Biostatistics Collaboration of Australia – Masters in Biostatistics

Workplace Project Portfolio

The MADbeds Project – The influence of bed positioning on delirium risk in the Alfred Hospital Intensive Care Unit

Richard Hardy

PREFACE

Overview

This Workplace Project Portfolio consists of a single unit project spanning July to November 2016.

The project was carried out with researchers from the Alfred Hospital, and Monash University's Department of Epidemiology and Preventive Medicine's Biostatistics Unit. It utilised data from the Alfred Hospital's Intensive Care Unit, curated by Dr David Pilcher of the Alfred Hospital, and provided to me as a Stata dataset by Dr Jessica Kasza of Monash University. Additional data on the Alfred Hospital ICU beds was provided by Dr Natalie Apelbaum of the Alfred Hospital. This report outlines my contribution to the statistical analysis, made under the supervision of Dr Kasza, who oversaw my analysis through all phases, from data manipulation and selection, to reporting of the final statistics and generating this report.

The project objectives focussed on the Alfred Hospital Intensive Care Unit (ICU) and compared the effects of each of its 45 ICU beds on the development of delirium for the patients treated in them, seeking to determine if there were any beds that had unusually low or high proportions of patients who developed delirium. This was achieved through the generation of statistics analogous to Standardised Mortality Ratios (SMRs) used in the comparison of hospital performance. The analysis produced Standardised "Madness" Ratios (SMRs for this context), which were based on a binary outcome indicating a clinical diagnosis of delirium for each patient during time spent in an ICU stay. SMRs were produced by hierarchical logistic regression models controlling for patient-level factors and treating each bed as a cluster, and were used to summarise the effects of each bed on the risks of occupant patients developing delirium.

Student's role

My role was to ultimately produce and present these SMRs using the dataset and following the background work of the Alfred Hospital and Biostatistics Unit teams. This involved:

- performing exploratory analysis of the dataset provided, and based on this initial analysis, performing the data management steps required to further clarify the sample for analysis.
- building a small range of hierarchical logistic regression models utilising patient-level and bed-level factors to predict delirium utilising a pre-selected group of clinically chosen candidate variables.
- comparing these models and generating a suite of SMRs.

- generating bootstrap confidence intervals for the predicted SMRs, modifying a Stata program written by Dr Kasza for a similar purpose.
- reporting and interpreting the final results.

Teamwork

In my role I collaborated primarily with Dr Kasza, who provided guidance in the step-by-step completion of the data analysis tasks outlined above, as well as updated data from the other members of the team. Dr Kasza also provided me with detailed feedback on my work, giving advice and direction throughout each stage of the project. We held meetings over the course of the semester, where we consolidated work to a certain stage and clarified the steps to come. These aided me immensely in my understanding of new statistical concepts that built on that of earlier subjects, notably Longitudinal and Correlated Data, and allowed me to contribute steadily to the project. Involvement in this project as a member of a team invested in its outcome has been a very rewarding experience!

Reflections on Learning

Communication skills

This project was helpful in improving my written and verbal communication skills, particularly in presenting and discussing statistical methods and results that I was previously unfamiliar with. It gave me the opportunity to improve upon my accuracy in the use of terms and concepts taught throughout the BCA course, and to improve towards succinctly presenting work in a clinically meaningful and accurate way.

Work Planning

I gained an insight into the stages involved in a relatively complex project in practice. As this was my first application beyond subject assignments, I was grateful for the experience of my supervisor in setting guidelines for the completion of the different stages of the project, and for providing a comprehensive background to the initial datasets used in the analysis. Working through it has given me a clearer idea of the time-frames and planning required for a research question to be addressed in practice. A key takeaway was that preliminary analysis and comprehensive data management is extremely important in allowing for the relevant techniques to stand up to their underlying assumptions, and to be accurately used to address a research question.

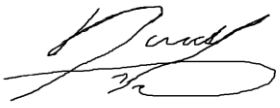
Statistical Principles, Methods, and Computing

Among the new methods introduced in this project were: building prediction models, bootstrap re-sampling to generate standard errors and confidence intervals, SMRs, working with programs within Stata, and comparing hierarchical logistic regression models. The majority of the analysis was performed using Stata, and this has given me a deeper understanding of its use, building on the data management practices covered in Data Management and Statistical Computing. In particular, I have improved on my ability to overcome issues with editing and sorting data of different kinds, merging them accurately, and outputting data. I have also built upon my abilities to perform exploratory analyses and make statistical decisions, generate loops and modify programs in a programming language, deal with missing values in context, understand multilevel models and create a reproducible analysis utilising each of these elements.

Ethical Consideration

The MADbeds project was approved by Dr David Pilcher, with ethical approval granted to Dr Pilcher by the Alfred Hospital.

The MADbeds Project – The influence of bed positioning on delirium risk in the Alfred Hospital Intensive Care Unit

Location and Dates:	Monash University Department of Epidemiology and Preventive Medicine: Biostatistics Unit., Alfred Hospital, Melbourne, VIC July 2016 - November 2016
Context:	This project was carried out with researchers from the Alfred Hospital, and Monash University's Department of Epidemiology and Preventive Medicine's Biostatistics Unit, utilising data on the Alfred Hospital Intensive Care Unit (ICU). ICU staff had suspected that patients assigned to one particular ICU bed seemed to consistently develop delirium, which motivated the present study. It sought to determine whether there were any beds within the Alfred Hospital's ICU that had unusual levels of delirium. Dr Jessica Kasza gave statistical and content-matter supervision for my work on this project.
Student Contribution:	- Data management: cleaned dataset and defined sample for analysis. - developed hierarchical logistic regression models - generated statistics analogous to standardised mortality ratios (SMRs). - generated their standard errors using bootstrapped datasets
Statistical issues involved:	- defining a sample for analysis - hierarchical logistic regression - comparing model performance - SMRs
Declaration:	I declare this project is evidence of my own work, with direction and assistance provided by my project supervisor. This work has not been previously submitted for academic credit.
Student Name:	Richard Hardy
Student Signature:	

Supervisor's Statement:	The objective of this project was to determine whether there were any beds within the Alfred Hospital's ICU that had unusual levels of delirium. This project was challenging, and required the application of many different statistical techniques, many of which Richard was unfamiliar with at the start of this project. Richard rose to the challenge admirably: generating a clean dataset ready for analysis; developing a prediction model for delirium; fitting hierarchical logistic regression models; getting to grips with standardised mortality ratios; and finally, bootstrapping to generate standard errors. Richard demonstrated the ability to work independently, manage timelines and articulate questions well, apply these complex statistical techniques, and write up his results. He has been a very dedicated student.
Supervisor name:	Jessica Kasza
Supervisor Signature:	

PROJECT REPORT

1. Introduction

Delirium is the most commonly found psychiatric syndrome in the general hospital setting [1], and is the subject of much recent investigation within ICU environments. It is characterised by impaired cognitive function and consciousness and is a source of concern as a negative prognostic indicator of morbidity, mortality and hospital length of stay [2]. The incidence of delirium in ICU environments has varied widely, ranging between 15-80% [3]. There exist many tools developed for diagnosing delirium, including a four-feature confusion assessment method for the ICU (CAM-ICU) developed by Ely and colleagues [4] and the NEECHAM Confusion Scale, both of which are widely used and have been shown to be valid and reliable [5].

The ICU environment itself is a potential source of modifiable factors that may exacerbate the onset of delirium ultimately leading to poorer outcomes for ICU patients. These notably include noise and light, both of which have been extensively studied in the context of critical care. Noise levels in ICU units frequently exceed WHO and EPA recommendations [1,6] and excessive levels have been found to be at least partly responsible for sleep fragmentation affecting normal sleep-wake cycles [6] with poor sleep strongly associated with an increased risk of delirium. As a result, multifaceted strategies that directly reduce noise and light to promote better sleep have been included in clinical guidelines to reduce the risks of ICU delirium [7].

The focus of this study is on comparing this risk of delirium between beds in the Alfred Hospital's ICU, after adjusting for patient-level and bed-level factors, in order to identify the pattern of delirium risk within the ICU. Bed-level factors included bed noise and light characteristics, included in the analysis to control for noise and light which potentially contribute to delirium risk in ICU settings.

2. Methods

2.1 Data

Data for this project were provided by the Monash University Biostatistics Unit and sourced primarily from The Australian and New Zealand Intensive Care Society Adult Patient Database (ANZICS APD), which is under continual review and development, and is managed by the ANZICS CORE management committee [8]. An initial 'patient-level' dataset was provided, which contained information on patient-level characteristics. In addition, a 'bed-level' dataset containing information on noise, light and location information for each of the 45 beds in the ICU was provided in excel format and merged later in the analysis.

The patient-level data included 148 variables describing patients admitted to the Alfred ICU between July 1st 2012 and June 30th 2015. The ANZICS APD includes comprehensive information on diagnosis, ICU and hospital admission dates/times, Glasgow Coma Scale (GCS) scores [9, 10], Acute Physiology, Age, Chronic Health Evaluation (APACHE III) scores [11, 12], ventilation status, and patient demographics such as age, gender, and smoking behaviour. This constituted the bulk of the patient-level information. Delirium outcome was recorded as a binary variable for each patient, indicating whether there was a diagnosis of delirium at any time during a stay in the ICU.

The bed-level data included measures of light levels for each bed, noise exposure, and an indication of whether beds were isolation units, or in a room with or without a door. Light measurements were taken with electric lights on or off, for each of four directions from the head of each bed: ahead, back, left and right. Means of these light measures were calculated under both lighting conditions, with the lights off measures quantifying levels of natural light. Noise exposure was described by variables that indicated which beds were located near busy areas of the ICU. Specifically, these included indications for whether a bed was located near exits, the pharmacy, or the nurse handover station.

2.2 Data Management

The first objective was to produce a cleaned dataset in order to build a predictive model that could take account of correlation structure within the original dataset and accurately predict delirium. Delirium was diagnosed as a binary measure of disease, hence logistic regression was chosen as a method well suited to developing a model for its prediction [13]. Furthermore, this could also be expanded to a hierarchical model to allow for the correlation of outcomes for patients assigned to the same bed.

Initial investigation involved determining how many patients had multiple ICU admissions; how many beds each patient spent time in over the course of a single ICU stay; and what proportion of each stay each patient spent in each bed. To define the sample for analysis, and address these questions, exploratory analysis of the patient-level dataset was performed using Stata (Version 13.1). This analysis is described below.

Patients may have been admitted to the ICU multiple times over the three-year data collection period. These multiple admissions may have occurred during one or more hospital admissions. Multiple ICU admissions for patients induced a within-patient correlation structure, as delirium outcomes for the same patient during separate ICU stays are expected to be correlated. The cleaned dataset was sorted by patient ID and subsequently by ICU admission date, with a variable generated

that corresponded to the count of ICU admissions for each patient over the study period. Based on the distribution of these counts, the first admission within the first hospital stay was retained, with any subsequent admissions discarded. The remaining questions were then addressed with analysis restricted to the first ICU admission per patient.

Patients may be moved between beds as their needs change, or because of operational needs of the ICU, leading to patient movement within the ICU. With data in wide form (i.e. with one observation per subject), binary variables were created for each of the 45 bed variables to flag instances of time spent in each bed. These were then counted to determine the number of beds patients spent time in. For example, a hypothetical patient who on their first ICU admission spent time in bed 23, and was then moved to bed 36, received a 1 for a measured time greater than zero in bed 23, a 1 for a measured time greater than zero in bed 36, and a 0 for all remaining beds, resulting in a bed count of 2 for this patient.

Patients with missing bed times, and those who weren't able to be identified as having spent time in any bed were excluded. This produced a dataset containing observations of the first ICU admission of independent patients who were treated in at least one ICU bed during their admission.

A significant challenge to overcome was to determine how to build a prediction model for delirium risk that utilised data from patients who spent time in more than one bed. A solution was to again focus analysis on a subset of the data, by assigning each patient to a single bed.

