

RESEARCH ARTICLE

Open Access

A review of statistical estimators for risk-adjusted length of stay: analysis of the Australian and new Zealand intensive care adult patient data-base, 2008–2009

John L Moran^{1*} and Patricia J Solomon² for the ANZICS Centre for Outcome and Resource Evaluation (CORE) of the Australian and New Zealand Intensive Care Society (ANZICS)

Abstract

Background: For the analysis of length-of-stay (LOS) data, which is characteristically right-skewed, a number of statistical estimators have been proposed as alternatives to the traditional ordinary least squares (OLS) regression with log dependent variable.

Methods: Using a cohort of patients identified in the Australian and New Zealand Intensive Care Society Adult Patient Database, 2008–2009, 12 different methods were used for estimation of intensive care (ICU) length of stay. These encompassed risk-adjusted regression analysis of firstly: log LOS using OLS, linear mixed model [LMM], treatment effects, skew-normal and skew-t models; and secondly: unmodified (raw) LOS via OLS, generalised linear models [GLMs] with log-link and 4 different distributions [Poisson, gamma, negative binomial and inverse-Gaussian], extended estimating equations [EEE] and a finite mixture model including a gamma distribution. A fixed covariate list and ICU-site clustering with robust variance were utilised for model fitting with split-sample determination (80%) and validation (20%) data sets, and model simulation was undertaken to establish over-fitting (Copas test). Indices of model specification using Bayesian information criterion [BIC: lower values preferred] and residual analysis as well as predictive performance (R^2 , concordance correlation coefficient (CCC), mean absolute error [MAE]) were established for each estimator.

Results: The data-set consisted of 111663 patients from 131 ICUs; with mean(SD) age 60.6(18.8) years, 43.0% were female, 40.7% were mechanically ventilated and ICU mortality was 7.8%. ICU length-of-stay was 3.4(5.1) (median 1.8, range (0.17-60)) days and demonstrated marked kurtosis and right skew (29.4 and 4.4 respectively). BIC showed considerable spread, from a maximum of 509801 (OLS-raw scale) to a minimum of 210286 (LMM). R^2 ranged from 0.22 (LMM) to 0.17 and the CCC from 0.334 (LMM) to 0.149, with MAE 2.2-2.4. Superior residual behaviour was established for the log-scale estimators. There was a general tendency for over-prediction (negative residuals) and for over-fitting, the exception being the GLM negative binomial estimator. The mean-variance function was best approximated by a quadratic function, consistent with log-scale estimation; the link function was estimated (EEE) as 0.152(0.019, 0.285), consistent with a fractional-root function.

Conclusions: For ICU length of stay, log-scale estimation, in particular the LMM, appeared to be the most consistently performing estimator(s). Neither the GLM variants nor the skew-regression estimators dominated.

* Correspondence: john.moran@adelaide.edu.au

¹Department of Intensive Care Medicine, The Queen Elizabeth Hospital, Woodville SA 5011, Australia

Full list of author information is available at the end of the article

Background

Length of stay during an intensive care unit (ICU) or hospital admission is a function of diverse patient and organisational input variables [1]. It is widely used as an indicator of performance [2] and is a determinant of costs, although resource allocation is also known to affect length of stay [3]. Not surprisingly, ICU length of stay has been the subject of frequent analysis [4-9], with the majority of studies presenting cross-sectional analyses over a relatively short periods of months [10] to 1-2 years [9].

ICU patient length of stay (and costs) demonstrate skewed distribution and various statistical modelling strategies have been employed in analysis of such data [11-14]; albeit linear regression (ordinary least squares regression, OLS) of the logged dependent variable has demonstrated a remarkable persistence [15]. Individual patient data, as accessed from ICU data-bases, have an intrinsic hierarchical structure (patients within ICUs) and due analytic consideration of this structure is also appropriate [16]. Using such data from the Australian and New Zealand Intensive Care (ANZICS) adult patient database (APD) [17], calendar years 2008-2009, the purpose of this paper was to: (i) compare the *predictive performance and model specification* of conventional estimators for skewed data (ICU length-of-stay); OLS (with both raw and log-scaled dependent variable) and generalised linear models (GLMs [14]), with more innovative approaches: multilevel or hierarchical linear mixed models (LMM) incorporating random effects [11,16]; extended generalised linear models (EEE) with flexible link and variance functions [18]; estimators utilising skew-normal and skew-*t* multivariate distributions [19]; and finite mixture (FMM) models which consider the dependent variable as a mixture of distributions [20-22]; and (ii) determine the *effect of (ICU) mortality outcome upon length of stay*, allowing for "endogenous variable bias" [23] by means of a treatment-effects model [24,25] where specific allowance is made for the correlation of independent predictors and error terms. Under such conditions, model coefficients are biased [26].

Methods

The ANZICS adult patient database (APD) is a bi-national (Australia and New Zealand) voluntary data collection of individual ICU admissions which commenced in 1990. The data set requirements are specified in a data dictionary [27]. The current data-base does not incorporate coronary care units. Data is collected at the individual ICU and uploaded to the central repository (ANZICS APD) for processing and quality assurance; which consists of a cycle of error and exception checks, site feed-back, resubmission and incorporation into a final reporting data-set.

The ANZICS APD was interrogated to define an appropriate patient set, over the time period 2008-2009. Physiological variables collected were the worst in the first 24 hours after ICU admission. All first ICU admissions to a particular hospital for the period 2008-2009 were selected. Exclusions: patients with an ICU length of stay ≤ 4 hours; and patients aged < 16 years of age. Specific attention was directed to the fidelity of severity of illness records; in particular the scoring of the Glasgow Coma Score (GCS). Records were used only when all three components of the GCS were provided. Records where all physiological variables were missing were excluded and for the remaining records, missing variables were replaced to the normal range and weighted accordingly [28]. ICU length of stay, initially recorded in hours was transformed to days; patients with an ICU length of stay > 60 days were not considered, no formal trimming methods were employed [29].

Statistical methods

Continuous variables were reported using mean (SD), except where otherwise indicated; categorical variables were analysed using the Chi-squared test. To avoid the confounding effect of calendar time, patient data over the two years 2008 and 2009 was pooled. Distributional form of continuous variables of interest was by means of histogram, hanging rootogram [30,31] or violin plots [32]. The "hanging root", an alternative to the histogram, compares an empirical distribution with a theoretical distribution by means of "hanging" spikes from the theoretical distribution. The spike lengths represent bin counts, using the square root of the frequencies to stabilise the sampling variation across bins and make deviations in the tails, where the counts are small, more visible. Deviations are shown as deviations from a horizontal line ($y=0$) instead of deviations from a curve (the density function) in order to facilitate identification of patterns of deviation. The violin plot is a modification of the box plot that adds plots of the estimated kernel density to the summary statistics displayed by box plots [33], incorporates a marker for the median of the data, a box indicating the interquartile range, and spikes extending to the upper- and lower-adjacent values; overlaid is a density plot. Kernel density estimates are modifications of the histogram (a "smoothed" histogram), where densities are the continuous analogues of proportions (derivatives of the cumulative distribution function, so that areas under the density function read off as probabilities) [34]. Kurtosis (a measure of the heaviness of the tails of the distribution) and skewness were quantitatively expressed (a normal distribution having a skewness of 0 and a kurtosis of 3) according to standard formulae [35].

Predictor variables considered were

- a) Continuous: age (in years), severity of illness score (Acute Physiology and Chronic Health Evaluation

(APACHE) III score [36]); both centred for computation. The predictive effect of these variables was entered initially as both linear and quadratic.

- b) Categorical: these were parameterized as indicator variables with the reference level ($\equiv 0$) indicated in parentheses in the following list: gender (female); mechanical ventilation (not ventilated); ICU level, as defined in the ANZICS database, as Rural/Regional, Metropolitan, Tertiary and Private (Tertiary); State of origin; that is New Zealand and the States of the Commonwealth of Australia (New South Wales (NSW), the largest contributor); patient surgical status as post-elective surgery, post-emergency surgery and non-surgical (non-surgical); descriptors of ICU admission primary organ system dysfunction, these being a consolidation of the “diagnostic categories” of the APACHE algorithms: cardiovascular, gastrointestinal, metabolic, neurologic, respiratory, trauma, renal/genitourinary (cardiovascular); mean annual ICU volume across all ICUs ($0 < \text{mean volume}$; $1 > \text{mean volume}$); Died-in-ICU ($0 = \text{Alive}$, $1 = \text{Died}$)

Clinically meaningful combinations of variables and their interactions were assessed for effect; to preserve clinical transparency, higher order interactions were explored, but not entertained in the final model. The potential for multiple colinearity was tested using the variance inflation factor (VIF) and condition number (CN); where $VIF < 10$ and CN less than “30 or more” [37] were desirable.

ICU length of stay was modelled (with a fixed covariate list, and clustering on ICU site with robust variance [38]) via the estimators;

1. OLS using length of stay as a raw and a logged dependent variable. For the log OLS the expected value, $E(y)$, is proportional to $\exp(x'\beta)$ and the variance is assumed to be proportional to the mean squared. Prediction is on the log-scale (the geometric mean) and re-transformation is dependent upon the distribution of the error term: if normally distributed $\hat{u}(x_i) = \exp(x'\hat{\beta}) + 0.5\hat{\sigma}_\epsilon^2$ [39,40]; under homoscedasticity $\hat{\mu}(x_i) = \hat{\phi} \cdot \exp(x'\hat{\beta})$ where $\hat{\phi}$ is the estimated smearing factor and is usually between 1 and 4 [41,42].
2. LMM (logged length of stay) using maximum likelihood for model estimates. Potential modifying covariates were computed as fixed effects; ICU-year units as random intercepts (or “levels”) and random coefficients (“slopes”: APACHE III and APACHE III squared, age and ventilation status) were incorporated into the model fit.
3. A treatment effects model, log-dependent variable, via the Stata™ module “treatreg” as initially described

by Cong and Drukker [43]. A treatment-effects model is a two-equation system estimator (non-linear probit [44] and linear OLS), in which the effect of an endogenous binary variable (in this case, “Died-in-ICU”) on the continuous dependent variable is estimated by maximum likelihood [24,45]. Formally, the model is expressed in two equations [43]: the regression equation $y_i = x_i\beta + \delta z_j + \epsilon_j$, where z_j is the endogenous dummy variable indicating treatment assignment ($=1$, if $z_j^* > 0$; or not, $z_j^* = 0$). The outcome treatment-effect, estimated by the probit equation, is understood as being determined by an unobservable latent variable (z_j^*), a linear function of covariates (w_j), and a random component (u_j): $z_j^* = w_j\gamma + u_j$. The covariates entered into the outcome treatment equation were: age mechanical ventilation status, APACHE III score and its square, patient surgical status and the interaction with ventilation status, descriptors of ICU admission primary organ system dysfunction. As these covariates were commensurate with those of the OLS equation, a question of model identification arises [25]. However, as shown by Maddala, such a model is identified even if the error terms are not independent and the variable list is identical in the two equations [45].

