu_i,\ 1.25\right)$$

With priors:

$$\alpha_u \sim N\left(0,\ \sigma_u\right);$$

$$\sigma_u = half - Cauchy\left(2\right);$$

$$\alpha_{u:\ i} \sim N\left(0,\ \sigma_{u:\ i}\right);$$

$$\sigma_{u:\ i} \sim half - Cauchy\left(2\right);$$

$$\beta_i \sim N\left(3.75,\ \sigma_e\right);$$

$$\sigma_e = half - Cauchy\left(2\right)$$

Similar methods could be established to examine other factors (such as *diagnosis* or *clustering*) that could affect admission length. *Age* and *gender* variables were added to models in a hierarchical fashion in conjunction with the *patient-type* variable to assess whether there were interactions which improved the model.

Models gave flexibility by allowing error variance to change between patient types, units and interaction between the two. Partial pooling of information could account for uncertainty when estimating group-level effects and provide stable estimates with the aid of weakly informative priors.

Data analysis

Data analysis was performed using the statistical package R[14] with specific additional packages.[15–26]

Model generation

$Stan$[20] is a programming language for specifying statistical models and used as a Markov chain Monte Carlo sampler for Bayesian analyses. The *brms* package[15] allows R code to interface with *stan* and was used within this project. Four chains were used in every model simulation, each run with a warmup of 1000 out of 5000 iterations. For reproducibility, a seed of 123 was used.

Model verification and comparisons

Autocorrelations of parameter values with previous draws, measured by successive lags from each chain, were obtained. Graphical displays of posterior predictive checks were performed and Q–Q plots of residuals (comparing observed and expected residuals) were generated.

In Bayesian statistics, rather than considering a posterior point estimate, it is more useful to work with posterior distributions, $p_{post}(\theta) = p(\theta \mid y)$ and summarise predictive accuracy by using *log pointwise predictive densities* (lppd). This is calculated by evaluating the expectation using draws from posterior simulations, $p_{post}(\theta)$ and labelling θ^s, $s=1,2,\ldots,S$. The expected log pointwise predictive density (elpd-loo) is[26] : $\hat{Ippd} = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p\left(y_i|\theta^s\right)\right)$

Model comparisons were performed by applying leave-one-out cross-validation (LOO-CV) using training and validation data sets: observations were left out one at a time, so the training set used N-1 observations and the other observation as the validation sample. With LOO-CV, predictive accuracy is evaluated by first computing a pointwise predictive measure and then taking the sum of these values over all observations to obtain a single measure.[27]

The LOO Information Criterion (LOO-IC) (using the *loo* package within *brms*[25]) *estimates* expected lppd by integrating over uncertainty in the parameters and, thus, does not assume that the posterior distribution is multivariate normal.

Although individually, the elpd-loo and LOO-IC estimates have little intrinsic value, it is helpful to compare values between models. Difference is computed relative to the model with the highest elpd-loo/LOO-IC. When that difference is positive, the expected predictive accuracy is higher for the model with the largest positive difference. The model with the largest negative valued difference is the worst-performing model.

Table 2 Patient records used in data analysis

Unit	Patient type 1	Patient type 2	Patient type 3	Total
1	40	22	38	**100**
2	40	20	28	**88**
3	40	23	17	**80**
4	40	23	37	**100**
Total	**160**	**88**	**120**	**368**

RESULTS

From the four units involved, data were acquired from 368 patients (table 2).

Data from 40 type 1 (typical) patients were easily collected from each unit. Of the other types, amount collected from units over the maximum 4-year time period varied.