Assignment to a single bed required understanding how patients moved through the beds, and would be a suitable simplification of the data if most patients had spent most of their stay in one bed. If most patients moved between multiple beds throughout their stay, then it would not be reasonable to assume that any case of delirium observed was associated with a stay in a specific bed, since the delirium outcome variable recorded delirium at any time during a stay in the ICU. To fully account for patients moving between beds, a more complex multiple membership model could be applied [14,15], however this was beyond the scope of this project.

For each patient, a total ICU time variable was calculated to correspond to the total sum of their time in each bed. Beds were then ranked from longest occupancy time to shortest, and the bed in which patients spent the greatest proportion of their total ICU stay was defined as a patient's 'first-ranked' bed, subsequently defined as their 'dominant bed'. This produced the final dataset for prediction modelling, containing observations of patients' first ICU stay, and including one dominant bed per stay. Time spent in beds other than the dominant bed was thus not considered for analysis.

2.3 Developing predictive models for delirium

The final dataset was split into an estimation set comprising 80% of the observations, and a validation set comprising the remaining 20%. Models were fitted on the 80% set and internally validated using the 20% set [16]. Three prediction models for delirium were then developed: the first did not include any patient or bed-level predictors (the "null" model (1)); the second included patient-level predictors only (2); and the third included patient- and bed-level predictors (3).

The first of these, the null logistic random effects model, is described as follows.

With the outcome of interest Y representing the binary indicator of delirium, and with patients $i = 1,2,\dots,n_j$ clustered in their dominant bed j;

Let:

$$Y_{ij} = \begin{cases} 1 & \text{if patient } i \text{ with dominant bed } j \text{ developed delirium during their first ICU admission} \\ 0 & \text{otherwise} \end{cases}$$

$$i = 1,\dots,n_j \ , \qquad j = 1,\dots,45.$$

Letting a random coefficient U_j represent the random effect of bed j, the null model is given by:

$$Y_{ij}|U_j \sim \text{Bernoulli}\left(P_{ij}\right), \ \log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \alpha + U_j, \ \ U_j \sim N\left(0,\sigma_j\right) \qquad (1)$$

where P_{ij} represents the probability of patient i in bed j developing delirium, and therefore its logit represents the log-odds of patient i in bed j developing delirium during their first ICU admission. α represents the model intercept, and U_j the random effect of bed j. Under the null model, estimation of $\alpha + U_j$ gives the estimated null log-odds of patients in bed j developing delirium. U_j is assumed to be normally distributed with mean 0, with σ_j to be estimated. Explicitly modelling these random effects takes account of some of the expected variation between the delirium outcome in each bed.

Patient-level factors considered for inclusion in the patient-level model and patient-bed-level model were age, the Verbal, Motor and Eye Opening components of total GCS score (each included separately), APACHE III score with age and GCS points removed, sex, diagnosis category (the grouped APACHE III diagnoses), ICU admission source (the mechanism by which patients were admitted to the ICU, represented by a code), ICU admission type (either elective or emergency), smoking status (a description of current and past smoking behaviour, represented by a code), the (natural log of) time spent in the dominant bed, English speaking ability (0/1), and an indication of time spent under invasive ventilation (0/1).

Variables treated as continuous were centred prior to their inclusion in the models to ensure baseline values were in the range of the measured data and could be meaningfully interpreted. The natural log of the measure used for time in the dominant bed was taken to improve its fit and interpretation, with its baseline value corresponding to 1 hour spent in the dominant bed.

The patient-level model adjusting for this patient-level risk is given by:

$$Y_{ij}|\boldsymbol{X_{ij}}, U_j \sim \text{Bernoulli}(P_{ij}),\ \log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \alpha + \boldsymbol{\gamma'X_{ij}} + U_j,\ \ U_j \sim N(0, \sigma_j) \qquad (2)$$

$\boldsymbol{\gamma'}$ represents the vector of patient-level coefficients for $\boldsymbol{X_{ij}}$, the vector of patient-level covariate values for patient i in bed j. U_j represents the random effect of bed j, normally distributed around mean 0 with standard error σ_j to be estimated, adjusted for the patient-level factors included in the model. In this case, the log odds of patient i in bed j developing delirium is predicted by the sum of the intercept, the patient-level factors included, and the random effect of bed j remaining after adjusting for the patient-level factors.

Bed level factors considered for inclusion in the patient-and bed-level model were: door type (one of curtains only, isolated room, door which can be closed); a binary indication of a bed being near a 'busy area' (if near nurse handover and/or pharmacy and/or by exits); and the mean of the natural light level measurements taken with electric lights off (centred at its mean across the 45 beds). All three factors were included in a model termed the "patient-bed-level model". This was fitted with and without two-way interactions and a three-way interaction between the three bed-level factors.

This model including both patient-level and bed-level factors (the "patient-bed-level model") is given by:

$$Y_{ij}|\boldsymbol{X_{ij}}, \boldsymbol{X_j}, U_j \sim \text{Bernoulli}(P_{ij}), \quad \log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \alpha + \boldsymbol{\gamma'X_{ij}} + \boldsymbol{\beta'X_j} + U_j, \quad U_j \sim N(0, \sigma_j) \qquad (3)$$

α represents the model intercept, $\boldsymbol{\gamma'}$ represents the vector of patient-level coefficients for $\boldsymbol{X_{ij}}$, the vector of patient-level covariate values for patient i in bed j. $\boldsymbol{\beta'}$ represents the vector of bed-level coefficients for $\boldsymbol{X_j}$, the vector of bed-level covariate values for bed j. U_j represents the random effect of bed j, normally distributed around mean 0 with standard error σ_j to be estimated, adjusted for both the patient-level and bed-level variables included in the model. In this case, the log odds of patient i in bed j developing delirium is predicted by the sum of the intercept, the patient-level and bed-level factors included, and the random effect of bed j remaining after adjustment for all patient and bed level factors.

In each of the three models, the U_j random effect term accounts for the clustering of patients within their dominant beds, thus accounting for the random variation in delirium expected in each model due to risk factors at both patient and bed level that were not accounted for in the model's specification.

The fit and performance of each model was then compared using AIC, BIC, area under the receiver operating curve (AU ROC), and p values for the Hosmer-Lemeshow $\hat{C}$ statistic (H-L p value) [17]. Models were internally validated using the 20% validation set reporting ROC and H-L p value. The fit of models (2) and (3) , and of the continuous variables included in these models, was also assessed using binned residual plots [18, 19].

2.4 Standardised 'Madness' Ratios

Standardised Mortality Ratios are used routinely in the comparison of mortality outcomes between health care providers [20]. They are defined as the ratio of observed to expected deaths for a given hospital. In the present study, delirium experience is measured using a binary outcome and compared between beds, an analogous situation to mortality experience compared between hospitals, in that a binary outcome measure is used in both cases, with clustering of patients within higher level units. Hence, a Standardised 'Madness' Ratio (SMR) is defined as the ratio of observed to expected delirium cases for a given bed. The expected number of cases of delirium is estimated using the predicted number of cases of delirium obtained from each of the 3 statistical models in turn. If the SMR for a given bed is significantly <1, then that bed has a favourable delirium experience, a ratio of >1 is unfavourable, and a ratio of 1 is neutral.

Four SMRs were generated using the methods of Mohammed et al [21]. Each of these four SMRs was generated for each of the three delirium prediction models: the null model, the patient-level model, and the patient-bed-level model, for a total of 12 SMRs. Constructing the range of models provided SMRs that gave three levels of adjustment and therefore three predicted delirium probabilities for each patient, which could be compared to ascertain the reasons for any discrepancies between them.

Formulations for SMRs generated using the patient-level model are shown as follows: formulations for the SMRs generated for the other models are similar. These adjust for the patient-level risk factors included as the coefficient-covariate matrix $\boldsymbol{\gamma}'\boldsymbol{X_{ij}}$ in (2) .

Patient-level SMR 1:

$$plSMR1_j = \exp(U_j) \qquad (4)$$

The exponential of the random effects gives the estimated odds of delirium due to the bed, similar to the observed/expected measure [20]. SMR1 reflects the odds of patients developing delirium in bed j compared to the counterfactual odds of developing delirium in an average bed [21].

Patient-level SMR2:

With y_{ij} as an observation of Y_{ij} ($y_{ij} = 1$ if patient i with dominant bed j developed delirium, 0 otherwise),

$$plSMR2_j = \frac{\sum_{i=1}^{n_j} y_{ij}}{\sum_{i=1}^{n_j}\left(1 + \exp\left[-\alpha - \boldsymbol{\gamma}'\boldsymbol{X_{ij}}\right]\right)^{-1}} \qquad (5)$$

The observed bed-specific number of delirium cases is given by $O_j = \sum_{i=1}^{n_j} y_{ij}$. This is the number of patients who developed delirium who were assigned the same dominant bed j.

The fixed-effects coefficients of the model are utilised to provide the expected number of delirium cases for bed j, given by $\sum_{i=1}^{n_j}\left(1 + \exp\left[-\alpha - \boldsymbol{\gamma}'\boldsymbol{X_{ij}}\right]\right)^{-1}$. Thus plSMR2 describes the ratio of observed cases associated with a bed, to that of the expected cases in an average bed as predicted by the patient-level model (where the random effects U_j account for the between-bed variation and therefore are taken to be equal to zero, thus not contributing to the prediction of the expected cases of delirium for the 'average bed' [22]).

Patient-level SMR3:

From the sum $\alpha + U_j$ (corresponding to the log-odds of delirium in the j^{th} bed after risk adjustment [23]), the probability of delirium for each bed is obtained, p_j. The mean of these terms (i.e. the overall intercept α) is used to estimate the overall probability of death, p_0. The relative risk ratio of p_j/p_0 is then used to derive the SMR, referred to as plSMR3 here, which is estimated for patients at $\boldsymbol{X_{ij}} = \mathbf{0}$. For the j^{th} bed, plSMR3 is given by:

$$plSMR3_j = \frac{\left(1 + \exp\left[-\alpha + U_j\right]^{-1}\right)}{1 + \exp[-\alpha]^{-1}} = \frac{[1 + \exp(\alpha)]\exp(U_j)}{1 + \exp(\alpha + U_j)} \qquad (6)$$

With $\boldsymbol{X_{ij}} = \mathbf{0}$, plSMR3 compares the risk of delirium between beds for an 'average' patient (that is, a patient with a central age and APACHE III score, a profile of categorical variables set at their baseline category, who has spent 1 hour in a bed).

Patient-level SMR4:

SMR4 is formulated in a similar manner to SMR2, with the exception that the values in the numerator corresponding to the observed cases of delirium are replaced by model-based predictors of observed cases of delirium [20]. This accounts for some of the sampling variation in observed cases in a bed. With $\sum_{i=1}^{n_j}\left(1+\exp\left[-\alpha-\boldsymbol{\gamma}'\boldsymbol{X_{ij}}-U_j\right]\right)^{-1}$ thus corresponding to the 'smoothed', predicted number of delirium cases, plSMR4 is given by:

$$plSMR4_j = \frac{\sum_{i=1}^{n_j}\left(1+\exp\left[-\alpha-\boldsymbol{\gamma}'\boldsymbol{X_{ij}}-U_j\right]\right)^{-1}}{\sum_{i=1}^{n_j}\left(1+\exp\left[-\alpha-\boldsymbol{\gamma}'\boldsymbol{X_{ij}}\right]\right)^{-1}} \qquad (7)$$

The preceding formulations describe the patient-level SMRs. SMR formulations for the null and patient-bed-level models are given in Appendix 1.

Standard errors for each SMR type were generated in Stata utilising the bootstrap method [24] for each model refit to the full patient sample. For SMRs generated using the null and patient-level models, standard errors were based on 100 bootstrap datasets, while for SMRs generated using the patient-bed-level model, standard errors were based on 50 bootstrap datasets (due to the greater complexity of this model leading to excessive program run-times). Patients with missing information for GCS score, sex and smoking status were excluded prior to bootstrapping. Confidence intervals for the SMR for each of the 45 beds were then calculated assuming normality of the SMRs.