4. GLMs [14] with a log link (relating the conditional mean to the covariates) and either a gamma, inverse Gaussian or Poisson family (specifying the relationship between the variance and the mean) [14,42] using maximum likelihood. The possibility of over-dispersion with the Poisson GLM was tested by calculating $z = \frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i / \sqrt{2}}$ and regressing z as a constant-only model; significance of z indicating over-dispersion [46]. In the presence of over-dispersion, the Poisson GLM model prediction and performance measures were replicated using a GLM negative binomial (NB2) model with log link [47]. For this model the variance function is given as $\mu + \alpha\mu^2$, where μ is the mean and α the heterogeneity or over-dispersion parameter, estimated by maximum likelihood (in GLM Poisson regression, $\alpha=0$). The negative binomial may be also be conceptualised as a “waiting time” distribution (the number of days to discharge), a generalisation of the geometric distribution [5].
5. EEE model allowing simultaneous estimation of flexible parametric link (Box-Cox function [48]) and variance functions, as implemented by Basu in the Stata™ user-written module “pglm”, using the raw dependent variable, scaled to the mean [18,49]. Expected values, conditional on covariates, are an inverse Box-Cox function of the linear predictor: $x'\beta = g(u_i; \lambda) = (\mu_i^\lambda - 1)/\lambda$ if $\lambda \neq 0$; $\log(\mu)$ if $\lambda = 0$. The default power variance function was used, which sets the family of variance functions ($h(\mu_i; \theta_1; \theta_2)$) to

$\theta_1\mu_i^{\theta_2}$. The variance functions include the Poisson, gamma and inverse Gaussian variance functions as special cases [50]; as with linear OLS, where $\lambda = 1$ and $\theta_2 = 0$ [51]. Estimation of regression and link parameters is via an extension of quasi-likelihood, and the variance parameters by additional estimating equations; that is, the EEE is a semi-parametric model. The link function transforms the mean of the outcome (not the outcome), overcoming the retransformation problem [50].

6. Regression utilising the skew-*t* and skew-normal distribution as implemented by Marchenko and Genton in the Stata™ user-written modules “skewtreg” and “skewnreg” respectively [19]. These distributions, fitted to the transformed (log) dependent variable, are (equivalently) the log-skew-*t* and log-skew-normal distributions by virtue of the correspondence which connects the normal and log-normal distributions [52,53]. The skew-normal distribution, beside location and scale parameters, has an additional shape parameter (α ; if $\alpha = 0 \equiv$ normal distribution) regulating distributional asymmetry; the skew-*t* distribution allows for asymmetry and heavier tails via shape (α) and degrees-of-freedom (ν) parameters [19]. Regression and other model parameters are estimated by maximum likelihood.

7. FMM, with two mixing components of a gamma and negative binomial (NB1 and NB2) densities, using the user-written Stata™ module “fmm”, version 2.0.0, by Deb [54]. In a FMM, the random variable is assumed to be drawn from a super-population that is an additive mixture of distinct subpopulations or classes (*C*) in proportions π_1, \dots, π_C , where $\sum_{j=1}^C \pi_j = 1$, $0 < \pi_j < 1 (j = 1, \dots, C)$ [20-22]. The *C*-point finite mixture model is given by $f(y_i|\Theta) = \sum_{j=1}^C \pi_j f_j(y_i|\theta_j)$, where the mixing probabilities π_j are estimated with other parameters (Θ) [20-22]. The θ_j is a parameter vector characterizing the component density function $f_j()$, the component density belonging to different parametric families (implemented in “fmm” are gamma, lognormal, negative binomial, normal, Poisson and Student-*t*). Estimation is by maximum likelihood.

Model adequacy was gauged by (i) progressive reduction in AIC (Akaike Information Criterion, for nested models) and BIC (Bayesian Information Criterion, for non-nested models) [55], both of which are penalised (with respect to number of observations and model parameters) likelihood methods for model determination (ii) likelihood ratio tests where appropriate and (iii) normality and lack of heteroscedasticity of residuals based upon graphical analysis [14].

Table 1 Demographics of ICU-hospital level, ventilation and surgical status: mean(SD)

ICU hospital Level/variable	Overall	Not-ventilated			Ventilated		
		Non-surgical	Elective surgical	Emergency surgical	Non-surgical	Elective surgical	Emergency surgical
Rural	(13.05%)						
ICU length of stay (days)	3.3 (4.6)	2.6 (3.1)	2.2 (2.6)	2.5 (3.6)	5.7(7.1)	4.7 (4.5)	5.4 (6.4)
APACHE III score	49.3(29.2)	42.6 (23.9)	36.4(16.1)	40.8 (20.8)	78.4 (33.6)	54.7 (25.3)	70.5 (30.8)
Age (years)	60.1 (19.5)	58.5 (20.4)	63.8 (16.1)	61.3 (20.5)	57.9 (18.3)	67.8 (15.1)	64.0(19.0)
% males	55	55	56	51	58	66	56
Metropolitan	(18.45%)						
ICU length of stay (days)	3.7 (5.2)	3.1 (4.3)	1.8 (2.2)	2.3 (3.4)	5.7 (6.8)	4.7 (5.8)	5.3 (6.6)
APACHE III score	55.5 (30.6)	50.4 (26.3)	37.3 (16.7)	44.9 (23.6)	76.4 (34.5)	55.5 (23.1)	65.5 (29.7)
Age (years)	59.9 (19.2)	58.6 (19.9)	63.8 (16.1)	61.3 (20.5)	57.7 (19.2)	66.0 (15.2)	63.3 (19.2)
% males	55	52	58	50	56	66	59
Tertiary	(44.49%)						
ICU length of stay (days)	3.9 (5.7)	3.0 (4.3)	1.5 (1.8)	2.4 (3.9)	5.9 (7.2)	2.7 (4.1)	5.5 (6.7)
APACHE III score	57.0(30.1)	51.6 (26.1)	38.7 (15.8)	45.2 (22.7)	73.4 (34.6)	48.3 (19.0)	60.8 (27.8)
Age (years)	58.3 (19.2)	57.3 (20.1)	62.8 (16.2)	59.7 (20.1)	54.6 (19.6)	64.2 (15.6)	57.9 (19.9)
% males	60	55	60	56	62	63	62
Private	(24.01%)						
ICU length of stay (days)	2.4 (3.7)	3.2 (4.5)	1.5 (1.8)	1.9 (2.7)	6.6 (7.9)	3.2 (3.9)	4.5 (6.6)
APACHE III score	43.7(22.4)	49.6 (22.9)	35.3 (15.2)	40.5 (19.1)	78.1 (32.1)	52.3 (21.7)	62.48 (29.4)
Age (years)	65.5 (16.0)	68.4 (16.9)	63.4 (16.0)	64.6 (17.9)	68.9 (15.7)	68.4 (13.8)	67.9 (16.1)
% males	57	52	54	47	54	61	49

Residuals were generated from the fitted model according to the scale of estimation [50]. The final model was developed on a determination set (80% of data) and validated on a validation set (20% of data); the random samples being stratified by (the 2) calendar-years. Predicted values were generated with continuous covariates centred and categorical covariates held at the reference category, as above, and in the presence of a logged dependent variable (length of stay) back-transformation to the original metric (days) utilised Duan's smearing estimate [41]. Distributional form of the raw-scale predictions (determination and validation) were displayed using violin plots. Final model performance was assessed by R^2 (on the "day" scale [56]) computed as the square of the correlation between predicted and observed length of stay [57]; Lin's concordance correlation coefficient (between raw and predicted raw length of stay [58]); mean absolute error (MAE); and root mean squared error (RMSE) [12,14]. Supplementary diagnostic analyses [15,40,50,51,59] were also performed, using:

- (i) user-written Stata™ code provided by Norton [60] and Jones [61]; the "GLM family" test (in GLMs there is accommodation of skewness, in particular, via the variance [42]: $\text{Var}[y|x] = \alpha \cdot [E(y|x)]$; therefore test if $\gamma = 0$ (Gaussian), $= 1$ (Poisson), $= 2$ (gamma), and $= 3$ (inverse Gaussian)); and the "COPAS" test: testing for over-fitting via split sample (50:50) cross-validation (1000x) using a version of the test initially formulated by Copas [62]; forecast predictions are regressed on

raw-scale length of stay and a significant difference of $\beta_{\text{predictions}}$ from unity suggests over-fitting [42]).

- (ii) With respect to the particular problem that raw-scale variance was a power function of the raw-scale mean function, the (modified [40,59]) Park test [63] was implemented. Squared residuals from a provisional model (GLM or OLS log-scale) were regressed (using a GLM with robust variance estimate and log-link) on the predictions (\hat{y}) from the same model, both log transformed:
 $\ln((y_i - \hat{y}_i)^2) = \gamma_0 + \gamma_1 \ln(\hat{y}_i) + v_i$; the coefficient on γ (see (i), above) indicating the GLM variance function.
- (iii) a modified Hosmer-Lemeshow goodness-of-fit (F) test [50,64], regressing the particular model residual distribution against deciles of the linear predictor (using OLS with robust variance). Graphical plots of the calibration of regression decile-parameter estimates against deciles of the linear predictor were undertaken; positive parameter estimates above the null-line ($\equiv 0$) indicate under-prediction and negative estimates indicate over-prediction [50].