Survival plots for *patient types* are shown in figure 1A. These show lengthier admissions for the longer secure care group (*patient type 2*). Assuming an exponential distribution, hazard of discharge for this group is significantly lower (at the 5% level) than the other two groups. A log-rank test for differences (using *patient-type* 1 as the reference) gave a χ^2 value of 97.8 on two degrees of freedom ($p \leq 2 \times 10^{-6}$). Survival plots of patients from the different *units* are shown in figure 1B. These plots are closer together than the *patient-type* survival distributions
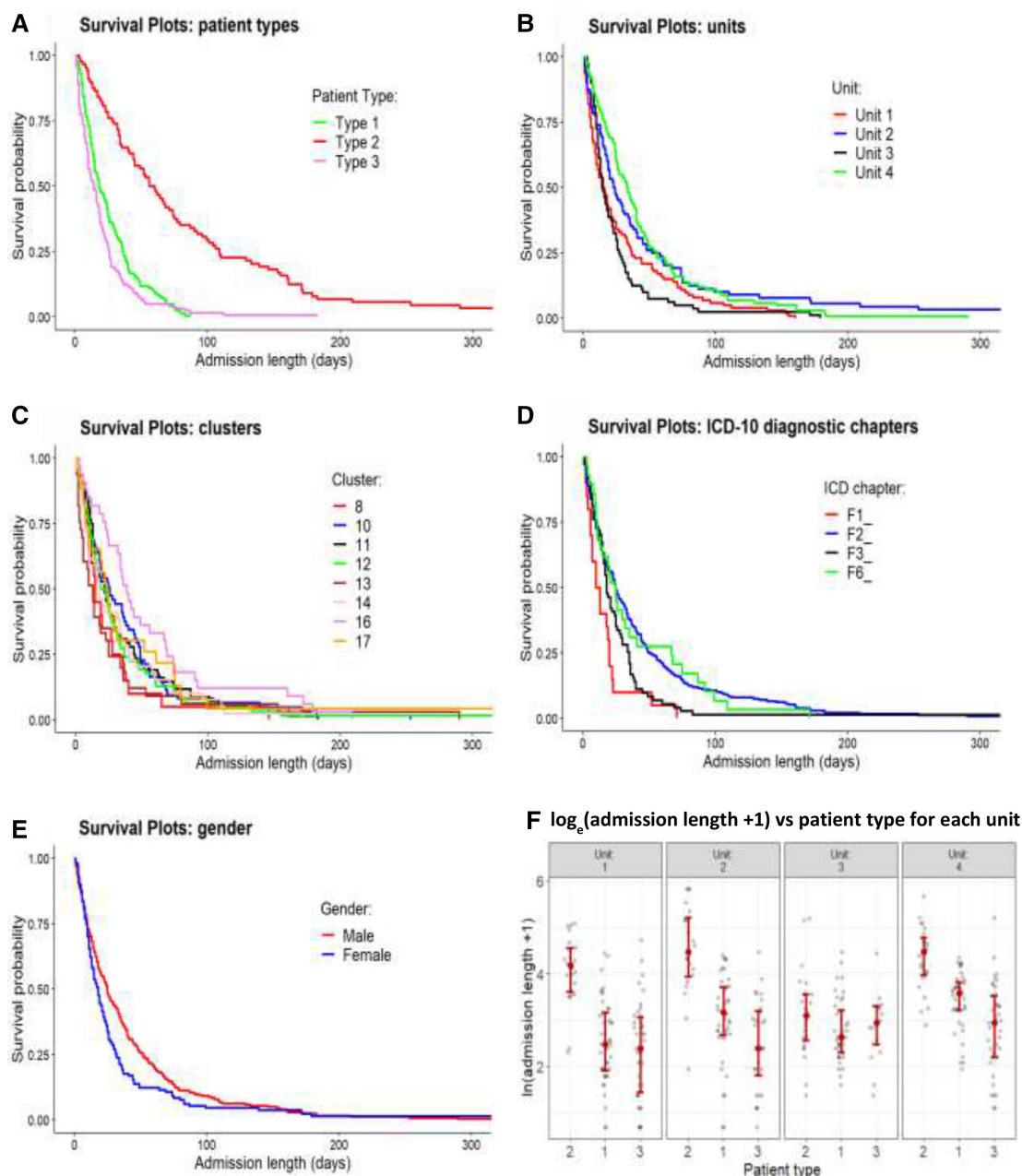

Figure 1 Plots of admission lengths and predictor variables (only groups containing at least 20 patients depicted).

but do indicate differences between *unit* distributions. Indeed, a log-rank test for differences (using unit 1 as reference) gave a χ^2 value of 20.6 on three degrees of freedom (p=0.001). As survival plots are initial inspection tools, to make them easier to view and to avoid doubts concerning the power of log-rank tests, only group subsets containing 20 or more patients were considered for other variables (figure 1C–E). The log-rank test for differences in survival distributions gave p=0.0007 for *ICD chapter* differences, p=0.1 for *cluster* and p=0.03 for *gender* differences.