3. Results

3.1 The data set

The initial data set consisted of 8,348 observations from 7,358 patients admitted to the Alfred Hospital ICU between July 1st 2012 and June 30th 2015. 18 of these were duplicate observations. Fourteen patients with duplicate observations were identified, 2 of which had 3 duplicates, and 12 of which had 1 duplicate. These 18 duplicate observations were excluded from further analysis. Results of the initial data management steps are summarised in Table 1.

Of the 7,358 patients comprising the dataset, 750 (10%) were admitted multiple times to the ICU over the 3 year study period (Table 2). Within a single ICU admission, a given patient may have been

Table 1: Summary of management steps; arriving at a final dataset for model construction

Data	Observations	Changes	
Original set	8,348	--	--
1st ICU Admission set	7,358	18 Duplicates excluded	972 Secondary admissions excluded
Final set	7,161	197 Unassignable admissions excluded	Patients assigned dominant beds

Table 2 : Counts and percentages of patients with each number of ICU admissions.

ICU admissions	Patients	% of all patients
1	6,608	89.81
2	596	8.10
3	111	1.51
4	28	0.38
5	9	0.12
6	3	0.04
7	2	0.03
8	1	0.01
Total	7,358	100

moved between beds over the course of the ICU stay. 45 of the 148 patient-level variables were records of length of stay, in hours, in each of the 45 beds. Records of length of stay in beds were given for each ICU admission, for a given patient. However, for patients who spent time in more than one bed, the bed in which delirium was diagnosed was not recorded.

Determining the number of ICU admissions per patient was required to outline the extent of correlation within the sample that arose via multiple admissions of single patients on separate occasions. Table 2 summarises the number of times patients were admitted to the ICU over the study period. This included patients admitted to the ICU during separate admissions to the Alfred Hospital within the 3 year period, as well as patients admitted multiple times to the ICU during a single hospital visit. 6,608 patients were admitted to the ICU only once. This constituted 89.8% of the patients under study. 596 (8.1%) were admitted twice. The remaining 154 patients (2.1%) were admitted 3 or more times.

Focussing on a patient's first ICU admission allowed for within-patient correlation to be eliminated and thus simplified further analysis. 7,358 observations were retained by isolating the first admission for each patient, while 972 observations corresponding to subsequent admissions were excluded.

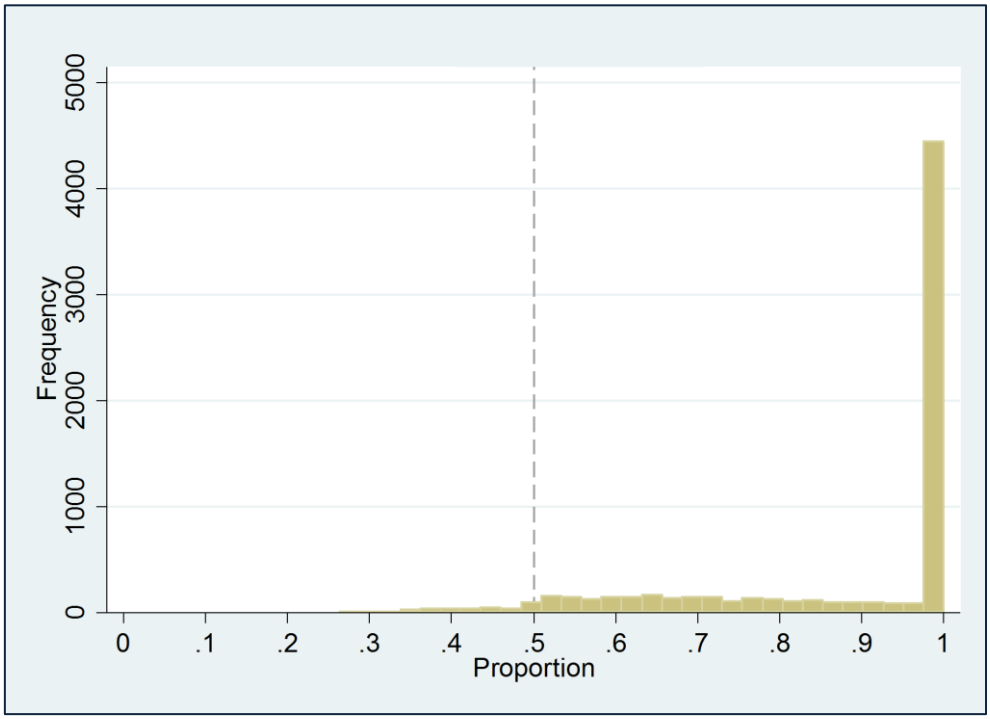
Table 3: Counts and proportions of the number of ICU beds patients spent time in during their first ICU stay

Beds on First ICU Admission	Patients	% of Raw Total	% of Retained Total
0	3	0.04	--
Missing	194	2.64	--
1	4,372	59.42	61.05
2	1,995	27.11	27.86
3	547	7.43	7.64
4	149	2.03	2.08
5	63	0.86	0.88
6	21	0.29	0.29
7	9	0.12	0.13
8	2	0.03	0.03
9	0	0.00	0.00
10	2	0.03	0.03
11	1	0.01	0.01
Raw Total	7,358	100	--
Retained Total	7,161	97.32	100

Of the 7,358 observations corresponding to the first ICU admission per patient, 3 indicated an admission to the ICU, but did not specify time in any bed (i.e. bed time variables 1-45 all equalled 0). These observations were excluded. 194 patients (2.64%) had missing values for time spent in all beds. These were also excluded. Hence data from 197 patients were omitted from further analysis, providing the final dataset. Table 3 summarises the extent of the patient movement within the ICU. Of patients who spent time in a bed; 6,367 (88.9%) spent time in either one or two beds, while a remaining 547 (7.6%) and 247 (3.5%) spent time in three beds, and four or more beds, respectively. 4,372 patients (61% of the final sample) spent their entire ICU stay in a single bed.

A patient's 'first-ranked' bed was defined as the bed in which they spent the majority of their time while in the ICU. Thus, the 4,372 patients (61%) who remained in a single bed for the duration of their ICU stay spent all of their time in their first-ranked bed. For these 61 per cent of patients, delirium could only have occurred during time spent in this bed. The time spent in the first-ranked bed as a proportion of total time spent in the ICU is shown in Figure 1 for all 7,161 patients. 6,874 patients (96% of all patients) spent over half of their ICU stay in one bed.

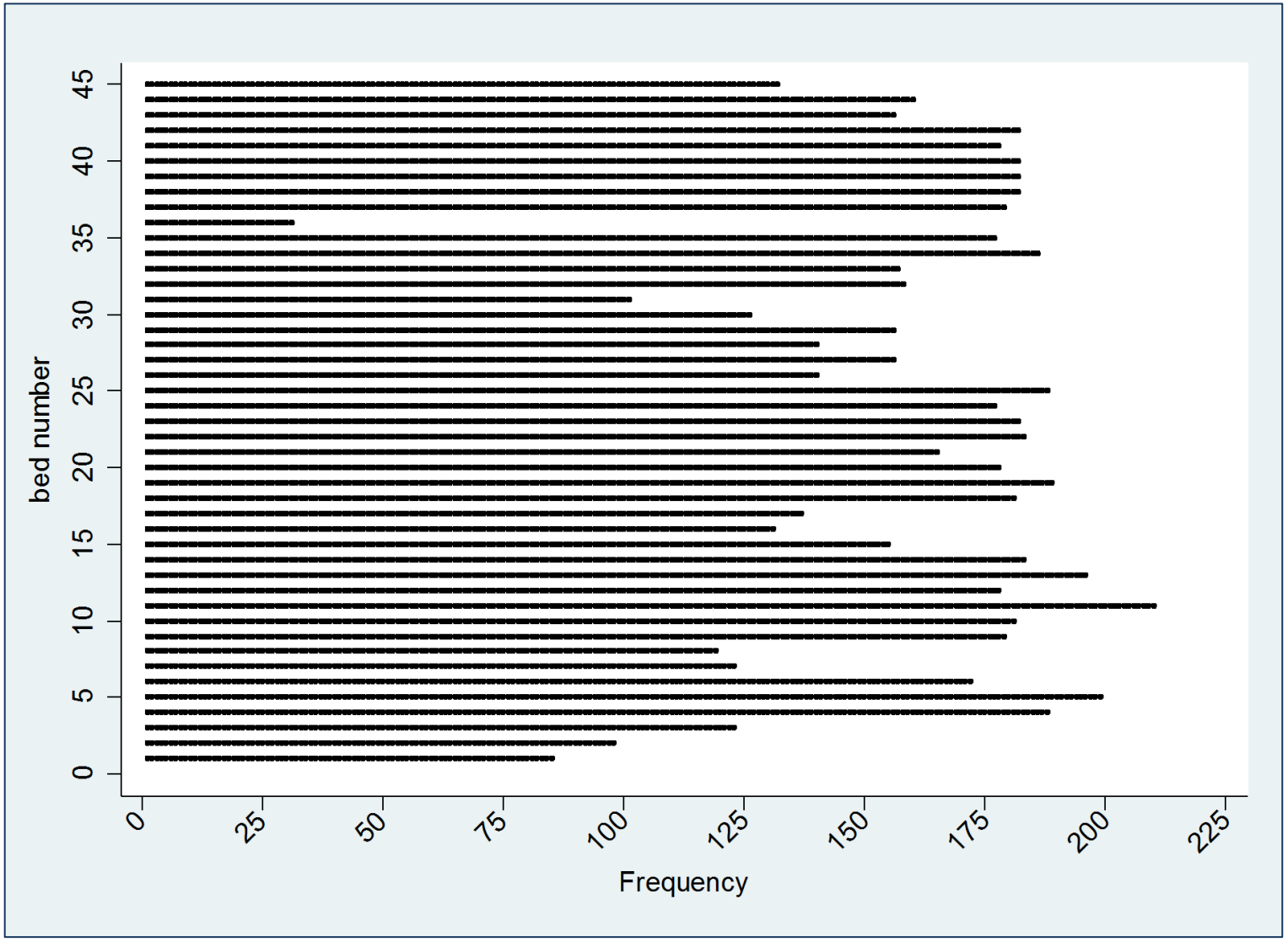
Figure 1: Proportions of total ICU time spent in the first-ranked bed for each patient (n = 7,161)



Given that the overwhelming majority of patients spend a majority of their ICU time in one bed, it is reasonable to assign patients to a 'dominant bed' for the purposes of delirium risk prediction, and generation of SMRs. Primary analysis is therefore restricted to the dominant bed for each patient. The time in this bed is assumed to be the patient's time at risk of delirium during their first ICU stay.

Figure 2 displays the number of patients for whom each bed was classified as a dominant bed. The number of patients assigned to each bed varies. Bed 36 had the lowest number of patients assigned to it (31 patients), while bed 11 had the highest (210 patients). The discrepancy between the numbers of patients assigned to the beds may be due to different uses for different beds within the ICU and preferences of ICU staff for the placement of patients in particular beds.

Figure 2: Dominant bed distribution; *counts of instances of patients being assigned to each of beds 1 to 45 (n = 7,161)*



3.2 Developing predictive models for delirium

1,144 (16%) of the total 7,161 patients on their first ICU admission developed delirium, nearing the lower end of the 15-80% range of the incidence of delirium reported in ICU environments [3]. 7,115 patients had complete data for all variables in Table 5.

Of this cohort, the estimation set on which the models were built consisted of complete data for 5,691 patients. The validation set consisted of complete data for the remaining 1,424 patients. The final selected models were re-fit to the data from the 7,115 patients with complete data.

All patient-level candidate variables were included in an initial patient-level model. These included age, the Verbal, Motor and Eye Opening components of total GCS score, APACHE III score with age and GCS points removed, sex, diagnosis category, ICU admission source, ICU admission type, smoking status, the (natural log of) time spent in the dominant bed, English speaking ability and the indication of time spent under invasive ventilation. There was no clinical justification for any patient-level interaction terms to be included.