As the interpretation of β coefficients is not invariant to model link functions [18], the average marginal effect [65,66] of ICU mortality upon length of stay ("Died-in-ICU") was estimated using the "margins" routine of Stata™ and the specific computations provided by Basu [49] for the EEE estimator. A marginal ("partial" or "incremental" [18]) effect

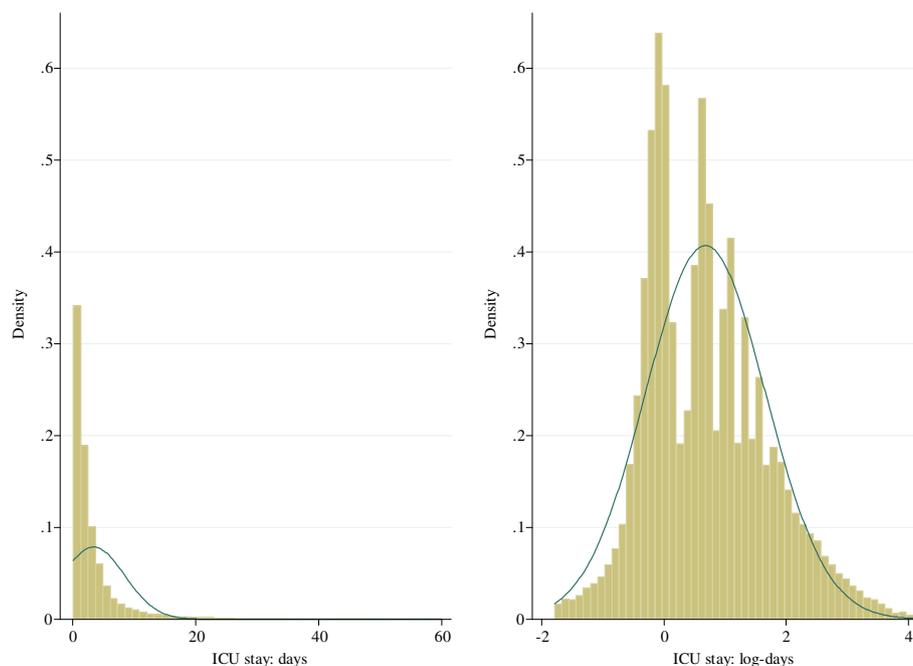
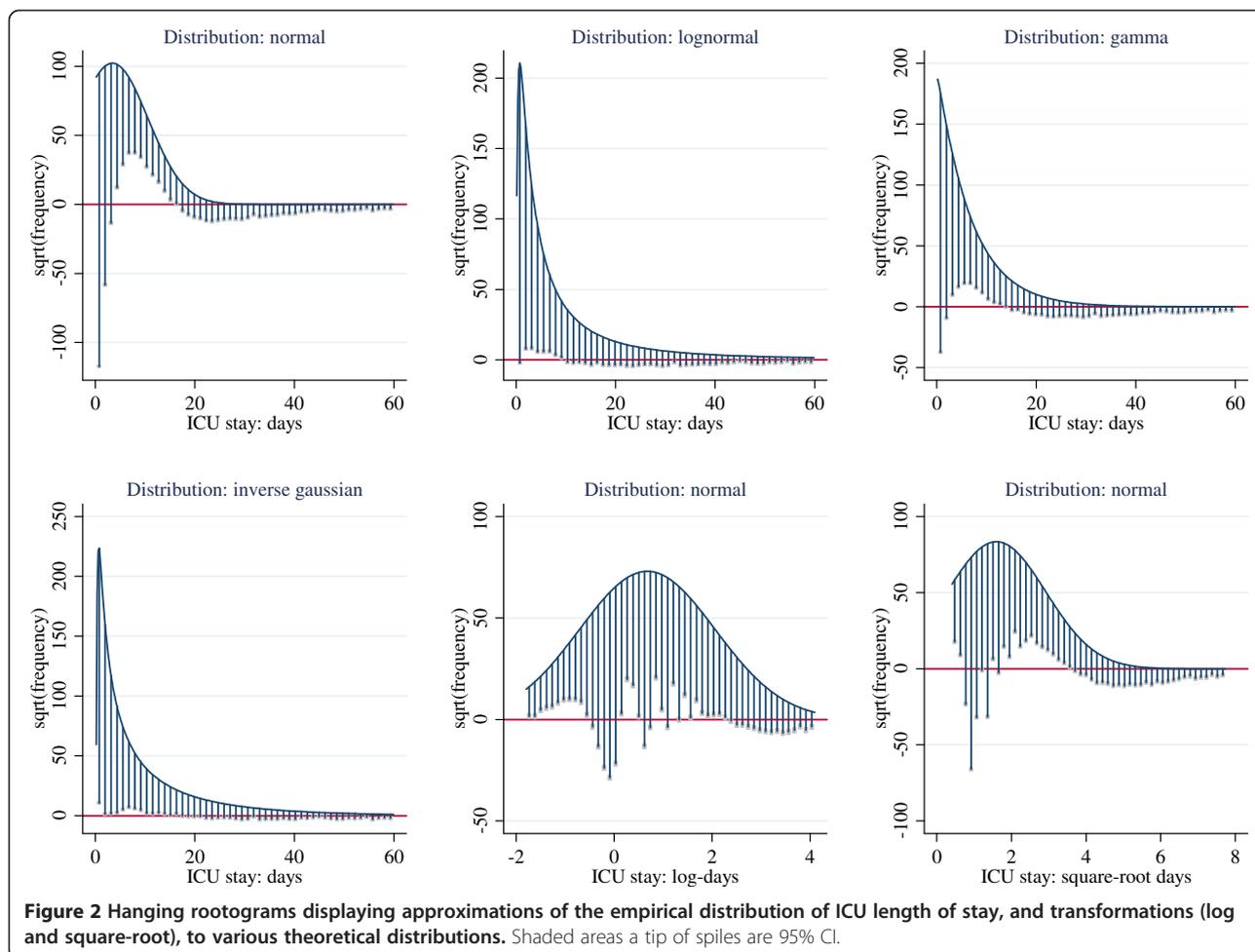


Figure 1 Histograms of ICU length of stay with normal curve overlay: left panel, raw scale; right panel, log transform.



measures the effect on the conditional mean of the regression dependent variable (“ y ”, \equiv ICU length-of-stay) of a change in a regressor(s) of interest (in this case the binary variable “Died-in-ICU”) [66]. The impact of progressive increases in patient severity of illness (APACHE III score) was elucidated by the “marginsplot” module of Stata™ [67], using the contrast across the 1 (\equiv Died)/0 (\equiv Alive) binary variable “Died-in-ICU”. Because of the technical problem of applying the Duan estimator directly to a contrast of the predicted logs in the log-scale estimators (exponentiating the difference and then multiplying by a function of the residuals would not produce an estimate directly comparable to the contrast produced after regression performed on raw-scale data), the illustrative plots were generated using raw-scale OLS and a log-link Gaussian family GLM.

Stata™ (Version 12 MP, 2011; Stata Corporation, College Station, Texas) statistical software was used.

Results

The database for the period 2008–2009 contained records of 114798 patients with exclusion (4.20%) of

patients having: ICU length of stay < 4 hours, incomplete GCS or APACHE III records. The final data set had missing ICU length of stay in 0.03% and 141 patients (0.001%) with ICU length of stay > 60 days (mean 81.98 (24.37), range 60.05-199.67) not considered in analysis. Patients ($n = 111663$) admitted to 131 ICUs in the *same* number of hospitals, were of mean(SD) age 60.6(18.8) years, APACHE III score 52.5(28.9); 43.0% were female, 40.7% were mechanically ventilated (on the day of admission) and ICU mortality was 7.8%. ICU length of stay was 3.4(5.1) (median 1.8, inter-quartile range 2.8 (0.93-3.7)) days. Summary statistics of length of stay, APACHE III score, age and gender by ventilation status and patient surgical type are shown in Table 1. Overall ICU length of stay and the log transform are displayed in histogram form with normal curve overlay in Figure 1 (left and right panels respectively); raw scale length of stay demonstrated marked kurtosis and right skew (29.4 and 4.4 respectively) whilst the log transform demonstrated a more symmetrical distribution (kurtosis 3.1 and skewness 0.5). Comparison of the empirical distribution of ICU length of stay and its’ log and square root

Table 2 Performance comparisons for various estimators

Performance index	BIC	R ²	R ²	CCC	CCC	MAE	MAE	RMSE	RMSE	Residual	Normality	Residual vs
Model	Determination set	Determination data-set (n = 86740)	Validation data-set (n = 21678)	Validation data-set	Determination data-set	Determination	Validation	Determination	Validation	P-P plot	Q-Q plot	fitted Symmetry
OLS: raw scale	509801.4	0.19	0.18	0.313	0.308	2.4	2.4	4.6	4.5	No	+ ve skew	No
OLS: log scale	216183.2	0.18	0.18	0.293	0.290	2.4	2.4	4.6	4.5	Yes	+ ve skew	Yes
LMM (log scale)	210285.8	0.22	0.20	0.334	0.328	2.3	2.3	4.5	4.5	Yes	+ ve skew	Yes
(Raw scale)	507492.8											
Treatment effects regression (log scale)	244883.2	0.18	0.17	0.296	0.294	2.4	2.4	4.6	4.5	Yes	+ ve skew	Yes
(Raw scale)	539291.4											
GLM: family(Poisson), link(log)	479143.3	0.19	0.18	0.321	0.314	2.4	2.4	4.5	4.5	No	+ ve skew	No
GLM: family(negbin), link(log)	375021.6	0.19	0.18	0.322	0.316	2.4	2.4	4.5	4.5	No	+ ve skew	No
GLM: family(gamma), link(log)	355449.9	0.19	0.18	0.322	0.317	2.4	2.4	4.6	4.5	No	+ ve skew#	No
GLM: family(inverse Gaussian), link(log)	365186.1	0.19	0.18	0.322	0.317	2.4	2.4	4.6	4.6	No	+ ve skew#	No
EEE:	Not estimable	0.19	0.18	0.320	0.316	2.4	2.4	4.6	4.6	No	+ ve skew#	No
Log-skew-t regression (Raw scale)	213465.5 349461.3	0.18	0.17	0.285	0.283	2.4	2.4	4.6	4.6	Yes	neg skew#	No
Log-skew-normal regression (Raw scale)	214216.4 436227.8	0.18	0.17	0.317	0.273	2.6	2.4	4.6	4.6	No	neg skew#	No
FMM: raw scale	329774.9	0.18	0.18	0.151	0.149	2.2	2.2	5.0	5.0	No	+ ve skew#	Yes

LLM; Linear mixed model (random coefficient model). CCC; concordance correlation coefficient. Negbin; negative binomial. P-P plot; standardized normal probability plot. Q-Q plot; plot of the quantiles of the residuals against the quantiles of the normal distribution . +ve skew#; marked positive skew. neg skew#; marked negative skew. Residuals vs fitted; increase spread of residuals with increment of fitted values (predicted length of stay). MAE; mean absolute error. RMSE; root mean square error. FMM; finite mixture model. R², CCC, MAE and RMSE were computed on the back-transformed "day" scale (for the linear mixed model, OLS-log, "treatreg" log-skew-t and log-skew-normal regression) using the Duan smearing estimate.

