Modelling admission lengths within psychiatric intensive care units

Stephen Dye, Faisil Sethi, Thomas Kearney, Elizabeth Rose, Leia Penfold, Malcolm Campbell, Koravangattu Valsraj

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 (0.058)). 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.

In England and Wales, PICU standards state admission length must be appropriate to clinical need and risk, but generally should not exceed 8 weeks. Despite this, lengths of stay range from a day to over a year. Some studies have shown association between diagnosis and admission length in acute psychiatric settings or PICU. 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). This is based on the Health of the Nation Outcome Scale and was put forward as a solution to assess local service needs. MHCT usage has been questioned by some clinicians who described distinct clinical patient types within PICUs. 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
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. 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. 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, median was deemed to be 6 weeks (42 days). So $\log_e$ (admission length+1) = $\log_e$ (1) = 0. Therefore, the 0.95 quantile of admission length was estimated to be 3.00, 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 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, 3σ=3.75, thus σ=1.25. The 0.95 quantile of admission length = $e^{3.75+(1.645\times 1.25)} = e^{3.75+1.645} = 492$ 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):

$$\text{patient type } = i = 1, 2, 3$$

mean $[\log_e (\text{length of stay} + 1)] = \beta_i$

Priors were:

$\log_e (\text{length of stay} + 1) \sim N(\beta_i, 1.25)$

$\beta_i \sim N(3.75, \sigma_e)$

$\sigma_e = \text{half-Cauchy}(2)$

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

---

Table 1 Predictor variables

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

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. 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-dependency between LoS for each unit. This was handled using hierarchical models. Here, priors were:

\[
\alpha_i \sim N(0, \sigma_u) ;
\]

\[
\sigma_u \sim \text{half-Cauchy}(2) ;
\]

\[
\beta_i \sim N(3.75, \sigma_e) ;
\]

\[
\sigma_e = \text{half-Cauchy}(2)
\]

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

\[
\log_e (\text{length of stay} + 1)_{i} \sim N(\mu_i, 1.25)
\]

With priors:

\[
\alpha_u \sim N(0, \sigma_u) ;
\]

\[
\sigma_u = \text{half-Cauchy}(2) ;
\]

\[
\alpha_{u, i} \sim N(0, \sigma_{u, i}) ;
\]

\[
\sigma_{u, i} = \text{half-Cauchy}(2) ;
\]

\[
\beta_i \sim N(3.75, \sigma_e) ;
\]

\[
\sigma_e = \text{half-Cauchy}(2)
\]

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 with specific additional packages.

**Model generation**

**Stan** is a programming language for specifying statistical models and used as a Markov chain Monte Carlo sampler for Bayesian analyses. The **brms** package 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, \( p_{post}(\theta) = p(\theta | y) \) and summarise predictive accuracy by using \( \log \text{pointwise predictive densities} (\text{lpdp}) \). This is calculated by evaluating the expectation using draws from posterior simulations, \( p_{post}(\theta) \) and labelling \( \theta^* \), \( s=1,2,\ldots,S \). The expected log pointwise predictive density (elpd-loo) is:

\[
\text{elpd} = \sum_{s=1}^{S} \log \left( \frac{1}{S} \sum_{i=1}^{n} p(y_i | \theta^*) \right)
\]

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.

The **LOO Information Criterion (LOO-IC)** (using the **loo** package within **brms**) estimates expected lpdp 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.
RESULTS

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

<table>
<thead>
<tr>
<th>Unit</th>
<th>Patient type 1</th>
<th>Patient type 2</th>
<th>Patient type 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>22</td>
<td>38</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>20</td>
<td>28</td>
<td>88</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>23</td>
<td>17</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>23</td>
<td>37</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>160</td>
<td>88</td>
<td>120</td>
<td>368</td>
</tr>
</tbody>
</table>

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 $\chi^2$ 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.
but do indicate differences between unit distributions. Indeed, a log-rank test for differences (using unit 1 as reference) gave a $\chi^2$ 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](image_url)  
**Figure 2** Results from model 3.
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.

Posterior predictive checks give the model's predictive distribution for a replication of \( y \), denoted \( y_{\text{rep}} \). 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. Here, for each observation, the predicted probability of generating a value less than or equal to the actual observation is calculated:

\[
P_{i}^{\text{pred}} = P\left(y_{rep} \leq y_{i}\right)
\]

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 (\( z_{\text{resid}} \)):

\[
z_{i}^{\text{resid}} = F_{\text{normal}}^{-1}\left(P_{i}^{\text{pred}}\right)
\]

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 \( \log(\text{days admitted}+1) \)). 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. 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 \( b_j \) 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 sigma) 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) gave rise to useful avenues to pursue.

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.
This pragmatic project was designed to maximise clinical patient types. 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.

Acknowledgements
This study formed part of an MSc project performed by SD with supervision from Dr Kevin Waters, Sheffield University. Rod Jones gave wonderful suggestions upon previous work, James Fitton (Mental Health Strategies) helpful encouragement and advice and Sai Kunphna (NHS&I Demand and Capacity Programme) an overview on aspects to consider.

Contributors
SD, TK and FS conceived this study. SD designed the study and acts as guarantor. Data were collected by SD, LP, ER and MC. FS, KV and TK oversaw governance. SD drafted the first version of the manuscript. All authors provided revisions for intellectual content and approved the final manuscript.

Funding
The authors have not declared a specific grant for this study from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests
None declared.

Patient consent for publication
Not applicable.

Ethics approval
This study used four, 10-bedded PICUs: two from Norfolk and Suffolk NHS Foundation Trust (NSFT) and two from South London and Maudsley NHS Trust (SLaM). The lead NHS organisation (NSFT) approved the study as service evaluation (registration: 2020M113-SE) and the University of Sheffield approved the study from an ethical perspective (registration: 190205032).

Provenance and peer review
Not commissioned; externally peer reviewed.

Data availability statement
Data are available upon reasonable request. R code and data text available upon reasonable request.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Stephen Dye http://orcid.org/0000-0002-0535-5332
Faisal Sethi http://orcid.org/0000-0003-4722-4549
Malcolm Campbell http://orcid.org/0000-0003-1710-9363
Koravangattu Valara http://orcid.org/0000-0001-6668-4365

REFERENCES
13 Gelman A. Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and draper). Bayesian Anal 2006;1:35–44.