All bed-level candidate variables were included in an initial patient-bed-level model. These included door type, the indication of a bed being near a busy area, and the mean of the natural light level measurements across the 45 beds. Bed-level 2-way and 3-way interaction terms were also included in this model. This model excluded English speaking and time under invasive ventilation, and included all other patient-level candidate variables.

Table 4 displays the performance statistics for these initial models, along with those for the final chosen models described below. Statistics corresponding to the final patient-level and patient-bed-level models are in bold. Overall, discrimination as assessed by the AU ROC was high, and the H-L p values indicate that the models were well calibrated. All models considered gave small estimates of the variance of the random effects of bed j.

Inclusion of English speaking and time under ventilation variables in the patient-level model did not contribute to meaningfully improved discrimination, and resulted in a model with a higher AIC and BIC (Model 1 vs Model 2). The model with these two variables excluded was chosen as the final patient-level model. Inclusion of the interactions in the patient-bed-level model did not result in markedly better discrimination. The model that omitted the interaction terms resulted in a lower AIC and BIC (Model 4 vs Model 3) and was thus chosen as the final patient-bed-level model. The final patient-level model included variables summarised in Table 5, and is described by equation (2).

Table 4: Performance statistics for selecting a risk-adjustment model; *AIC, BIC, Area under the ROC curve and p values for the Hosmer-Lemeshow $\hat{C}$ statistic (H-L p value)*

Numbers in brackets represent values based on the 20% test set.

Model 1: Model containing all patient-level candidate variables.
Model 2: Model 1 with the indicators of invasive ventilation and English speaking excluded. (2)
Model 3: Model 2 with all bed-level main effects (3)
Model 4: Model 3 with bed-level 2-way and 3-way interactions.

Patient-level and Patient-bed-level models selected for the production of SMRs are shown in bold.

Model	AIC	BIC	AU ROC	H-L *p* value
1	4162.64	4401.919	0.7888 (0.7577)	0.6785 (0.2604)
2	**4160.783**	**4386.769**	**0.7885 (0.7585)**	**0.7775 (0.2073)**
3	**4167.197**	**4419.769**	**0.7887 (0.7588)**	**0.7752 (0.2443)**
4	4175.369	4474.468	0.7893 (0.7582)	0.9387 (0.5358)

Appendix 2 Table 2 provides the model parameter estimates for the final patient-level model, and Appendix 2 Table 1 provides the final null model. The final patient-bed-level model included the variables summarised in both tables 5 and 6 and is described by equation (3). Appendix 2 Table 3 provides the model parameter estimates for the final patient-bed-level model.

Table 5: Summary of the patient-level and outcome variable(s) included in models (2) and (3); *summary statistics are reported on their measured scales (cont. over page)*

	n (%)		n (%)
Total	7,161	**Total**	7,161
GCS – Eye component		**Age (years)**	
1[a]	946 (13)	<45	1,870 (26.1)
2	245 (3.4)	45 to 59	1,736 (24.2)
3	1,000 (14)	60 to 69	1,528 (21.3)
4	4,942 (69)	70+	2,027 (28.3)
Missing[b]	28 (0.4)	Median [IQR]	59 [44,71]
GCS – Motor component		**Smoking Status**	
1[a]	621 (8.7)	Current[a]	1,594 (22.3)
2	62 (0.9)	Ex Smoker	2,043 (25.5)
3	91 (1.3)	Never Smoked	1,963 (27.4)
4	197 (2.8)	Unknown	1,552 (21.7)
5	486 (6.8)	Missing[b]	9 (0.13)
6	5,676 (79)		
Missing[b]	28 (0.4)		

GCS – Verbal component			**ICU source category**		
1[a]	946	(13)	Accident and emergency[a]	2,683	(37.5)
2	376	(5.3)	OT / recovery	3,238	(45.2)
3	188	(2.6)	Other hospital	189	(2.64)
4	837	(12)	Other hospital ICU	195	(2.72)
5	4,786	(67)	Unknown	9	(0.13)
Missing[b]	28	(0.4)	Ward	847	(11.8)
Sex			**Admission type category**		
Female[a]	2,379	(33.2)	Elective[a]	2,097	(29.3)
Male	4,762	(66.5)	Emergency	5,055	(70.6)
Missing[b]	20	(0.28)	Unknown	9	(0.1)
Diagnosis Category			**APACHE III score (without the age and GCS component)**		
Trauma (other than head)[a]	963	(13.5)			
Cthr	1,289	(18)			
Head injury	673	(9.4)	<20	923	(12.9)
Med	1,583	(22)	20 to 39	3,516	(49.1)
Surgical	468	(6.5)	40 to 59	1,924	(26.9)
Other cthrop	367	(5.1)	60 to 79	583	(8.1)
Neuro	642	(9.0)	80+	215	(3.0)
Cvs	680	(9.5)	Median [IQR]	35	[25,46]
Od/metabolic	308	(4.3)			
Burn	188	(2.6)			
Time spent in the dominant bed (hours)			**Delirium diagnosed (outcome)**		
<1	993	(13.9)	Yes	1,144	(16)
1 to 3	3,049	(42.6)	No	6,017	(84)
3 to 5	1,385	(19.3)			
5+	1,734	(24.2)			
Median [IQR]	2.67	[1.5,4.9]			

[a] baseline value of a categorical variable used in both models (2) and (3).

[b] values excluded when generating the SMR bootstrap standard errors. ∴ models (2) and (3) used $n = 7{,}115$ for SMR estimates and CIs.

Table 6: Summary of bed-level variables included in model (3); *summary statistics report variables on their measured scales, patients with missing patient-level data excluded.*

	n (%)		n (%)
Total	7,115	**Total**	7,115
Busy Area		**Door type**	
No[a]	5,787 (80.8)	Closing[a]	2,845 (39.7)
Yes	1,374 (19.2)	Isolated	501 (7.0)
		Open	3,815 (53.3)
		Mean light measurement: lights off (lux)	
		2 to <4	747 (10.5)
		4 to <6	3,044 (42.8)
		6 to <8	2,157 (30.3)
		8 to <10	1,028 (14.4)
		10+	139 (2.0)
		Median [IQR]	5.94 [5.0,7.52]

[a] baseline category of a categorical variable used in model (3).

Variables treated as continuous were centred before being included in the models. Age was centred at 50 years, and divided by 5; APACHE III score without the age and GCS component was centred at 37; Mean light measurement with lights off was centred at 6.18 lux. Categorical variables were included as summarised in Tables 5 and 6. The natural log was taken of the measure of the time spent in the dominant bed.

With time in beds other than the dominant bed ignored, time spent in the dominant bed was used as the indication of time spent at risk of developing delirium during the first ICU stay. While this may not be appropriate for a minority of patients who were moved often between beds, it was not likely to affect the predictive ability of the models, or the resulting SMRs in any clinically meaningful way. Sensitivity analysis for this is beyond the scope of this report, however it is important to note this as a limitation of the current analysis (see Discussion). Increasing (logged) time in the dominant bed was highly associated with increased delirium risk in both patient-level and patient-bed level models, $(p < 0.001)$ after adjusting for patient-level and patient-bed-level risk factors in each model respectively.

Binned residual plots used to assess model fits are shown in Figure 3: these plots are similar to the residual plots for linear regression models. Approximately 95% of the points are expected to lie within the confidence limits if the model is correct, with a random scatter of points. In both models shown in Figure 3, 95% of the points representing the binned residuals are randomly scattered and lie within the confidence limits, indicating good model fit to the data. The distributions of outlying points (indicated by the triangles beyond the limits) are not suggestive of any systematic problems

with the model fit. For each bin, approximate 95% confidence limits are $\pm 2\sqrt{p(1-p)/n}$, estimated using the standard deviation of each bin's residuals, with bins of size $n = 75$ approximately equal to the square root of the sample size (5,728). Observations outside the limits at low values of the average predicted delirium probability are not necessarily suggestive of poor model fit, and are instead caused by predicted probabilities approaching zero shrinking confidence limits at these lower values.

There is no indication that more complex forms of the continuous variables included in each model are required, as shown in the binned residual plots for these variables in figures 4 and 5.

Figure 3: Binned residual plots to assess the fit of the final models; *(A) patient-level model, (B) patient-bed-level model*

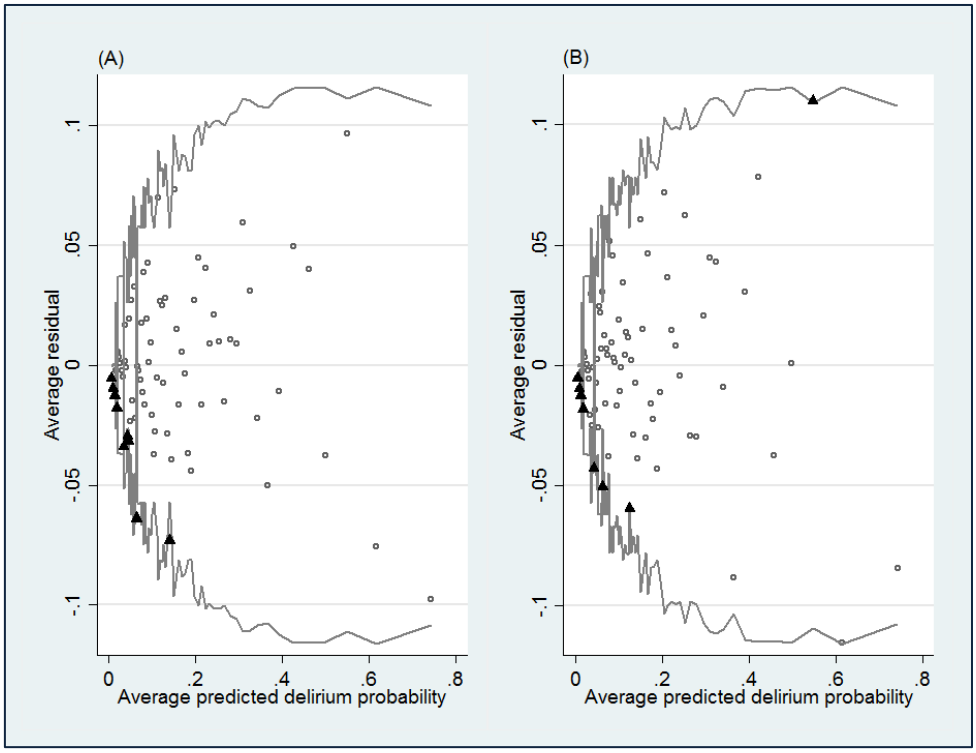


Figure 4 : Binned Residual Plots to assess the fit of continuous variables in the patient level model: *(A) Centred, scaled age for APACHE III score, (B) Time in the dominant bed on the log scale, (C) centred APACHE III score (without the age and GCS components)*