Table 3 Specification tests for various estimators

	Correlation	Residuals		Hosmer-Lemeshow (F) test		Copas test
	LP vs residuals	Skewness	Kurtosis	Determination	Validation	
OLS raw scale [Ⓢ]	1.00	4.33	32.74	0.0001	0.0001	0.996(0.993, 0.998)
OLS log scale [Ⓢ]	0.03	0.25	3.58	0.0001	0.0001	1.079(1.063, 1.094)
LMM log scale [Ⓢ]	0.008	0.25	3.76	0.0001	0.0001	1.054(1.036, 1.072)
Treatment effects model: log scale	0.150	0.27	3.58	0.0001	0.0001	1.043(1.025, 1.062)
GLM: family(Poisson), link(log) [#]	0.0001	2.37	13.41	0.0001	0.0001	0.940(0.899, 0.981)
GLM: family(negative binomial), link(log) [#]	0.0001	1.59	8.14	0.0001	0.0001	0.967(0.926, 1.008)
GLM: family(gamma), link(log) [#]	0.0001	1.26	6.60	0.0001	0.0001	0.923(0.882, 0.965)
GLM: family(inverse Gaussian), link(log) [#]	0.0002	0.20	4.98	0.0001	0.0001	0.776(0.717, 0.835)
EEE: raw scale	0.002	4.25	32.44	0.0001	0.002	0.956(0.957, 0.958)
Skew-t regression: log scale	0.0001	0.31	3.59	0.0001	0.0001	1.085(1.065, 1.105)
Skew normal regression: log scale	0.0001	0.32	3.56	0.0001	0.0001	1.134 (1.113, 1.155)
FMM: raw scale	0.0001	4.68	33.70	0.0001	0.0001	0.795(0.737, 0.852)

LP, linear predictor. [Ⓢ], standardised residuals [13]. [#], Deviance residuals [68].

transformation showed poor approximation to standard theoretical distributions (normal, lognormal and gamma; Figure 2).

Comparative performance of the estimators for ICU length of stay is seen in Tables 2 and 3; the regression models had 52 fixed parameters, including the constant term. Over the range of measures assessing predictive performance and model specification (including goodness-of-fit), the LMM demonstrated a consistently better performance compared with the other 11 estimators; albeit all estimators

revealed some problematic aspects of residual analyses. BIC for all the log-scale estimators (LMM, treatment effects- and log-skew-regressions) were considerably less than for estimation on the raw scale (Table 2). The final model variable set displayed a condition number of 19.4 and a mean VIF of 3.76; the best R² and concordance correlation coefficient was 0.22 and 0.334 respectively.

The outcome discrimination (Died-in-ICU) of the probit sub-equation of the treatment effects regression was excellent for both determination and validation sets (ROC curve

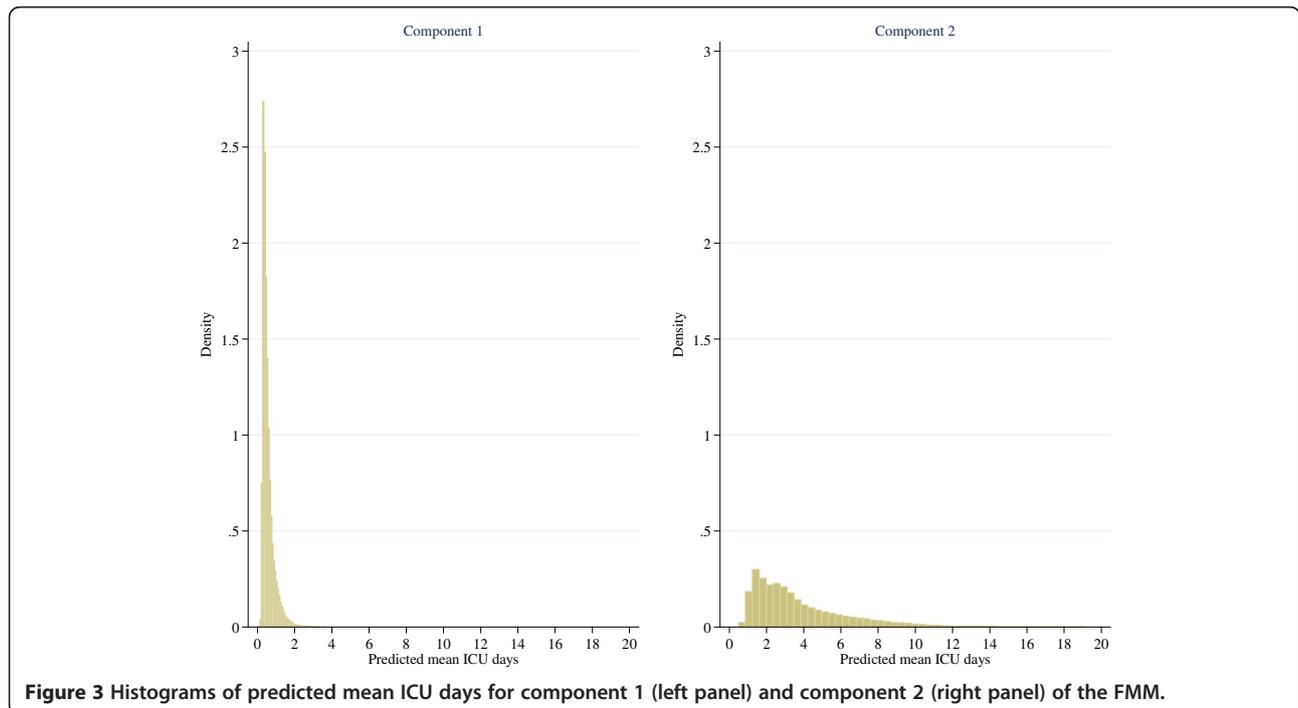


Figure 3 Histograms of predicted mean ICU days for component 1 (left panel) and component 2 (right panel) of the FMM.

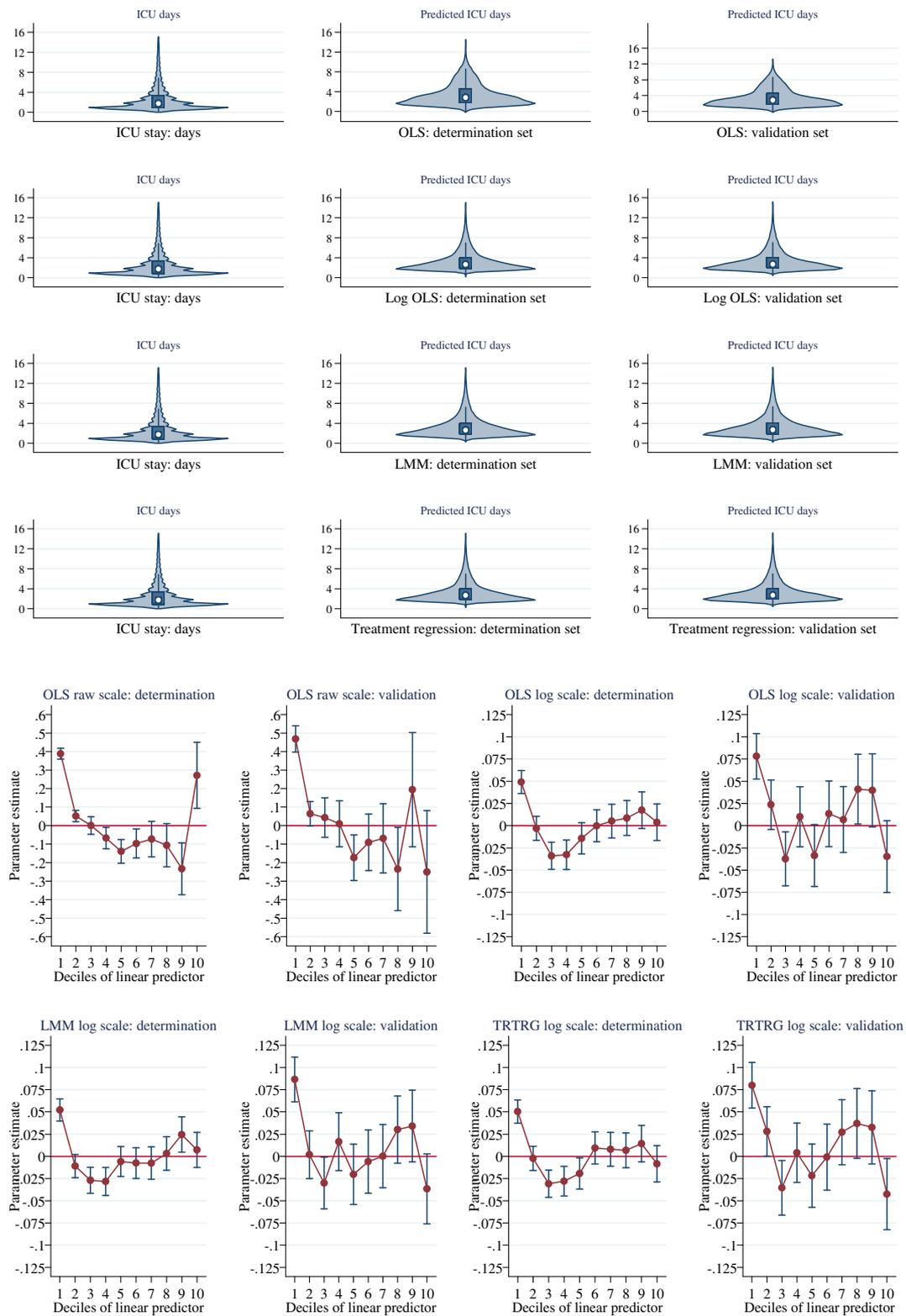


Figure 4 Upper panel: Violin plots of ICU length of stay (raw-scale, length of stay < 16 days) and predicted length of stay (raw-scale, < 16 days; determination and validation data sets) for estimators: OLS (raw-scale), log-scale OLS, LMM, and treatment effects regression. Lower panel: Hosmer-Lemeshow (F) test of residuals across deciles of linear-predictor, for estimators: OLS (raw-scale), log-scale OLS, LMM, and treatment effects regression. X-axis, linear predictor; Y-axis, parameter estimates.