Intra-class correlation (ICC) measures how similar outcomes of individuals within a group are likely to be, relative to those from other groups. Measurement is based on ANOVA, assuming normal distributions. Using log_e (admission length$_{(days)}$+ 1) as the outcome variable, the ICC for *patient type* was 0.37, *unit*=0.0585, *cluster*=0.0425, *ICD-10 chapter*=0.0259 and *gender*=0.014. None can be regarded as good, however, in the context of this data set, *patient type* was an order of magnitude better than any other.

ICC does not account for effect of interactions between groups, for example, the effect that *patient type* has may be confounded by the effect that *unit* has on admission length. Figure 1F shows how the *patient-type* variable varies with *unit*. This indicated that any model should include an interaction term between them.

From the three models described above, both elpd-loo and LOO-IC favoured model 3. Compared with model 1, the LOO-IC difference for model 2 was 25.5 and for model 3, 42.4. Differences in elpd-loo were 12.8 and 21.2, respectively. In models analogous to model 1, *patient type* was favoured above using either *ICD_10 chapters* or *clustering* measurements. There was an elpd-loo difference of 49.3 for patient type versus clustering and 50.5 versus ICD_10 chapter. LOO-IC differences were 98.5 and 100.9, respectively.

Adding further parameters to the model did not improve model comparison measures and there was some suggestion of model overfitting (with some models having a higher R^2 value but negative differences in elpd-loo and LOO-IC compared with model 3).
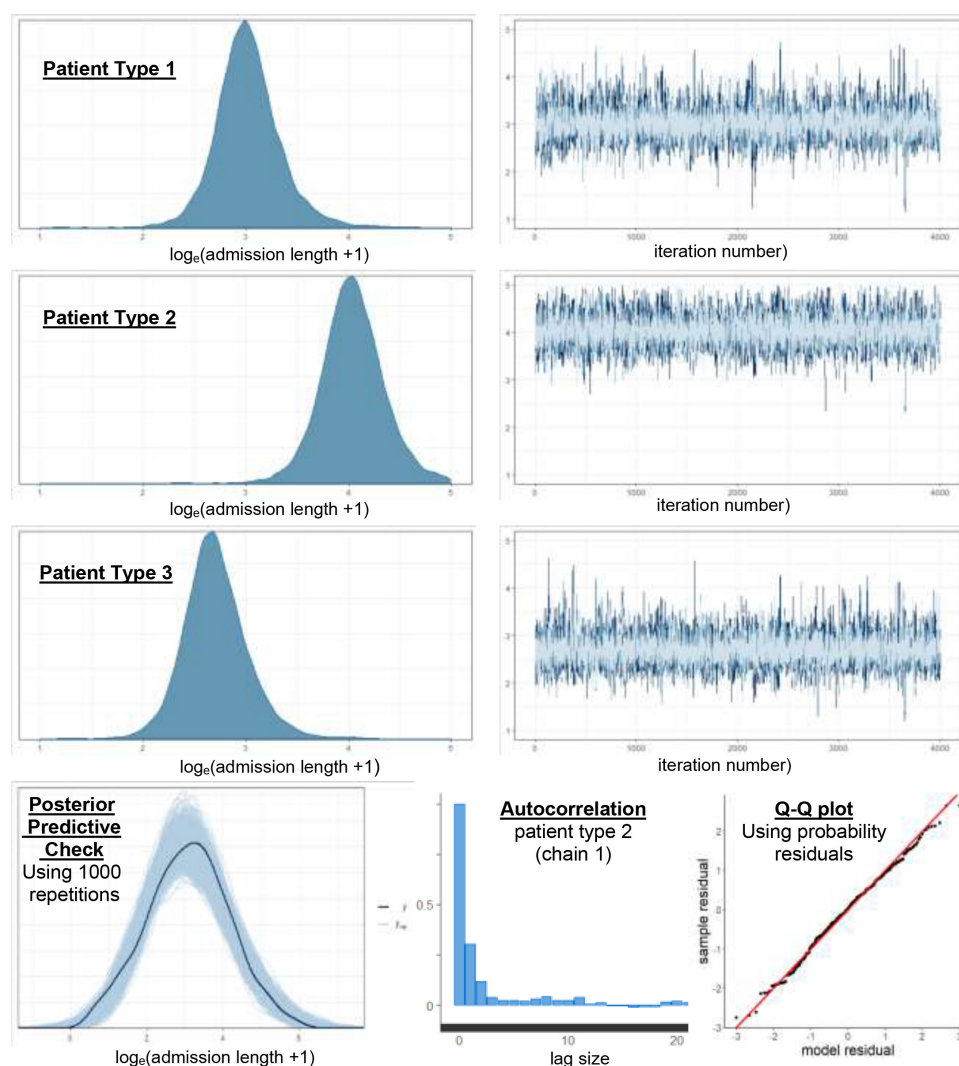