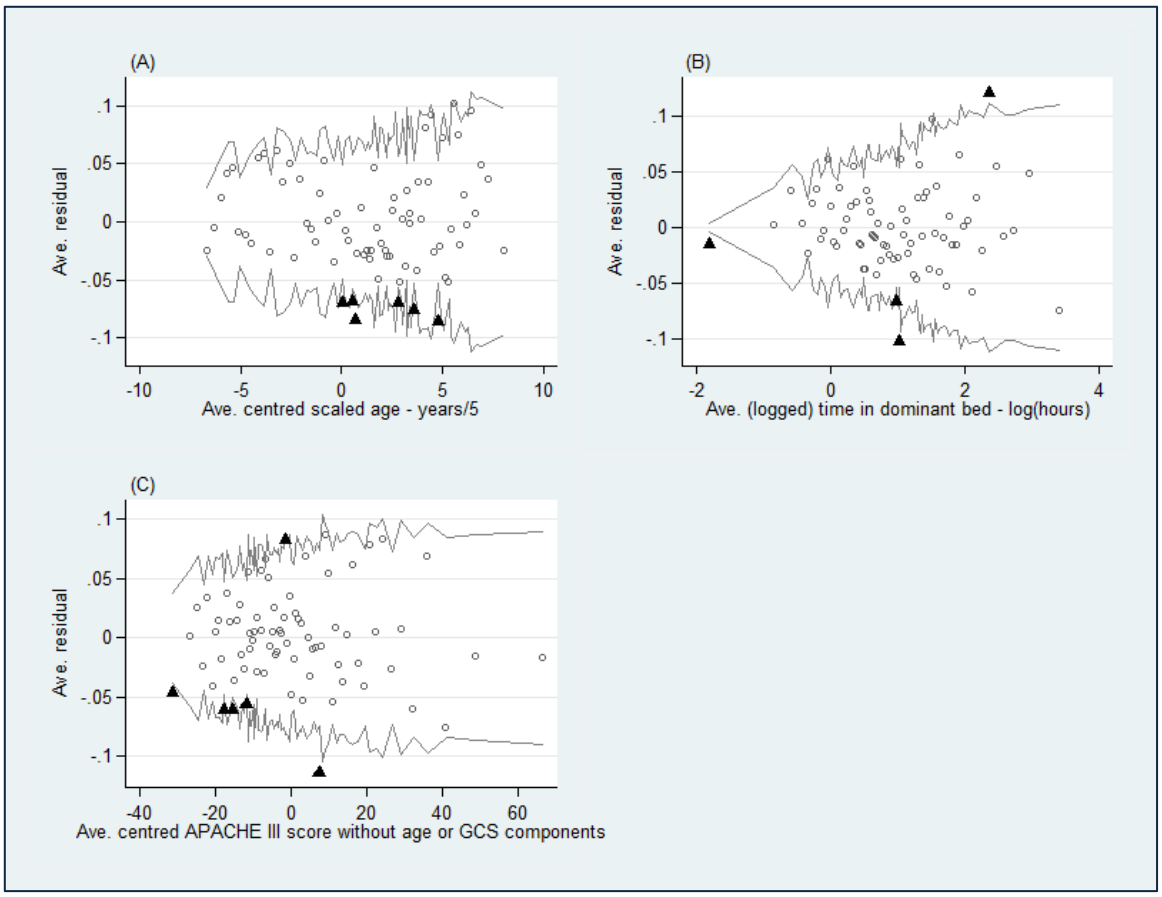
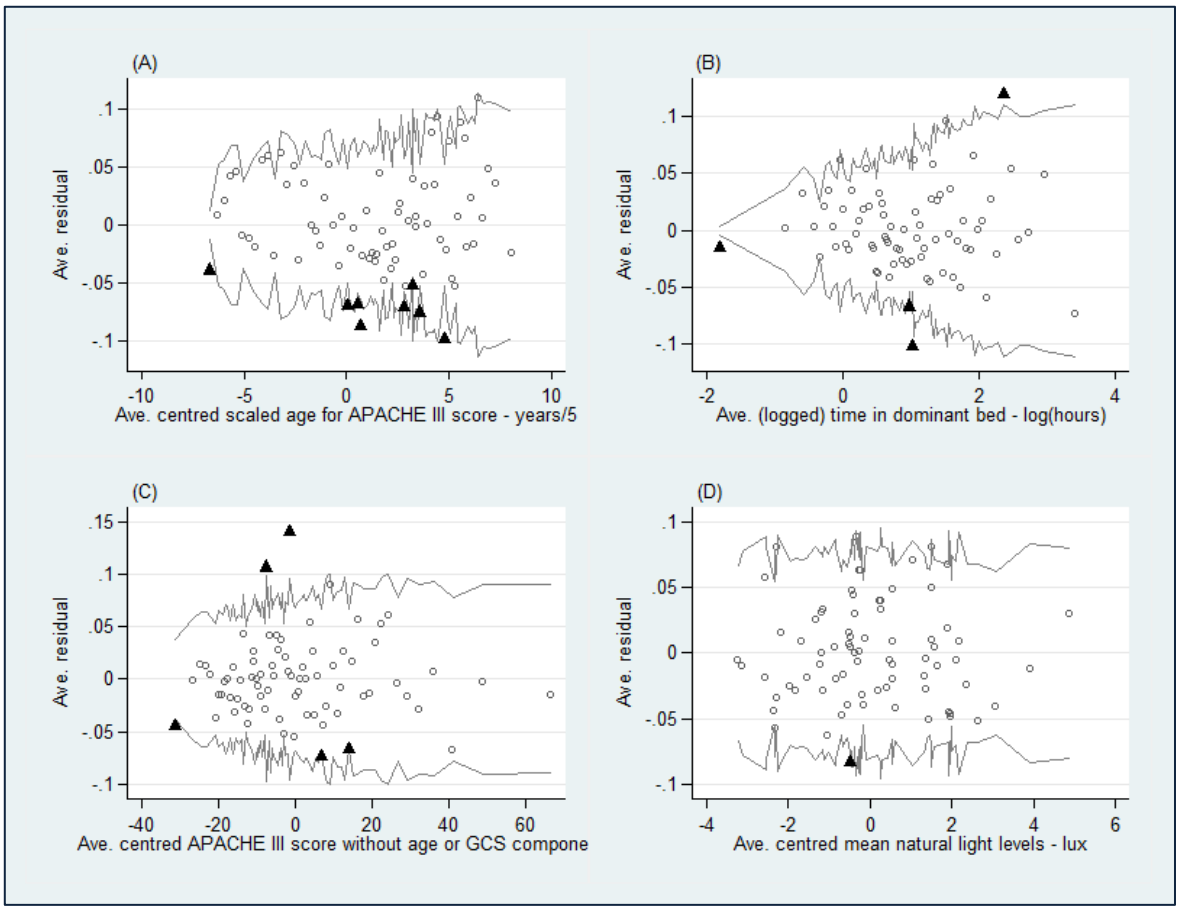


Figure 5: Binned Residual Plots to assess the fit of continuous variables in the patient-bed-level model: *(A) Centred, scaled age for APACHE III score, (B) Time in the dominant bed on the log scale, (C) Centred APACHE III score (without the age and GCS components), (D) Centred mean lights-off measurements*



3.3 Generating SMRs

Figure 6 displays the four SMR types generated using the patient-level model. Those generated from the null and patient-bed-level model are described in Appendix 3: Figures 1 and 2. For the patient-level model, the SMR3s are almost identical to the SMR1s, as the estimated odds of delirium in bed j given by SMR1 are similar to the risk ratios, p_j/p_0 given by SMR3. The SMR2s did not account for sampling variation in the observed number of delirium cases (due to its numerator being the sum of the observed cases from this sample), hence the variance in the estimates is wider for these, and they are less correlated with the other three SMR types, as shown in Figure 6. The SMR4s are highly correlated with the SMR3s. SMR4s are considered further, as they account for sampling variation and are intuitively interpreted as the ratio of observed cases associated with a bed, to that of the expected cases in an average bed.

Figure 6: Scatter matrix; *From top to bottom, patient-level SMRs: plSMR1, plSMR2, plSMR3, plSMR4calculated by (4), (5), (6) and (7), respectively.*

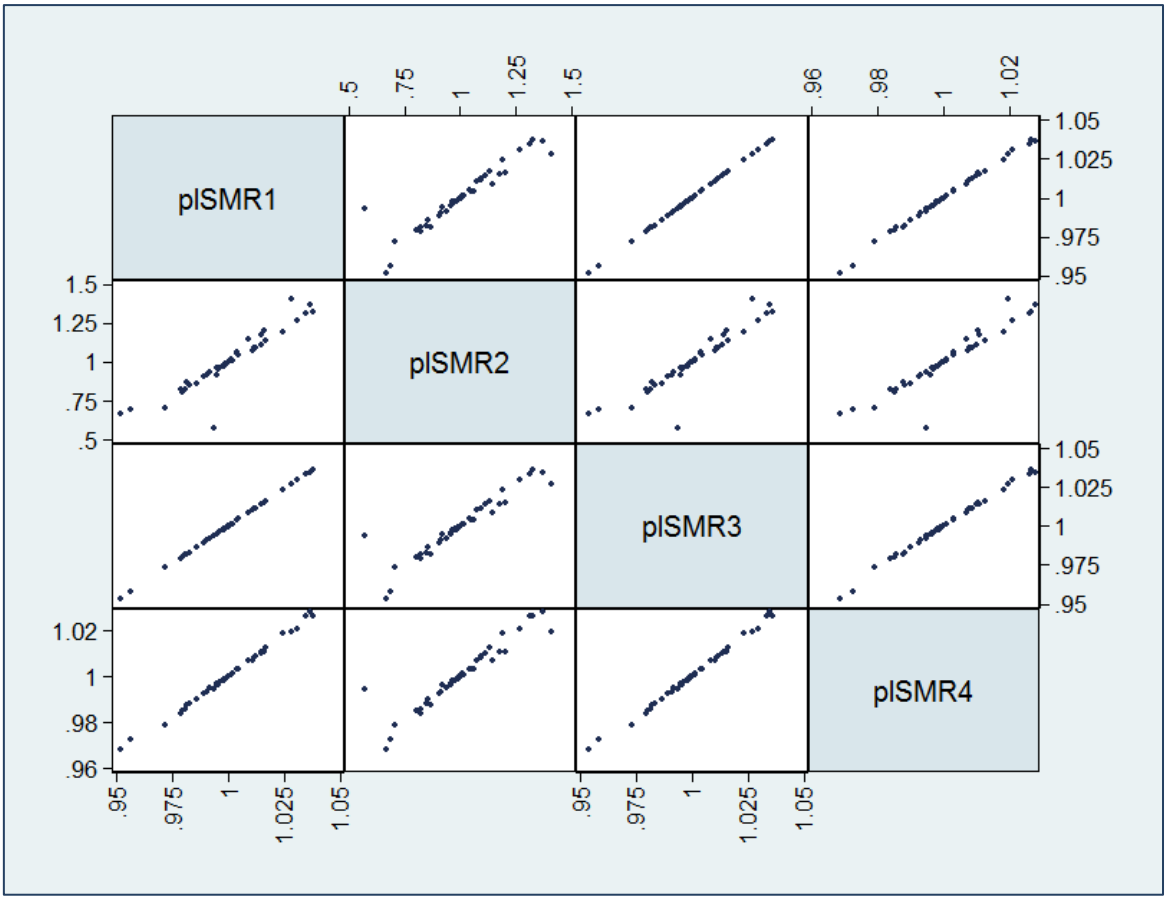


Figure 7 shows strong correlation between the SMR4s generated from each of the null, patient-level and patient-bed-level models; Figure 8 shows strong correlation between the SMR3s. SMRs generated by the patient-bed-level model are less variable than those produced by the null and patient-level models, indicating that inclusion of the bed level variables accounts for some of the variation in SMRs otherwise accounted for by the random effects in the null and patient level models.

Figures 9, 10 and 11 show forest plots of the SMR4s generated by each of the null, patient-level and patient-bed-level models respectively. Contrasting the SMR4 estimates and confidence intervals between the three models highlights the movement towards an estimated value of 1 as patient-level and bed-level variables are included in the models.