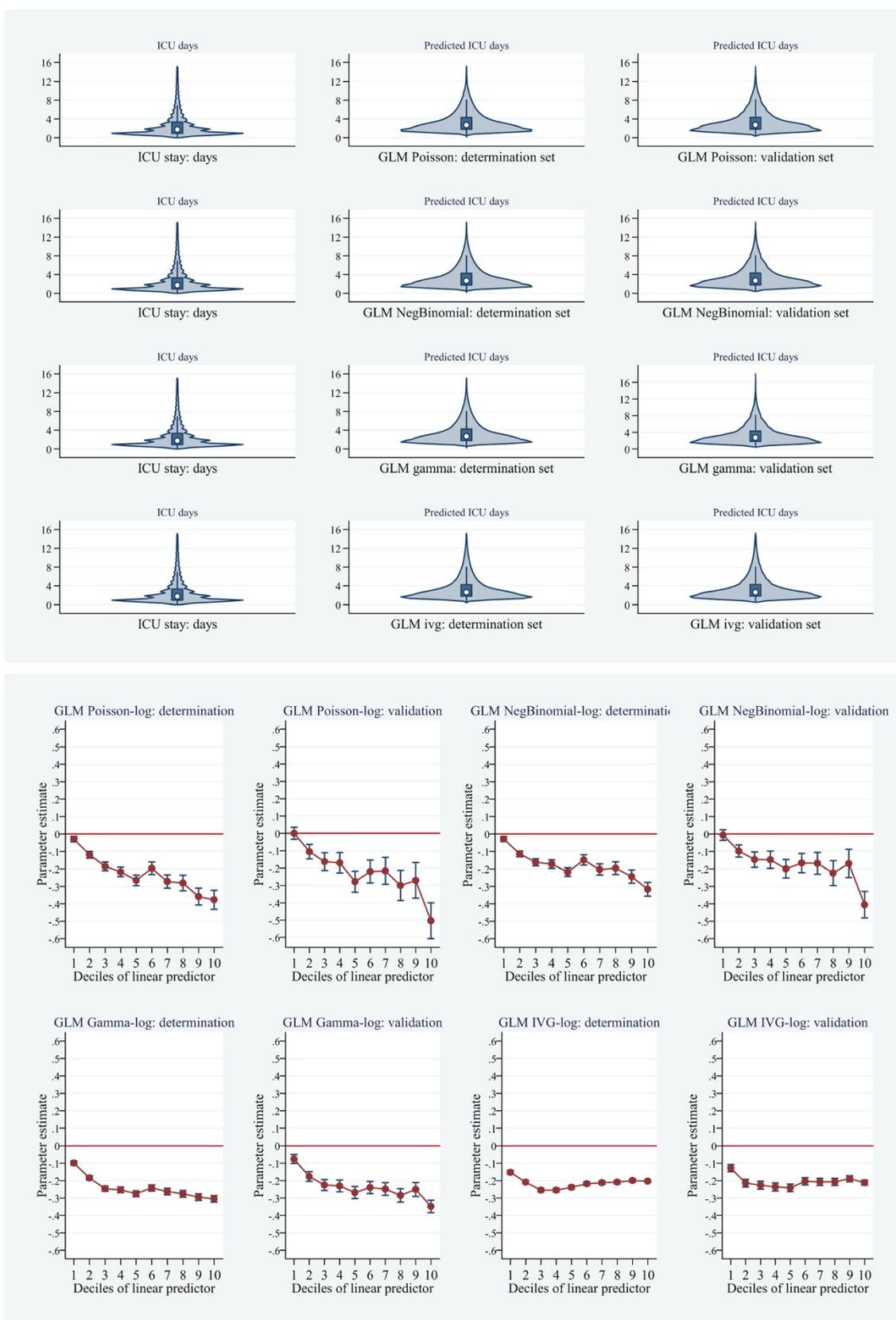


Figure 5 (See legend on next page.)

(See figure on previous page.)

Figure 5 Upper panel: Violin plots of ICU length of stay (raw-scale, length of stay < 16 days) and predicted length of stay (raw-scale, < 16 days; determination and validation data sets) for estimators: GLM Poisson-log, GLM negative binomial-log, GLM gamma-log, GLM inverse Gaussian-log. Lower panel: Hosmer-Lemeshow (F) test of residuals across deciles of linear-predictor, for estimators: GLM Poisson-log, GLM negative binomial-log, GLM gamma-log, GLM inverse Gaussian-log. X-axis, linear predictor; Y-axis, parameter estimates. determination. Validat., validation.

area 0.92). The null hypothesis, that the correlation ($\rho=0.065$) between the error terms of the separate estimations is zero, was rejected at $P=0.0001$. The Poisson GLM exhibited overdispersion (P value for “ z ” = 0.0001), the negative binomial GLM demonstrating an expected decrease in BIC; other performance measures were comparable between these two models. However, over-fitting, as assessed by the Copas test, was demonstrated in all other models except for the negative binomial GLM (Table 3). The “GLM family test” of the variance function of GLMs ($\text{Var}[y|x] = \alpha \cdot [E(y|x)]^\gamma$ [60]) rejected tests of γ for 0, 1, 2 and 3 ($P \leq 0.0001$). The modified Park test generated estimates for γ of: 0.348(0.333, 0.364) for log-scale OLS; 1.659(1.609, 1.708) for GLM Poisson-log; 2.106(2.005, 2.207) for GLM gamma-log; and 2.352(2.088, 2.616) for GLM inverse Gaussian-log. The EEE link parameter (λ) estimate was 0.152(0.019, 0.285) and the variance function (θ_2) was 2.115 (2.013, 2.218). The log-skew-normal and log-skew- t regressions reported α values of 1.073(0.933, 1.213) and 1.490 (1.323, 1.659) respectively, suggesting positive-skew; the degrees-of-freedom for the log-skew- t regression was 10.53 (7.97, 13.91) implying heavier-than-normal tails for the conditional distribution of (log) ICU length-of-stay. The 2 component FMM gamma model generated predicted mean ICU days of 0.56(0.32), range 0.06, 3.64 (component 1) and 3.69(2.45), range 0.49, 19.1 (component 2) as illustrated in Figure 3. The FMM negative binomial models demonstrated non-convergence.

Figures 4, 5, 6 show (i) upper panels: violin-plot distributional form of predictions and (ii) lower panels: Hosmer-Lemeshow-test parameter estimates, for determination and validation sets for each of the estimators, grouped by increasing scalar values of the Hosmer-Lemeshow parameter estimates. The graphical displays show a progressive lack-of-fit (Hosmer-Lemeshow test estimates) across Figures 4, 5, 6 in concert with corresponding over- and under-prediction as revealed by the violin plots. The LMM, log-scale OLS and treatment effects regression were the best calibrated; the EEE showed good calibration across the linear predictor deciles except for the upper deciles where substantial discordance was evident.

The β regression estimates for the binary variable “Died-in-ICU” are seen in Table 4, final column, and range from a low of 29% (treatment effects regression) to 295% (raw-scale OLS). The average marginal effects are also tabulated, and were generally decreased in magnitude compared with the β estimates. Figure 7 illustrates

the contrasts of predictive margins for “Died-in-ICU” across increments of APACHE III score for raw-scale OLS and log-link Gaussian family GLM (see also Table 4). Similar decreases (not shown) of the contrasts of predictive margins across the APACHE III score range were evident for the (log) linear predictor for estimators using log-ICU length of stay.

Discussion

Methodological recommendations for non-normal data analysis have usually been subsumed under the rubric of cost analysis [12,15,69]. The distribution of positive health expenditures (and length of stay) exhibit skew, kurtosis and heteroscedasticity, with a non-constant and increasing variance, albeit particular cost problems, such as the accommodation of a probability mass at “zero”, may not have immediate relevance to length of stay analysis [51,69]. This review has suggested that a LMM had better overall performance compared with 11 other estimators. However, a number of issues pertain to comparator studies of possible estimators; in particular data structure and the full assessment, or otherwise, of model specification.

The hierarchical nature of some clinical and/or administrative data sets and the requirement for appropriate analysis has been commented upon above. As opposed to the volume-outcome literature [70], mixed models would appear to have been applied in a relatively small number of studies for length of stay analysis [16,71-75]. The use of administrative data [5,16] may also be associated with integer based calendar-day recording which lacks the accuracy of fractional day based “exact” times [10,71].