Figure 2 Results from model 3.

Table 3 Model 3—admission lengths

Patient type	Unit	Mean admission length (days)	0.025 quantile	0.975 quantile
1	1	12.1 (*12.1*)	0.9 (*1.7*)	71.2 (*52.3*)
1	2	20.1 (*20.0*)	2.3 (*4.1*)	125.5 (*90.7*)
1	3	15.1 (*15.2*)	1.3 (*2.4*)	90.8 (*65.5*)
1	4	30.0 (*29.8*)	4.4 (*6.4*)	201.4 (*141.9*)
2	1	55.8 (*55.5*)	7.4 (*10.2*)	319.5 (*247.7*)
2	2	81.0 (*80.7*)	13.3 (*17.3*)	522.2 (*392.5*)
2	3	25.3 (*25.6*)	2.6 (*3.9*)	146.9 (*112.8*)
2	4	74.2 (*73.6*)	12.5 (*16.1*)	501.7 (*383.6*)
3	1	9.7 (*9.4*)	0.6 (*1.1*)	58.1 (*44.0*)
3	2	11.4 (*11.5*)	1.1 (*1.8*)	74.2 (*56.2*)
3	3	14.6 (*14.7*)	1.2 (*2.2*)	91.8 (*70.2*)
3	4	16.1 (*16.0*)	2.1 (*3.0*)	108.9 (*86.3*)

Figures in brackets are from a modified model 3 with narrower priors (see Discussion).

The middle plot in the last row of figure 2 displays autocorrelation from chain 1 of model 3 (other chains are similar). It shows that correlation settled after 3–4 lags (ideally, it should be around zero from lag 1 onwards). It was decided that this was satisfactory and that no thinning or increase in draws was needed.[27]

Posterior predictive checks give the model's predictive distribution for a replication of *y*, denoted y_{rep}. The first plot in figure 2's final row shows posterior predictive distributions of 1000 draws from model 3. It shows a reasonable fit.

One method of plotting residuals, using the posterior predictive distribution, is to use what Kay has termed a probability residual.[28] Here, for each observation, the predicted probability of generating a value less than or equal to the actual observation is calculated: $p_i^{resid} = P\left(y_i^{rep} \leq y_i | y\right)$. If the predictive distribution is well calibrated, these probabilities should be uniform and if the inverse cumulative distribution function of the standard Normal distribution is applied to these probability residuals, the result should be approximately standard normal. These are quantile residuals (z_residual): $z_i^{resid} = F_{normal}^{-1}\left(p_i^{resid}\right)$ and the final plot in figure 2 shows a Q–Q plot of these residuals using model 3 to be acceptable.