Figure 7: Scatter matrix; *SMR4s generated by the three models: SMR4pl derived from the patient-level model, SMR4n from the null model, SMR4pbl from the patient-bed-level model*

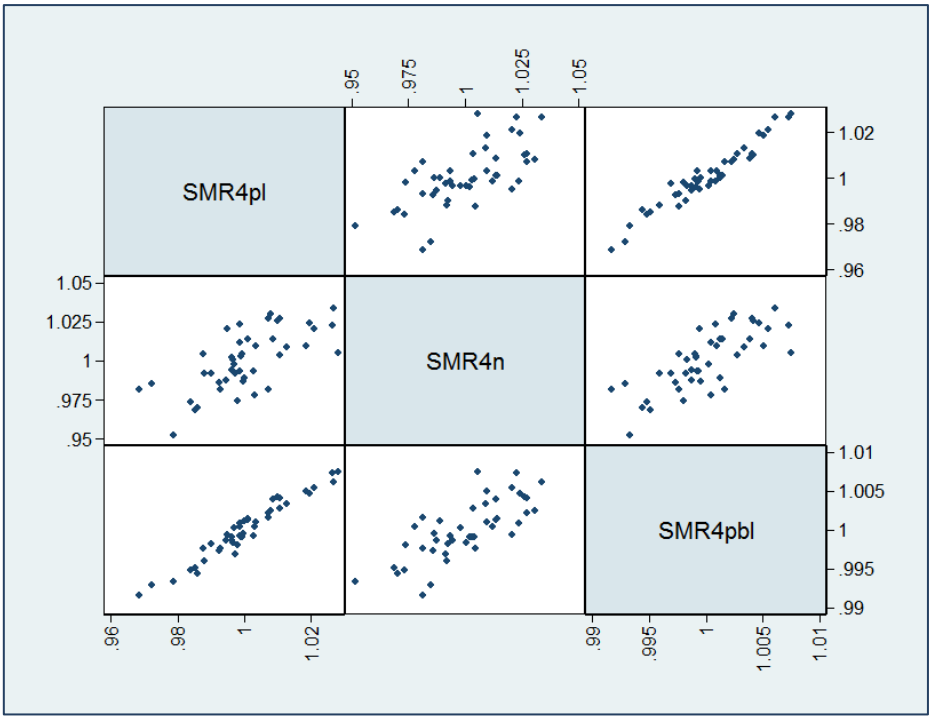


Figure 8: Scatter matrix; *SMR3s generated by the three models: SMR3pl derived from the patient-level model, SMR3n from the null model, SMR3pbl from the patient-bed-level model*

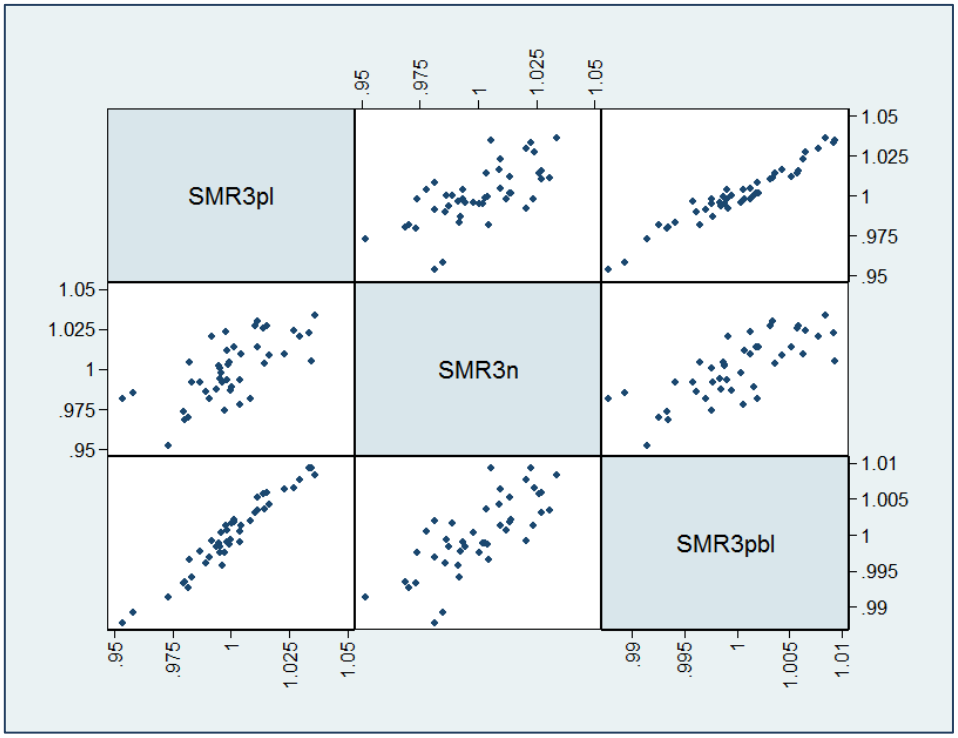


Figure 9: SMR4s generated by the null model: *95% confidence limits derived using standard errors obtained from 100 bootstrap samples.*

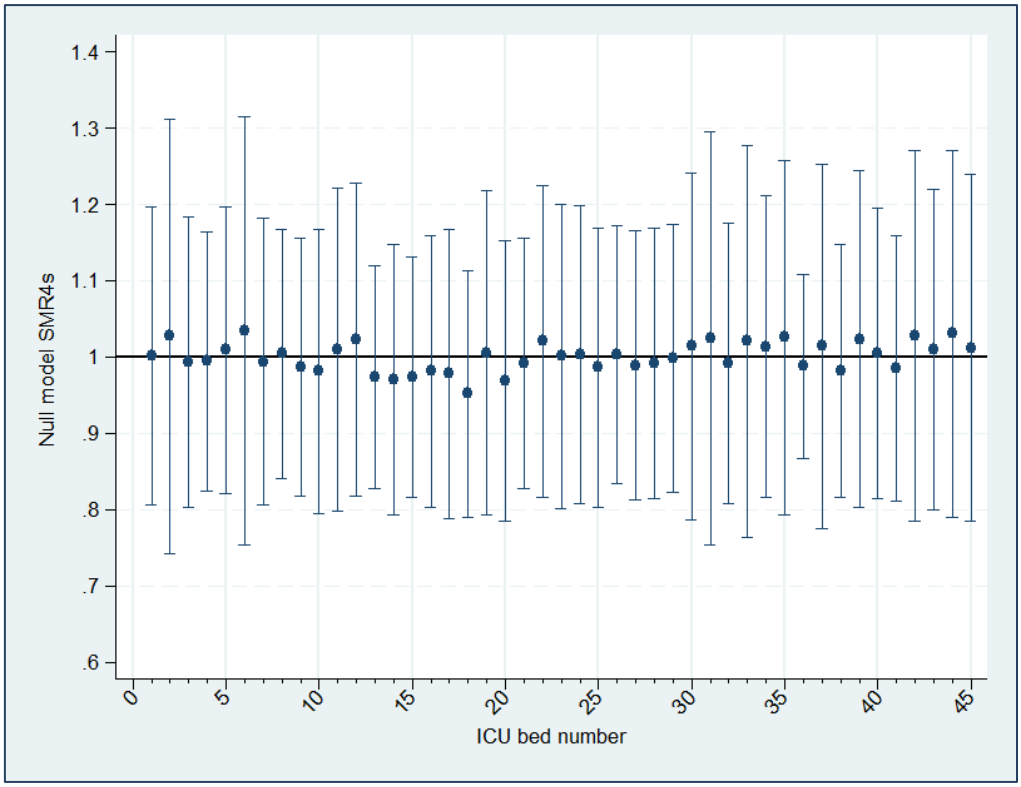


Figure 10: SMR4s generated by the patient-level model; *95% confidence limits derived using standard errors obtained from 100 bootstrap samples.*

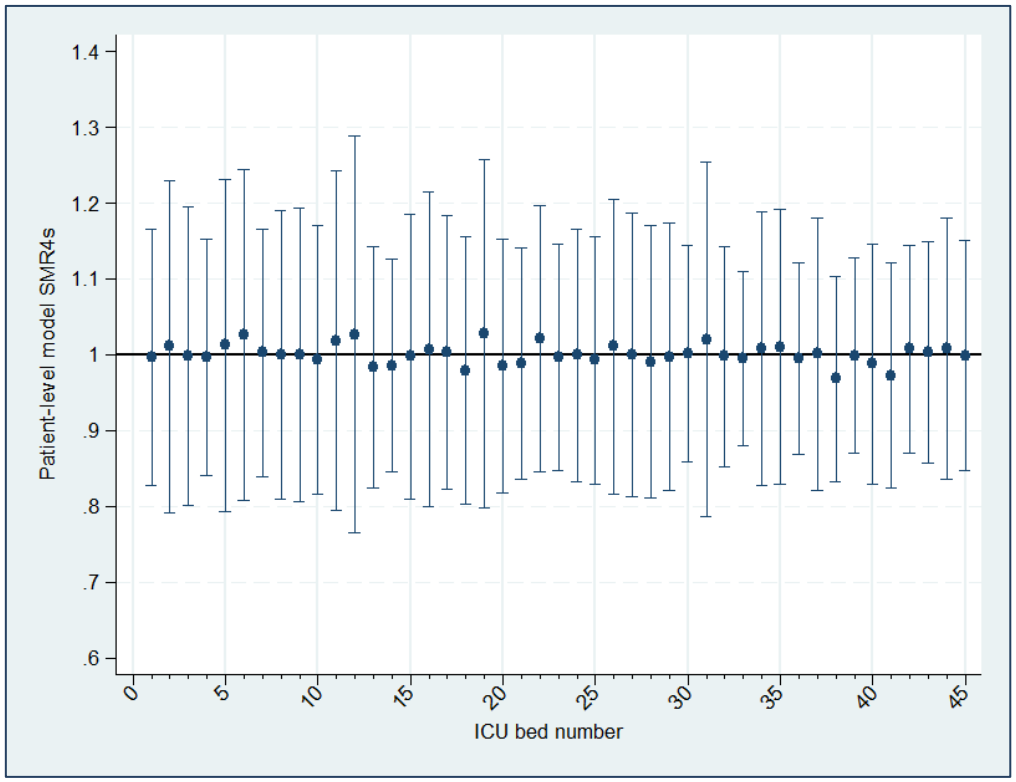
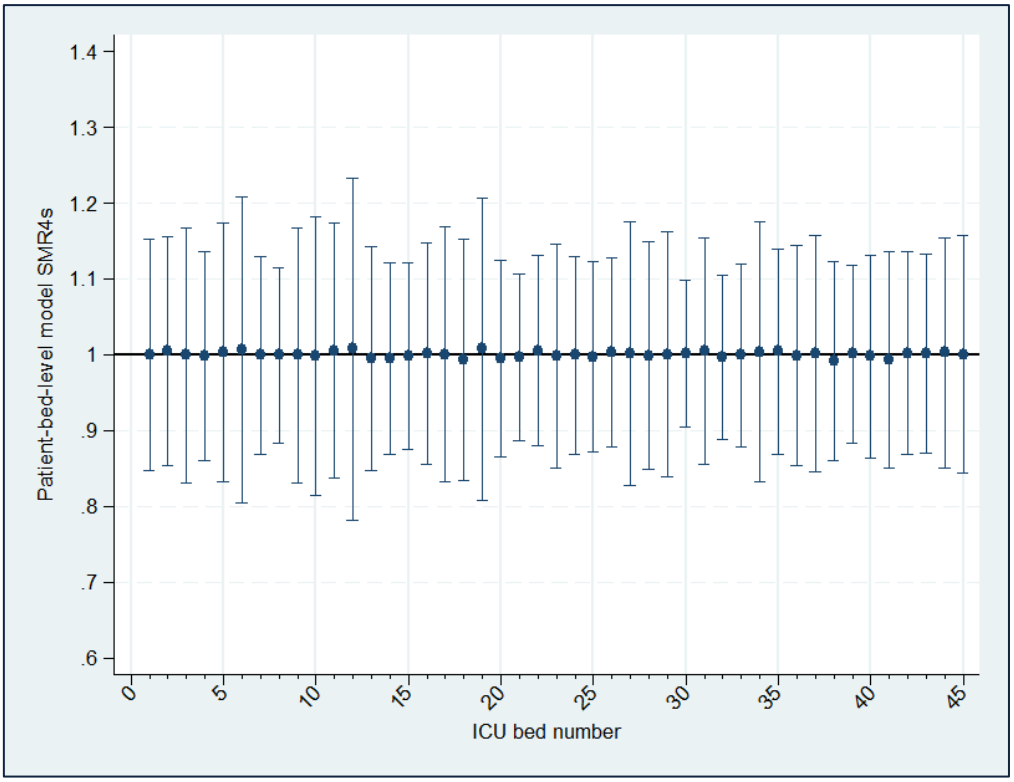


Figure 11: SMR4s generated by the patient-bed-level model; *95% confidence limits derived using standard errors obtained from 50 bootstrap samples.*



In each case, using each model in turn, the confidence intervals included 1 for all SMR4s. This finding held regardless of the SMR type (3 or 4), or model used. Hence, this dataset does not provide evidence that any of the 45 beds had unusually high or low levels of delirium.