Because of the non-normality of length of stay, formal trimming [76,77] or truncation [71,77,78] of the data, or deleting [5,73,79] “outliers” has been undertaken prior to possible data transformation. For instance, studies implementing the APACHE III [78] or IV [71] algorithms for predicting ICU length of stay have truncated the latter at 30 days, at a 1% data level; the same fraction as seen with data deletions [73]. The current study used a deletion fraction of 0.001%. The motivation for such strategies are various [80], but the theoretical basis of these data revisions has been questioned [29] and bias in the estimation of the mean has also been suggested [29,69]. More importantly, large values of length of stay are “true” values [69] and data reduction cannot necessarily be defended as

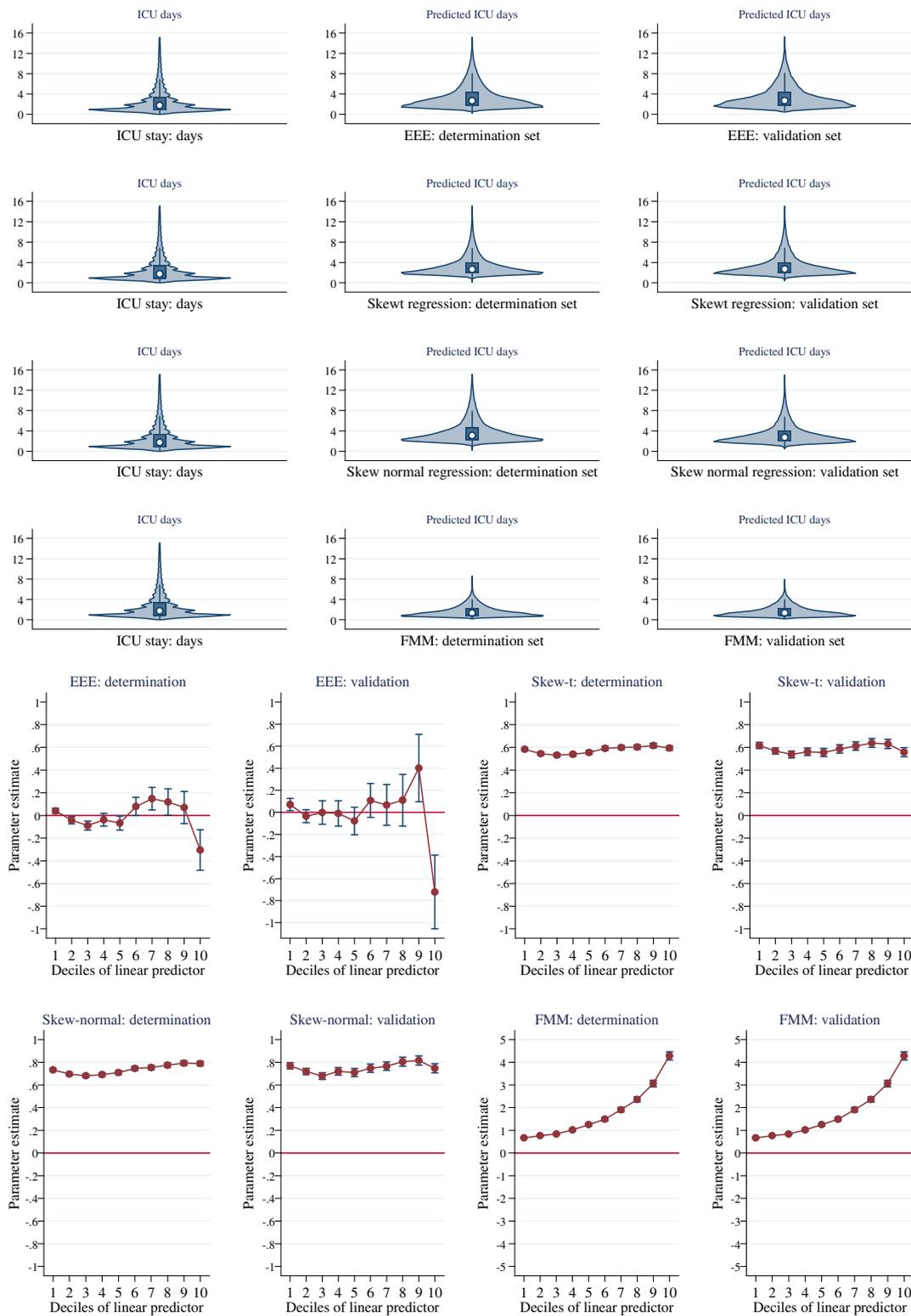


Figure 6 Upper panel: Violin plots of ICU length of stay (raw-scale, length of stay < 16 days) and predicted length of stay (raw-scale, < 16 days; determination and validation data sets) for estimators: EEE, log-skew-t regression, log-skew-normal regression and FMM. Lower panel: Hosmer-Lemeshow (F) test of residuals across deciles of linear-predictor, for estimators: EEE, log-skew-t regression, log-skew-normal regression and FMM. X-axis, linear predictor; Y-axis, parameter estimates.

Table 4 β regression estimate and average marginal effect for “Died in ICU”

	β estimate	Average marginal effect
OLS raw scale	2.947 (2.368, 3.526)	2.486 (2.060, 2.912)
OLS log scale	0.404 (0.302, 0.507)	0.353 (0.268, 0.437)
LMM log scale	0.367 (0.318, 0.416)	0.320 (0.285, 0.355)
Treatment effects regression: log scale	0.297 (0.194, 0.401)	0.215 (0.125, 0.302)
GLM: family(Poisson), link(log)	0.600 (0.444, 0.676)	1.456 (1.064, 1.848)
GLM: family(negative binomial), link(log)	0.596 (0.476, 0.715)	1.679 (1.239, 2.119)
GLM: family(gamma), link(log)	0.619 (0.497, 0.740)	1.802 (1.336, 2.269)
GLM: family(inverse Gaussian), link(log)	0.757 (0.608, 0.906)	2.648 (1.914, 3.382)
EEE: raw scale	0.678 (0.539, 0.817)	1.507 (1.042, 1.972)
Skew-t regression: log scale	0.376 (0.268, 0.483)	0.335 (0.250, 0.420)
Skew normal regression: log scale	0.314 (0.214, 0.413)	0.271 (0.195, 0.348)
FMM: raw scale, component 1	0.234 (0.121, 0.347)	0.130 (0.066, 0.195)
FMM: raw scale, component 2	0.648 (0.513, 0.784)	2.395 (1.871, 2.920)
GLM: family(Gaussian), link(log)	0.541 (0.421, 0.662)	1.285 (0.902, 1.667)

“...represent[ing] unusual cases...”, or because of “...analytical problems...” [73]. That trimming will cause decreases in residual variance in LMM was pointed out by Lee et al. [72] and was easily demonstrated in the current data set; where the residual variance was 0.646 for the full data, 0.600 at the 99th percentile trim (< 26 days) and 0.508 for the 95th percentile trim (< 12 days) of ICU length of stay. Model assessment must take note of data revisions.

Similarly, the repeated assertions [71,72,76,77,80,81] that the objective of such data manipulations are model (in particular, OLS) requirements for normality of the dependent variable are problematic. As observed by Buntin and Zaslavsky “Plotting of the data with various transformations is thus a useful preliminary step, but ultimately the distribution of the residuals from the transformed model is critical because the model assumptions concern the distribution of the residuals, not of the data” [59]. That the residuals “...usually have a similar distribution to the original data...” [12] does not absolve the analyst from model based residual interrogation; the simple reporting of overall measures such as R^2 [77] or agreement indices of observed versus predicted [71] are not satisfactory proxies. Moreover, inference on β -coefficients is biased under heteroscedasticity.

Effect of ICU death on length of stay.

The β coefficient for “Died-in-ICU”, as a main effect, ranged from +29.7% (treatment effects regression, log scale) to +300% (OLS, raw scale), consistent with the β estimates for the factor “Death” of +0.536 (age < 65 years) and +0.439 (age \geq 65 years) reported by Angus et al. [5], using a GLM (geometric family, identity link). However, the results of the marginal analysis (Table 4) implied over-prediction of the β -coefficient of “Died in ICU” across the

estimators, albeit the average marginal effect was considered, as opposed to the marginal effect at the mean or at a particular value [65,66]. Furthermore, there was a progressive decrease in the contrast, died in ICU versus alive in ICU, of predicted mean ICU days with an increase in the APACHE III score (Figure 7); this effect being consistent with findings from our previous study [82] and those of Rapaport et al. [8] and Woods et al. [9]. Of more interest was the demonstration, using the treatment effects regression model, that the covariate “Died-in-ICU” (which also has the status of an “outcome”) was associated with a significant correlation (ρ) between the error terms of the separate estimations (probit and linear regression) and was appropriately considered as endogenous [26]. That ρ was positive at 0.065 also indicated that OLS overestimated the “treatment” effect (Died-in-ICU); that is, a face-value interpretation of the β mortality coefficient was problematic. Endogeneity may also be suspected in mortality models where length of stay [7] or mortality probability [83] are entered as predictive covariates. Such regression of a variable upon its components has been termed a “dubious practice” [15]. This being said, the extension of 2-stage methods to non-linear models may not be without its own issues [15,84,85].

Estimators of length of stay.

The statistical basis of the various estimators used in the current paper has been canvassed in “Statistical methods”, above. These estimators, assessing the mean function $u(x)$, have seen extensive application and assessment in the cost data literature, but not necessarily for length of stay. The recent introduction of log-skew-normal and log-skew- t estimators has not afforded opportunity for extensive assessment. Within the ICU literature, the predominant estimators have been