Actual mean values can be calculated (noting the model gives values of $\log_e$(days admitted+1)). Table 3 displays means, 0.025 and 0.975 quantiles of actual admission lengths for types of patients from different units.

DISCUSSION

Results showed that patient typing based on clinical pathways has better predictive ability of admission length than clinical diagnosis or a specific tool developed to identify patient needs.

In one analysis of acute psychiatric admissions using ICD criteria and patient demographics, a model accounted for 15% of admission length variation.[29] This current project's less detailed and perhaps more pragmatic model accounted for over double that figure (37%). Although still not sizeable, this type of modelling could well give rise to useful avenues to pursue.

One question to be asked is whether correct data were collected and used appropriately? As always, balance needed to be made between pragmatism and level of detail. There may well be other clinical or non-clinical variables that are useful (eg, level of patient engagement with treatment, or severity and amount of behavioural disturbance, number of previous admissions), but the ones collected are routinely measured by clinicians, thus making it easier to compare different PICUs. One variable measured but not used in this project was from where an individual was admitted. To check this was not an 'important' variable, a classification tree method was used to ensure that it did not separate the data. As expected, the most important independent variables were *patient type* and *unit*. These were followed by *age* and *ICD chapter*. It seems reasonable to conclude that the variables collected were the easiest to obtain and most relevant for this project.

Models were developed hierarchically. Partial pooling of units was implemented. As well as defining intercept LoS parameters for each *patient type*, this model allows each *unit* to have a mean LoS associated with it that comes from a global distribution. Since the *units* are a random sample in themselves, interaction terms modelling the pattern of their effect on LoS from one *patient type* to another were expressed as random effects. There are then two levels of random effects: the b_i for *unit* and the b_{ij} for the *patient type* within each *unit*.

It may be that the choice of priors (especially for the SD of means) could have been better. This is reflected by wide intervals between 0.025 and 0.975 quantiles in table 3. Clinical judgement and a wish to have weakly informative priors gave rise to initial priors. When performing analysis using other priors, *brms* default settings (flat priors for *patient type* and student t distribution centred on zero with 3° of freedom and a 2.5 SD for the *sigmas*) gave rise to divergent transitions. However, more defined prior specifications with *patient type* means of N(3.0, 0.5), N(4.0, 0.75) and N(3.0, 0.75) ln(admission+1)$_{days}$, respectively, and *half-Cauchy(1)* distributions for σ_e, σ_u and $\sigma_{u: i}$ gave figures within brackets in table 3. This showed stability of means with noticeable decreases of widths between 0.025 and 0.975 quantiles. Therefore, in any further analysis, narrower priors should be considered.

Despite this limitation, it was satisfying to use a clinically pragmatic model that separated at least one of the patient types and had greater accuracy than more complicated models. Discharge planning is integral in patient management and it has transpired that discharge destination gives some indication of LoS. This could be used within service planning and monitoring for groups of patients.

Limitations

- Sampling over different time periods could have confounding effects: changes in unit practice or fabric may have occurred over the time needed to collect data for one specific group, compared with time needed for another.
- The data collection time period limit may have been too wide. This required balancing against numbers needed to perform effective analysis.
- This pragmatic project was designed to maximise clinician input. Unless explained appropriately, Bayesian analysis could potentially alienate clinicians.
- Prior specifications were too vague. However, the project provides a useful starting point for development of clinical patient types.

CONCLUSIONS

This study aimed to assess contribution of clinical patient typing to PICU admission length. Variables previously proposed to be associated with clinical severity and, therefore, time spent on PICUs were found not to be as useful as patient typing, which contributed most to admission length. The specific unit to which a patient was admitted also had influence on LoS. A further avenue to explore would be why units differ in distributions of admission lengths for specific patient types.