4. Discussion

This project aimed to compare the risk-adjusted levels of delirium in each of the 45 beds in the Alfred Hospital's ICU, to identify any beds with unusual levels of delirium. In addition to adjusting delirium risk for patient-level characteristics, bed level measures of noise and light were included in a model to predict the risk of delirium. It is possible that differences in delirium risk may arise through differences in noise and light levels between beds. ICU staff had previously suspected that patients assigned to certain beds seemed to develop delirium more often than those assigned to other beds. Specifically, ICU staff noted that patients assigned to one particular bed seemed to consistently develop delirium, anecdotal observations that provided the incentive for this study. Results of this analysis however, show that there were no beds with unusual delirium risk as measured by the SMRs. Overall the incidence of delirium was low in this ICU. Accounting for noise and light characteristics reduced the standard errors of the SMR estimates, supporting results of previous studies that have explored relationships between delirium outcome and noise and light exposure [7].

There were limitations to this study. Delirium was diagnosed during the course of an ICU stay and patients were assigned to one bed only. In most instances this was reasonable, since most patients remained in one bed for their entire ICU stay; however it was not known exactly when delirium was diagnosed during an ICU admission. Therefore, in the worst case, some patients who were assigned to their dominant bed may have developed delirium while in a different bed. The impact of this potential discrepancy is likely to be small for the vast majority of the patient sample, given 61 percent of patients remained in one bed for their entire ICU stay. Very few patients were moved frequently and the likelihood of contributing a count of diagnosed delirium to a bed other than the dominant bed is likely to be small. Although more complex modelling is a viable option, it is not clear how useful a multi-membership model would have been in comparison to the method used.

Although the predictive models had good discrimination, there may have been other additional variables that could have been included to better predict delirium, which may have resulted in models with greater discriminative ability than those applied. For example, the length of a patient's stay in hospital before being admitted to the ICU was not accounted for and was unknown in many cases. Patients with longer stays before or between ICU admissions may have been at a greater risk of developing delirium, and this risk was not able to be accounted for. Additionally, there were no variables that directly described noise levels at each bed in decibels. The location variable, although indicating exposure to noise for a bed in proximity to busy areas, was not a direct measure of decibel levels. Such direct measures of sound level may have provided better information for the patient-bed-level model. Models with a greater number of interactions could also have been considered, as there were sufficient delirium events to support such more complex models. However, since the focus here was on the estimation of SMRs and their standard errors, these more complex models were not explored. A further limitation of the models regards the bootstrap method used to provide the SMR standard errors. The complexity of the patient-level and patient-bed-level models, with their inclusion of random effects, led to long times taken to fit to the re-samples and produce the bootstrapped standard errors. This limited the number of bootstrap samples possible, especially for the patient-bed-level model SMRs.

This analysis has not taken into account a complex relationship between delirium and mortality, or a potential relationship between differing risks of death and assignment to particular beds. Patients with similar chances of dying are likely to be grouped into a bed by the condition for which they require ICU treatment, and those with higher chances of death are more likely to die before a diagnosis of delirium can be made, particularly if death is preceded by a period of unconsciousness. This relationship could not be accounted for by this study.

Finally, a potential limitation was that once ICU staff became aware of the study, they may have avoided placing patients in the beds that were thought to be associated with higher incidences of delirium. However, given that the majority of the data was collected prior to the study commencing, this is not likely to have had a large impact.

The SMRs generated relied on the predictive ability of the models, and the discrepancy between the SMR2s and the other SMR types may be partly due to the models used, in addition to the sampling error that this difference was attributed to. Using the number of observed delirium cases in the patient sample divided by the expected values derived by the models (as defined by SMR2) gave a significantly wider range of SMR estimates than for the other SMR types.

There were strengths to the study. The data collected was of a high quality, provided a large sample, and was collected for the purposes of predicting delirium. Light measurements were taken expressly for the purposes of this analysis. Despite the limitations discussed, results are robust to the prediction models that were used, each of which had good discrimination and fit. SMRs 1, 3 and 4 had good agreement and the study conclusions were robust to using any of these SMR types produced by any of the three models.

In conclusion, the reported patient-level model SMRs suggest that there are no beds with unusually high or low risks of delirium after controlling for patient-level delirium risk factors. This result holds when noise and light factors are included in the delirium prediction model that account for some variation in the patient-level SMRs. These findings may lend some support to the use of methods mitigating noise and light levels in ICUs to reduce delirium risk, given the known association between higher levels of noise and light on increased risks of developing delirium.

Appendix 1. SMR formulations

In the following SMR formulations, U_j is model-specific, describing the random effects after risk adjustment for the relevant covariates.

Null:

$$nSMR1_j = \exp(U_j) \quad (1)$$

$$nSMR2_j = \frac{\sum_{i=1}^{n_j} y_{ij}}{\sum_{i=1}^{n_j}(1+\exp[-\alpha])^{-1}} \quad (2)$$

$$nSMR3_j = \frac{\left(1+\exp[-\alpha+U_j]^{-1}\right)}{1+\exp[-\alpha]^{-1}} = \frac{[1+\exp(\alpha)]\exp(U_j)}{1+\exp(\alpha+U_j)} \quad (3)$$

$$nSMR4_j = \frac{\sum_{i=1}^{n_j}(1+\exp[-\alpha-U_j])^{-1}}{\sum_{i=1}^{n_j}(1+\exp[-\alpha])^{-1}} \quad (4)$$

Patient-Level:

$$plSMR1_j = \exp(U_j) \quad (5)$$

$$plSMR2_j = \frac{\sum_{i=1}^{n_j} y_{ij}}{\sum_{i=1}^{n_j}(1+\exp[-\alpha-\boldsymbol{\gamma'X_{ij}}])^{-1}} \quad (6)$$

$$plSMR3_j = \frac{\left(1+\exp[-\alpha+U_j]^{-1}\right)}{1+\exp[-\alpha]^{-1}} = \frac{[1+\exp(\alpha)]\exp(U_j)}{1+\exp(\alpha+U_j)} \quad (7)$$

$$plSMR4_j = \frac{\sum_{i=1}^{n_j}(1+\exp[-\alpha-\boldsymbol{\gamma'X_{ij}}-U_j])^{-1}}{\sum_{i=1}^{n_j}(1+\exp[-\alpha-\boldsymbol{\gamma'X_{ij}}])^{-1}} \quad (8)$$

Patient-bed-level:

$$pblSMR1_j = \exp(U_j) \quad (9)$$

$$pblSMR2_j = \frac{\sum_{i=1}^{n_j} y_{ij}}{\sum_{i=1}^{n_j}(1+\exp[-\alpha-\boldsymbol{\gamma'X_{ij}}-\boldsymbol{\beta'X_j}])^{-1}} \quad (10)$$

$$pblSMR3_j = \frac{\left(1+\exp[-\alpha+U_j]^{-1}\right)}{1+\exp[-\alpha]^{-1}} = \frac{[1+\exp(\alpha)]\exp(U_j)}{1+\exp(\alpha+U_j)} \quad (11)$$

$$pblSMR4_j = \frac{\sum_{i=1}^{n_j}(1+\exp[-\alpha-\boldsymbol{\gamma'X_{ij}}-\boldsymbol{\beta'X_j}-U_j])^{-1}}{\sum_{i=1}^{n_j}(1+\exp[-\alpha-\boldsymbol{\gamma'X_{ij}}-\boldsymbol{\beta'X_j}])^{-1}} \quad (12)$$

Appendix 2. Results of the null, patient-level and patient-bed-level models

Table 1: Null model results; fitted for the production of SMRs

Null model (*n=7,115)*					
		Fixed Effects			
Var.	**Coef. (95% CI)**	***p* value**			
Constant	-1.66 (-1.73 -1.60)	<0.001			
		Random Effects	**Random effects std. dev. (std. error)**	0.072	(0.0764)

Table 2: Patient-level model results; fitted for the production of SMRs

Patient-level model (*n=7,115*)					
Fixed Effects					
Var.	**Coef. (95% CI)**	***p* value**	**Var.**	**Coef. (95% CI)**	***p* value**
GCS – Eye comp.					
1	ref.		**Age (years/5)**	0.154 (0.13 0.18)	<0.001
2	0.061 (-0.43 0.55)	0.808			
3	0.379 (-0.07 0.83)	0.101			
4	0.241 (-0.23 0.71)	0.312			
GCS – Motor comp.			**Smoking Status**		
1	ref.		Current	ref.	
2	0.188 (-0.63 1.00)	0.651	Ex Smoker	-0.202 (-0.41 0.00)	0.056
3	-0.233 (-0.96 0.49)	0.527	Never Smoked	-0.316 (-0.53 -0.11)	0.003
4	0.079 (-0.47 0.63)	0.781	Unknown	-0.317 (-0.54 -0.10)	0.005
5	0.153 (-0.34 0.64)	0.540			
6	0.552 (-0.01 1.12)	0.057			
GCS – Verbal comp.			**ICU source category**		
1	ref.		Accident and Emergency	ref.	
2	0.320 (-0.08 0.72)	0.121	OT / recovery	-0.125 (-0.37 0.12)	0.312
3	0.323 (-0.19 0.84)	0.220	Other hospital	-0.062 (-0.52 0.39)	0.789
4	0.087 (-0.39 0.57)	0.723	Other hospital ICU	0.347 (-0.05 0.74)	0.087
5	-1.029 (-1.53 -0.53)	<0.001	Ward	0.313 (0.07 0.55)	0.010
Sex			**Admission type cat**		
Female	ref.		Elective	ref.	
Male	0.361 (0.20 0.52)	<0.001	Emergency	-0.354 (-0.64 -0.07)	0.013
Diagnosis Category			**APACHE III score (without the age and GCS component)**	0.007 (0.002 0.011)	0.004
Trauma (other than head)	ref.				
Cthr	-0.051 (-0.43 0.33)	0.793			
Head injury	-0.340 (-0.68 0.00)	0.050			
Med	0.116 (-0.17 0.40)	0.428			
Surgical	0.570 (0.18 0.95)	0.004			
Other cthrop	0.399 (-0.01 0.81)	0.056			
Neuro	0.033 (-0.31 0.37)	0.848			
Cvs	-0.021 (-0.39 0.31)	0.905			
Od/metabolic	0.937 (0.52 1.36)	<0.001			
Burn	0.737 (0.24 1.23)	0.004			
Time spent in dominant bed (log(hours))	0.909 (0.82 1.00)	<0.001	**Constant**	-3.08 (-3.61 -2.55)	<0.001
Random Effects			**Random effects std. dev. (std. error)**	0.0678 (0.0918)	