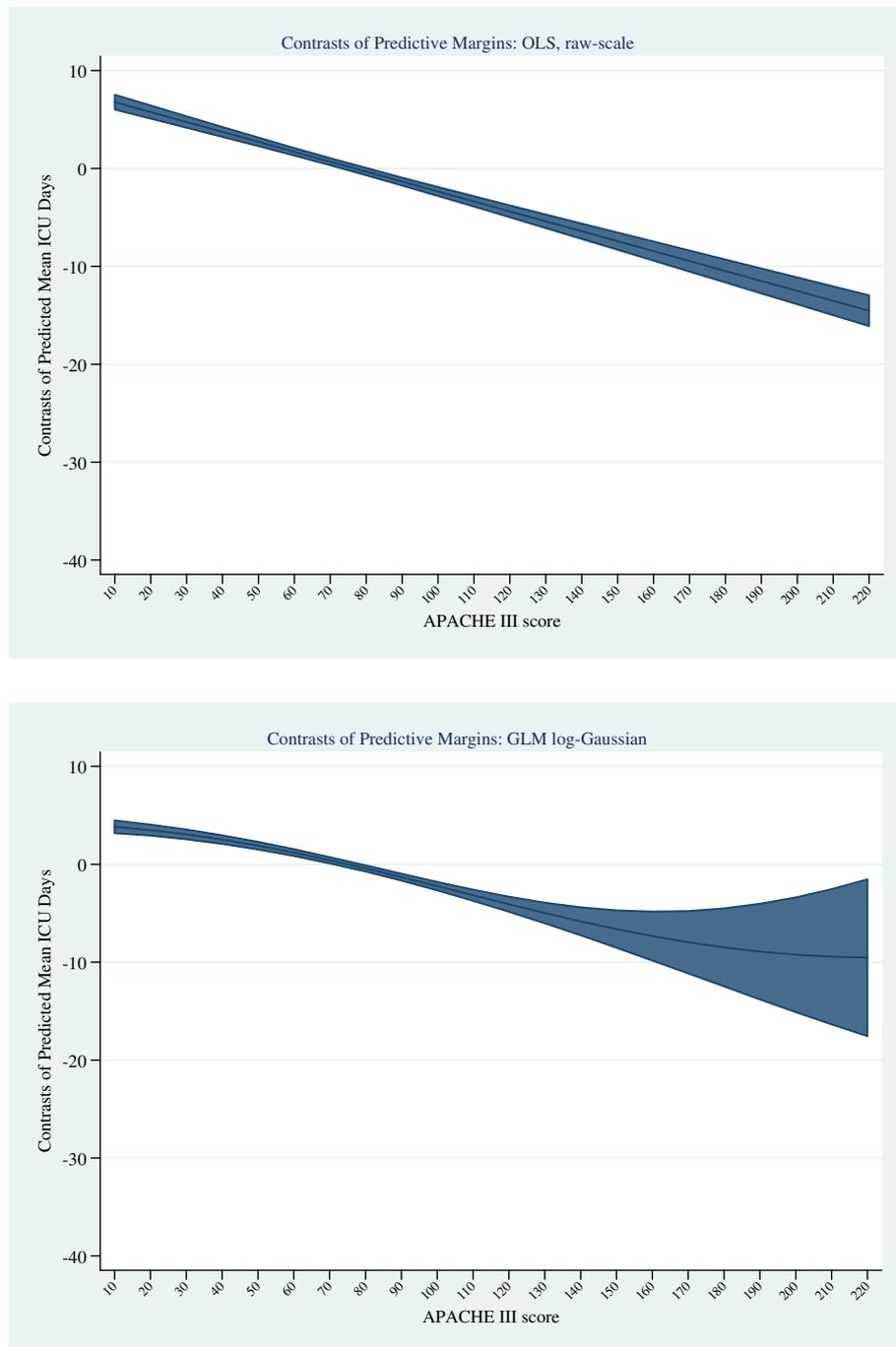


Figure 7 Contrasts of predictive margins (ICU days) across the binary variable “died in ICU” for OLS raw scale (upper panel) and GLM, Gaussian family and log link (lower panel).

ordinary least squares regression, with [8] and without [7,71,78] log transformation; generalized linear model (geometric family with identity link) [5] and random intercept linear model [83]. Lee and co-workers have explored alternative approaches suited to the analysis of hospital length of stay [74,86].

It is perhaps not surprising that in a large hierarchically structured data set that the LMM appeared to dominate. The suggestion of lack of discriminatory power across alternative estimators in cost analysis, albeit with small sample size [15], also appears to be vindicated in the current analysis. However, the further suggestion

that large sample size together with "...simple methods..." [69] and an identity link [12,15] may be analytically sufficient seems not to be the case with our data, although the LMM and log-OLS models were relatively easy to fit and interpret. The estimated link function ($\lambda=0.152(0.019, 0.285)$) supports neither an identity ($\lambda=1$) nor log link ($\lambda=0$), rather a fractional- but not a square-root ($\lambda=0.5$) function. Although the Park test and the EEE estimation of the variance function approximated 2, the GLM variants demonstrated poor goodness-of-fit (Figure 5) with the exception of the EEE (not withstanding the upper most decile of the linear predictor, Figure 6) where the link and variance functions were estimated, not fixed. The narrow confidence limits in the Copas test (Table 3) for the EEE estimator presumably reflects this re-estimation of link and variance functions in the cross-validation process. The residual behaviour of (i) the log scale estimators (Table 3) was generally satisfactory and empirical estimation suggested a quadratic mean-variance relationship which was consistent with the log-OLS estimator and (ii) the GLM models, was variable and with heavy tails. The use of gamma distribution in the presence of heavy tails has recently been cautioned [69]. Estimators on the raw scale (OLS and FMM) demonstrated consistent but sub-optimal residual behaviour; and despite modest BIC values, the log-skew estimators had an overall poor residual performance.

In the current study, the range of R^2 was modest (0.18-0.22; Table 2); consistent with that previously reported, 0.13 [78] to 0.21 [71,83], and the observations of Diehr et al. that R^2 for cost utilisation data are usually ≤ 0.20 [12]. Increases in R^2 have been reported, not surprisingly, across groupings; for instance, ICU units ($R^2 = 0.78$; [7]). In the current study, R^2 across ICU units ($n = 118$) was 0.92 (LMM, determination set).

Critique of methodology

The hierarchical structure of data under analysis is not uncommon in clinical or administrative data sets, but may not be seen in cost utilisation studies where survey and administrative data predominate. Thus cross paradigm performance assessments may lack validity, although the skew and kurtosis indices (4.4 and 29.4 respectively, in the current data) at least were comparable (4.1 and 25.6 [42], 4.4 and 50 [50]). The significance of the residual tests (Table 3) and the tendency to over-fitting must be interpreted in the context of the large data sets [87] and the reported adverse effect of outliers and skewed data on these tests [15,42]. We chose not to benchmark goodness of fit against the behaviour of raw residuals from the log-scale estimators. This preference was dictated by the application of homoscedastic retransformation from log-scale estimation which may

not have been optimal, leading to the generation of raw-scale residuals ($y_i - \hat{y}_i$) of uncertain status. Such uncertainty may be contrasted with the fixed relationship (via the inverse logit function) between the scale of estimation (log-odds) and the scale of interest (probability) in logistic regression where the Hosmer-Lemeshow test was first described. No generally applicable method of heteroscedastic retransformation recommends itself with large data sets and multiple predictor variables [39,42,51,59] and was not undertaken. Our analysis suggests that percentage changes in days are important, not absolute changes; this follows from the fact that the data satisfy important optimal statistical properties on the log scale, and that we can therefore estimate and interpret effects on this scale with some confidence.

Comparisons between estimators with respect to the significance (or otherwise), the effect magnitude and form (for continuous variables) of other covariates entered into the regressions (the covariates were fixed [11,59]) were not a focus of interest and were not reported. The relatively poor performance of the GLM model variants may be subject to enhancement by the use of generalised linear mixed models [88] which could incorporate the random effects structure of the LMM. Similarly, our use of the Poisson and negative binomial GLM fails to account for the structural exclusion of zero counts (both distributions include zeros) and use of zero-truncated count models may have been appropriate [47].

Conclusion

Model choice under the conditions of the current data set would appear to favour the LMM and log-OLS. The treatment effects model afforded extra explanation with respect to the covariate "Died-in-ICU" and had a good fit to the data. Mechanistic approaches to estimation are represented by the FMM which captures the bi-modal nature of the data (Figures 1 and 3), although the fit was not optimal; and the negative binomial GLM which represented the length of stay as "waiting times". Neither the log-skew-normal nor log-skew- t estimator would appear to recommend themselves as the preferred analytic approach for the current data set.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The study was conceived, designed, (data)-analysed, written and critically revised jointly by both authors (JLM, PJS). Both authors read and approved the final manuscript.

Author details

¹Department of Intensive Care Medicine, The Queen Elizabeth Hospital, Woodville SA 5011, Australia. ²School of Mathematical Sciences, University of Adelaide, Adelaide SA 5000, Australia. ³Australian and New Zealand Intensive Care Society, Carlton Victoria 3053, Australia.