ORCID iDs
Stephen Dye http://orcid.org/0000-0002-0535-5332
Faisil Sethi http://orcid.org/0000-0003-4722-4549
Malcolm Campbell http://orcid.org/0000-0003-1710-9363
Koravangattu Valsraj http://orcid.org/0000-0001-6669-4365

REFERENCES

1 Brown S, Chhina N, Dye S. The psychiatric intensive care unit: a prospective survey of patient demographics and outcomes at seven english picus. *J Psych Intensive Care* 2008;4:17.
2 National Association of Psychiatric Intensive Care Units. National minimum standards for psychiatric intensive care in general adult services. 2014. Available: www.napicu.org.uk/wp-content/uploads/2014/12/NMS-2014-final.pdf
3 Pereira SM, Walker LM, Dye S, *et al*. National survey of psychiatric intensive care, low secure and locked rehabilitation services: NHS patient characteristics. *J Psych Intensive Care* 2021;17:79–88.
4 Crossley N, Sweeney B. Patient and service-level factors affecting length of inpatient stay in an acute mental health service: a retrospective case cohort study. *BMC Psychiatry* 2020;20:438.
5 Raaj S, Navanathan S, Matti B, *et al*. Admission patterns in a psychiatric intensive care unit in ireland: a longitudinal follow-up. *Ir J Psychol Med* 2021:1–8.
6 NHS England. Mental health clustering booklet (V5.0 2016/17). 2016. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/499475/Annex_B4_Mental_health_clustering_booklet.pdf
7 Wing J, Curtis RH, Beevor A. Health of the nation outcome scales (honos). glossary for honos score sheet. *Br J Psychiatry* 1999;174:432–4.
8 Monitor. Guidance on mental health currencies and payment. NHS England 2013 Publications Gateway Reference 00883;
9 Kearney T, Dye S. Lean thinking and more: development of patient needs types in psychiatric intensive care. *J Psych Intensive Care* 2010;6:57–63.
10 Kearney T, Dye S, Sethi F. Patient journeys within psychiatric intensive care: payments by results versus lean pathways development. *J Psych Intensive Care* 2013;9:4–11.
11 Dye S. Can mental health clusters be replaced by patient typing? *British Journal of Healthcare Management* 2017;23:229–37.
12 WHO. *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Switzerland, Geneva: World Health Organisation, 1993.
13 Gelman A. Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and draper). *Bayesian Anal* 2006;1:515–34.
14 R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available: www.R-project.org/
15 Burkner P. brms: an R package for bayesian multilevel models using stan. *J Stat Softw* 2017;80:1–28.
16 Wickham H, François R, Henry L, *et al*. Dplyr: A grammar of data manipulation. 2020. Available: https://dplyr.tidyverse.org
17 Fife D. Flexplot: graphically-based data analysis. *Psychol Methods* 2022;27:477–96.
18 Fernández-i-Marín X. **ggmcmc**: analysis of MCMC samples and bayesian inference. *J Stat Soft* 2016;70:1–20.
19 Wickham H. *Ggplot2: elegant graphics for data analysis*. Cham: Springer-Verlag, 2016.
20 Stan Development Team. RStan: the R interface to stan. 2020. Available: http://mc-stan.org/
21 Therneau TM. A package for survival analysis in R. 2021. Available: https://CRAN.R-project.org/package=survival
22 Ucar I, Smeets B, Azcorra A. Simmer discrete-event simulation for R. *J Stat Softw* 2019;90:1–30.
23 Wickham H. Modelr: modelling functions that work with the pipe. 2020. Available: https://CRAN.R-project.org/package=modelr
24 Komsta L, Novomestky F. Moments: moments, cumulants, skewness, kurtosis and related tests. 2015. Available: https://CRAN.R-project.org/package=moments
25 Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 2017;27:1413–32.
26 Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput* 2014;24:997–1016.
27 Link WA, Eaton MJ. On thinning of chains in MCMC: thinning of MCMC chains. *Methods Ecol Evol* 2012;3:112–5.
28 Kay M. Extracting and visualizing tidy residuals from bayesian models. 2021. Available: https://mjskay.github.io/tidybayes/articles/tidybayes-residuals.html
29 Newman L, Harris V, Evans LJ, *et al*. Factors associated with length of stay in psychiatric inpatient services in London, UK. *Psychiatr Q* 2018;89:33–43.