Table 3: Patient-bed-level model results; fitted for the production of SMRs

Patient-bed-level model (*n=7,115*)					
Fixed Effects					
Var.	**Coef. (95% CI)**	***p* value**	**Var.**	**Coef. (95% CI)**	***p* value**
GCS – Eye comp.					
1	ref.		**Age (years/5)**	0.154 (0.13 0.18)	<0.001
2	0.056 (-0.44 0.54)	0.823			
3	0.374 (-0.08 0.83)	0.105			
4	0.235 (-0.23 0.70)	0.325			
GCS – Motor comp.			**Smoking Status**		
1	ref.		Current	ref.	
2	0.208 (-0.61 1.02)	0.617	Ex Smoker	-0.205 (-0.41 0.00)	0.052
3	-0.214 (-0.94 0.51)	0.562	Never Smoked	-0.317 (-0.53 -0.11)	0.003
4	0.086 (-0.47 0.64)	0.762	Unknown	-0.320 (-0.54 -0.10)	0.004
5	0.168 (-0.32 0.66)	0.501			
6	0.563 (-0.01 1.13)	0.053			
GCS – Verbal comp.			**ICU source category**		
1	ref.		Accident and Emergency	ref.	
2	0.313 (-0.09 0.72)	0.130	OT / recovery	-0.137 (-0.38 0.11)	0.269
3	0.319 (-0.20 0.84)	0.225	Other hospital	-0.069 (-0.52 0.38)	0.766
4	0.082 (-0.40 0.56)	0.738	Other hospital ICU	0.354 (-0.04 0.75)	0.081
5	-1.038 (-1.53 -0.54)	<0.001	Ward	0.314 (0.076 0.55)	0.010
Sex			**Admission type cat**		
Female	ref.		Elective	ref.	
Male	0.364 (0.21 0.52)	<0.001	Emergency	-0.347 (-0.63 -0.065)	0.016
Diagnosis Category			**Door type**		
Trauma (other than head)	ref.		Closing	ref.	
Cthr	-0.072 (-0.46 0.32)	0.718	Isolated	0.155 (-0.14 0.45)	0.298
Head injury	-0.337 (-0.68 0.00)	0.053	Open	0.125 (-0.07 0.31)	0.197
Med	0.123 (-0.18 0.42)	0.422			
Surgical	0.598 (0.20 1.00)	0.003			
Other cthrop	0.414 (-0.00 0.83)	0.053			
Neuro	0.049 (-0.30 0.39)	0.783			
Cvs	-0.041 (-0.39 0.31)	0.814			
Od/metabolic	0.952 (0.52 1.38)	<0.001			
Burn	0.771 (0.26 1.29)	0.003			
Time spent in dominant bed (log(hours))	0.914 (0.82 1.00)	<0.001	**APACHE III score (without the age and GCS component)**	0.007 (0.002 0.011)	0.004
Busy Area			**Mean light measurement: lights off (lux)**	-0.003 (-0.06 0.05)	0.906
No	ref.				
Yes	0.069 (-0.12 0.26)	0.469			
Constant	-3.17 (-3.72 -2.63)	<0.001			
Random Effects			**Random effects std. dev. (std. error)**	0.036 (0.1637)	

Appendix 3. Null and patient-bed-level SMRS

Figure 1: Scatter Matrix; *Null SMRs; From top to bottom, null SMRs calculated by Appendix 1- Equations (1), (2), (3) and (4).*

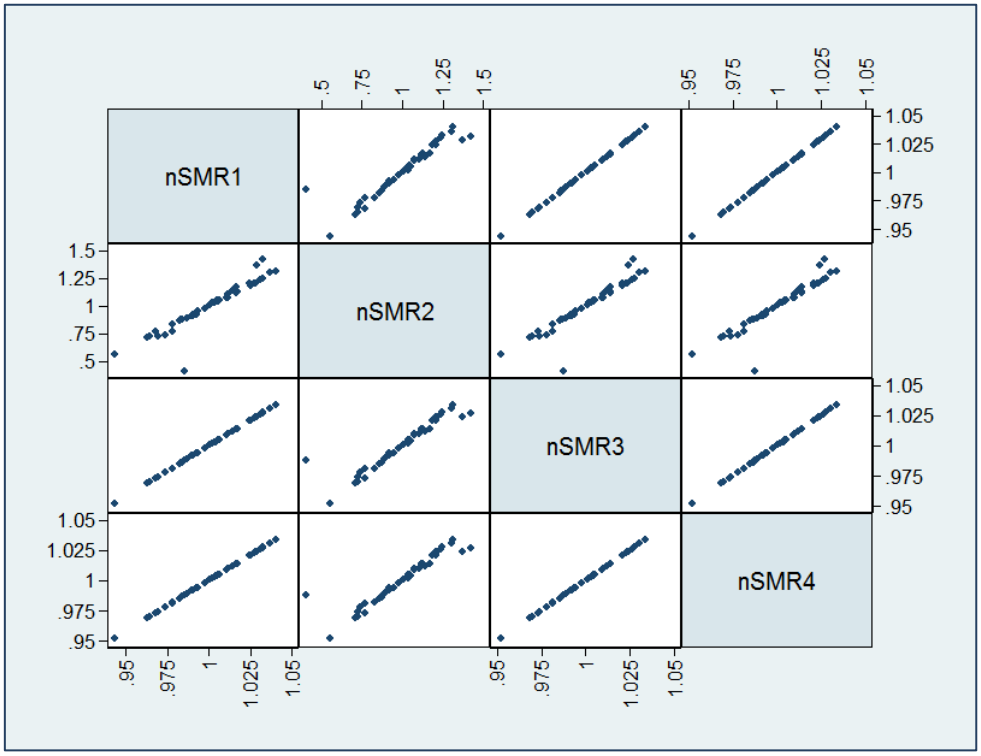
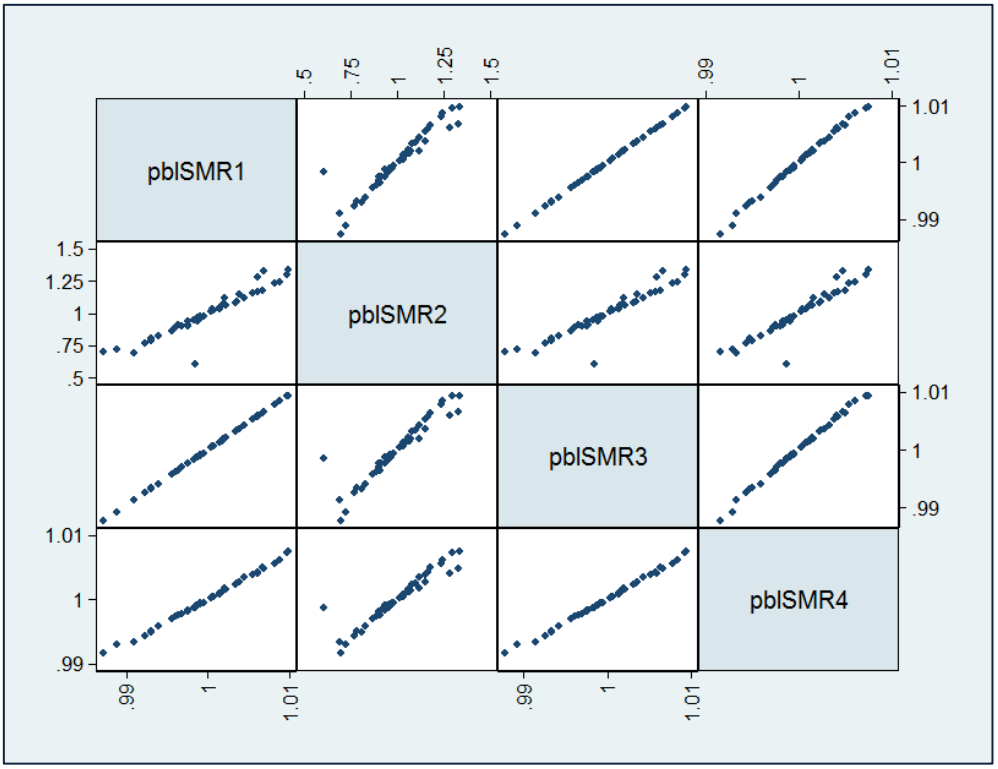


Figure 2: Scatter Matrix; *Patient-bed-level SMRs; From top to bottom, patient-bed-level SMRs calculated by Appendix 1- Equations (9), (10), (11) and (12)*



References.

[1] Maldonado JR. (2008) Delirium in the Acute Care Setting: Characteristics, Diagnosis and Treatment. *Critical Care Clinics,* 24 (4), 657-722.

[2] Ely E, Guatam S, Margolin R, Francis J, May L, Speroff T, et al. (2001) The impact of delirium in the intensive care unit on hospital length of stay. *Intensive Care Medicine,* 27 (12), 1892-1900

[3] Wenham T, Pittard A. (2009) Intensive Care Unit Environment. *Continuing Education in Anaesthesia, Critical Care & Pain*, 9 (6), 178-183

[4] Ely W, Inouye SK, Bernard GR, Gordon S, Francis J, May L, et al. (2001) Delirium in mechanically ventilated patients. Validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *The Journal of the American Medical Association*, 286 (21), 2703–10

[5] Van Rompaey B, Schuurmans MJ, Shortridge-Baggett LM, Truijen S, Elseviers M, Bossaert L. (2008) A comparison of the CAM-ICU and the NEECHAM Confusion Scale in intensive care delirium assessment: an observational study in non-intubated patients. *Critical Care*, 12 (1), 131

[6] Meyer TJ, Eveloff SE, Bauer MS, Schwartz WA, Hill NS, Millman RP. (1994) Adverse environmental conditions in the respiratory and medical ICU settings. *Chest,* 105 (4), *1211-6*

[7] Biren B, Kamdar BB, Knauert MP, Jones SF, Parsons EC, Parthasarathy S, Pisani MA (2016) Perceptions and Practices Regarding Sleep in the Intensive Care Unit: A Survey of 1,223 Critical Care Providers. *Annals of the American Thoracic Society*, 13 (8), 1370-1377

[8] ANZICS APD data dictionary version 4.1 (cited 2016 October 11); Available from: http://www.anzics.com.au/Pages/CORE/data-tools.aspx

[9] Teasdale G, Jennett B. (1974) Assessment of coma and impaired consciousness. A practical scale. *Lancet*, 2, 81-4.

[10] Teasdale G, Maas A, Lecky F, Manley G, Stocchetti N, Murray M. (2014) The Glasgow Coma Scale at 40 years: standing the test of time. *The Lancet Neurology*, 13, 844 – 54

[11] Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. (1991) The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*, 100 (6), 1619-36

[12] Paul E, Bailey M, Van Lint A, Pilcher DV. (2012) Performance of APACHE III over time in Australia and New Zealand: a retrospective cohort study, *Anaesthesia and Intensive Care, 40 (6), 980-994*

[13] Kasza J, Wolfe R. (2014) Interpretation of commonly used statistical regression models. *Respirology, 19, 14-21*

[14] Goldstein H. (2010) *Multilevel Statistical Models*, 4th ed. Wiley series in probability and statistics. Chichester, West Sussex: Wiley

[15] Chung H, Beretvas N. (2012) The impact of ignoring multiple membership data structures in multilevel models. *British Journal of Mathematical and Statistical Psychology,* 65 (2), 185-200

[16] Steyerberg EW. (2009) *Clinical Prediction Models*, New York: Springer.

[17] Kasza J, Moran JL, Solomon PJ (2013) Evaluating the performance of Australian and New Zealand intensive care units in 2009 and 2010. *Statistics in Medicine*, 32, 3720–3736

[18] Kasza J. (2015) Stata tip 125: Binned residual plots for assessing the fit of regression models for binary outcomes. *The Stata Journal,* 15 (2), 599–604

[19] Gelman A, Hill J. (2007) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press

[20] DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S, Mark DB. (1997) Comparing risk-adjustment methods for provider profiling. *Statistics in Medicine*, 16 (23), 2645–2664.

[21] Mohammed AM, Mankeltow BN, Hofer TP. (2016) Comparison of four methods for deriving hospital standardised mortality ratios from a single hierarchical logistic regression model. *Statistical Methods in Medical Research*, 25 (2), 706–715
[22] Glance LG, Dick AW, Osler TM, Mukamel D. (2003) Using hierarchical modeling to measure ICU quality. *Intensive Care Medicine*, 29 (12), 2223–2229
[23] Render ML, Kim M, Deddens J, Siva S, Welsh DE, Bickel K, et al. (2005) Variation in outcomes in Veterans Affairs intensive care units with a computerized severity measure. *Critical Care Medicine*, 33 (5), 930–939
[24] Carpenter J, Bithell J. (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine,* 19 (9), 1141-1164