Received: 14 October 2011 Accepted: 16 April 2012
Published: 16 May 2012

References

1. Burns LR, Wholey DR: The effects of patient, hospital, and physician characteristics on length of stay and mortality. *Med Care* 1991, **29**:251–271.
2. Thomas JW, Guire KE, Horvat GG: Is patient length of stay related to quality of care? *Hosp Health Serv Adm* 1997, **42**:489–507.
3. Needham DM, Anderson G, Pink GH, McKillop I, Tomlinson GA, Detsky AS: A province-wide study of the association between hospital resource allocation and length of stay. *Health Serv Manage Res* 2003, **16**:155–166.
4. Afessa B, Keegan MT, Hubmayr RD, Naessens JM, Gajic O, Long KH, Peters SG: Evaluating the performance of an institution using an intensive care unit benchmark. *Mayo Clin Proc* 2005, **80**:174–180.
5. Angus DC, Linde-Zwirble WT, Sirio CA, Rotondi AJ, Chelluri L, Newbold RC III, Lave JR, Pinsky MR: The effect of managed care on ICU length of stay: implications for medicare. *JAMA* 1996, **276**:1075–1082.
6. Becker RB, Zimmerman JE, Knaus WA, Wagner DP, Seneff MG, Draper EA, Higgins TL, Estafanous FG, Loop FD: The use of APACHE III to evaluate ICU length of stay, resource use, and mortality after coronary artery by-pass surgery. *J Cardiovasc Surg (Torino)* 1995, **36**:1–11.
7. Knaus WA, Wagner DP, Zimmerman JE, Draper EA: Variations in mortality and length of stay in intensive care units. *Ann Intern Med* 1993, **118**:753–761.
8. Rapoport J, Teres D, Zhao Y, Lemeshow S: Length of stay data as a guide to hospital economic performance for ICU patients. *Med Care* 2003, **41**:386–397.
9. Woods AW, MacKirdy FN, Livingston BM, Norrie J, Howie JC: Evaluation of predicted and actual length of stay in 22 Scottish intensive care units using the APACHE III system. *Acute Physiology and Chronic Health Evaluation. Anaesthesia* 2000, **55**:1058–1065.
10. Marik PE, Hedman L: What's in a day? Determining intensive care unit length of stay. *Crit Care Med* 2000, **28**:2090–2093.
11. Austin PC, Rothwell DM, Tu JV: A comparison of statistical modeling strategies for analyzing length of stay after CABG surgery. *Health Services & Outcomes Research Methodology* 2002, **3**:107–133.
12. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY: Methods for analyzing health care utilization and costs. *Annu Rev Public Health* 1999, **20**:125–144.
13. Manning WG, Basu A, Mullahy J: Generalized modeling approaches to risk adjustment of skewed outcomes data. *J Health Econ* 2005, **24**:465–488.
14. Moran JL, Solomon PJ, Peisach AR, Martin J: New models for old questions: Generalized Linear Models for cost prediction. *J Eval Clin Pract* 2007, **13**:381–389.
15. Basu AP, Manning WGP: Issues for the next generation of health care cost analyses. *Med Care* 2009, **47**:S109–S114.
16. Carey K: Hospital length of stay and cost: a multilevel modeling analysis. *Health Services & Outcomes Research Methodology* 2002, **3**:41–56.
17. Stow PJ, Hart GK, Higlett T, George C, Herkes R, McWilliam D, Bellomo R: Development and implementation of a high-quality clinical database: the Australian and New Zealand intensive care society adult patient database. *J Crit Care* 2006, **21**:133–141.
18. Basu A, Rathouz PJ: Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostat* 2005, **6**:93–109.
19. Marchenko YV, Genton MG: A suite of commands for fitting the skew-normal and skew-t models. *The Stata Journal* 2011, **10**:507–539.
20. Conway KS, Deb P: Is prenatal care really ineffective? Or, is the 'devil' in the distribution? *J Health Econ* 2005, **24**:489–513.
21. Deb P, Holmes AM: Estimates of use and costs of behavioural health care: a comparison of standard and finite mixture models. *Health Econ* 2000, **9**:475–489.
22. Singh CH, Ladusingh L: Inpatient length of stay: a finite mixture modeling analysis. *Eur J Health Econ* 2010, **11**:119–126.
23. Terza J: Estimating endogenous treatment effects in retrospective data analysis. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research* 1999, **2**:429–434.
24. Guo S, Fraser MW: Sample selection and related models. In *Propensity Score Analysis: Statistical Methods and Applications*. Edited by Guo S, Fraser MW. Thousand Oaks: SAGE Publications, Inc; 2010:85–124.
25. Wooldridge JM: Limited dependent variable models and sample selection corrections. In *Introductory econometrics: A modern approach*. Third edition. Edited by Wooldridge JM, Mason OH. Thomson South-Western; 2006:582–631.
26. Hilbe JM: Handling endogeneity and latent class models. In *Negative Binomial Regression*. 2nd edition. Cambridge: Cambridge University Press; 2011:407–446.
27. ANZICS: Adult Data Base. Data Dictionary Version 1.5. 2005. @ http://www.anzics.com.au/admc/files/data_dictionary.pdf; Accessed, June 2006.
28. Wagner D, Knaus W, Bergner M: Statistical-Methods. *Crit Care Med* 1989, **17**:S194–S198.
29. Lee AH, Xiao J, Vemuri SR, Zhao Y: A discordancy test approach to identify outliers of length of hospital stay. *Stat Med* 1998, **17**:2199–2206.
30. Buis ML: HANGROOT: Stata module creating a hanging rootogram comparing an empirical distribution to the best fitting theoretical distribution. @ <http://econpapers.repec.org/scripts/search/search.asp?ft=hangroot> 2011, Accessed June 2011.
31. Wainer H: The suspended rootogram and other visual displays: an empirical validation. *The American Statistician* 1974, **28**:143–145.
32. Winter N, Nichols A: VIOPLOT: Stata module to produce violin plots. @ <http://econpapers.repec.org/scripts/search/search.asp?ft=hangroot> 2010, Accessed June 2010.
33. Hintze JL, Nelson RD: Violin plots: a box plot-density trace synergism. *The American Statistician* 1998, **52**:181–184.
34. Cox NJ: Speaking Stata: density probability plots. *Stata Journal* 2005, **5**:259–273.
35. Cox NJ: Speaking Stata: the limits of sample skewness and kurtosis. *The Stata Journal* 2010, **10**:492–495.
36. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A: The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991, **100**:1619–1636.
37. Belsley DA: *Conditioning diagnostics, collinearity and weak data in regression*. New York: John Wiley & Sons; 1991.
38. Rogers W: sg17: Regression standard errors in clustered samples. *Stata Technical Bulletin Reprints* 1993, **3**:88–94.
39. Manning WG: The logged dependent variable, heteroscedasticity, and the retransformation problem. *J Health Econ* 1998, **17**:283–295.
40. Manning WG, Mullahy J: Estimating log models: to transform or not to transform? *J Health Econ* 2001, **20**:461–494.
41. Duan N: Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc* 1983, **78**:605–610.
42. Jones AM: Models for Health Care. *Health Econometrics and Data Group Working Paper 10/01* 2010. @ york.ac.uk/res/herc/hedgwp, Accessed May 2011.
43. Cong R, Drukker DM: Treatment-effects model. *Stata Technical Bulletin Reprints* 2000, **10**:159–169.
44. Cheung YB: Adjustment for selection bias in cohort studies: an application of a probit model with selectivity to life course epidemiology. *J Clin Epidemiol* 2001, **54**:1238–1243.
45. Maddala GS: Multivariate qualitative variables. In *Limited-Dependent and Qualitative Variables in Econometrics*. Edited by Maddala GS. Cambridge: Cambridge University Press; 1983:93–148.
46. Hardin J, Hilbe J: The Poisson family. In *Generalized linear models and extensions*. 2nd edition. College Station: TX: Stata Press; 2007:183–198.
47. Hilbe JM: *Negative Binomial Regression*. 2nd edition. Cambridge, UK: Cambridge University Press; 2011.
48. Box GEP, Cox DR: An Analysis of Transformations. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 1964, **26**:211–252.
49. Basu A: Extended generalized linear models: Simultaneous estimation of flexible link and variance functions. *Stata Journal* 2005, **5**:501–516.
50. Basu A, Arondekar BV, Rathouz PJ: Scale of interest versus scale of estimation: comparing alternative estimators for the incremental costs of a comorbidity. *Health Econ* 2006, **15**:1091–1107.
51. Hill SC, Miller GE: Health expenditure estimation and functional form: applications of the generalized gamma and extended estimating equations models. *Health Econ* 2010, **19**:608–627.
52. Azzalini A, dalCapello T, Kotz S: Log-skew-normal and log-skew-t distributions as models for family income data. *Journal of Income Distribution* 2003, **11**:12–20.
53. Walls WD: Modelling heavy tails and skewness in film returns. *Applied Financial Economics* 2005, **15**:1181–1188.

54. Deb P, FMM: *Stata module to estimate finite mixture models*: 2008. @ <http://fmwww.bc.edu/repec/bocode/ffmm> 2008, Accessed November 2010.
55. Kuha J: **AIC and BIC. Comparisons of assumptions and performance.** *Sociological Methods & Research* 2005, **33**:188–229.
56. Scott A, Wild C: **Transformations and R^2 .** *The American Statistician* 1991, **45**:127–129.
57. Thomas JW, Ashcraft ML: **Measuring severity of illness: six severity systems and their ability to explain cost variations.** *Inquiry* 1991, **28**:39–55.
58. Lin LI: **A concordance correlation coefficient to evaluate reproducibility.** *Biometrics* 1989, **45**:255–268.
59. Buntin MB, Zaslavsky AM: **Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures.** *J Health Econ* 2004, **23**:525–542.
60. Norton EC: **Stata code for modeling health care costs.** @ <http://www.unc.edu/~enorton/> 2005, Accessed January 2011.
61. Jones AM: **Modelling non-normal outcomes using linear models.** @ <http://melbourneinstitute.com/health/WorkshopApr202010.html> 2010, Accessed May 2011.
62. Copas JB: **Regression, prediction and shrinkage.** *Journal of the Royal Statistical Society Series B* 1983, **45**:311–354.
63. Park RE: **Estimation with heteroscedastic error terms.** *Econometrica* 1966, **34**:888.
64. Hosmer DW, Taber S, Lemeshow S: **The importance of assessing the fit of logistic regression models: a case study.** *Am J Public Health* 1991, **81**:1630–1635.
65. Bartus T: **Estimation of marginal effects using margeff.** *Stata Journal* 2011, **5**:309–329.
66. Cameron CC, Trivedi PK: **Nonlinear regression methods.** In *Microeconomics using Stata*. College Station: Stata Press; 2010:319–362.
67. Stata Corp: *Reference Manual G-M: Release 12*. College Station: Stata Corp LP; 2011.
68. Hardin J, Hilbe J: **Analysis of fit.** In *Generalized linear models and extensions*. Second edition. Edited by Hardin J, Hilbe J. College Station, TX: Stata Press; 2007:47–62.
69. Mihaylova B, Briggs A, O'Hagan A, Thompson SG: **Review of statistical methods for analysing healthcare resources and costs.** *Health Econ* 2010, **20**:897–916.
70. Panageas KS, Schrag D, Riedel E, Bach PB, Begg CB: **The effect of clustering of outcomes on the association of procedure volume and surgical outcomes.** *Ann Intern Med* 2003, **139**:658–665.
71. Kramer AA, Zimmerman JE: **The relationship between hospital and intensive care unit length of stay.** *Crit Care Med* 2011, **39**:1015–1022.
72. Lee AH, Gracey M, Wang K, Yau KKW: **A robustified modeling approach to analyze pediatric length of stay.** *Ann Epidemiol* 2005, **15**:673–677.
73. Leung KM, Elashoff RM, Rees KS, Hasan MM, Legorreta AP: **Hospital- and patient-related characteristics determining maternity length of stay: a hierarchical linear model approach.** *Am J Public Health* 1998, **88**:377–381.
74. Yau KKW, Lee AH, Ng ASK: **Finite mixture regression model with random effects: application to neonatal hospital length of stay.** *Computational Statistics & Data Analysis* 2003, **41**:359–366.
75. Yau KKW, Wang K, Lee AH: **Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros.** *Biom J* 2003, **45**:437–452.
76. Lee AH, Fung WK, Fu B: **Analyzing hospital length of stay - Mean or median regression?** *Med Care* 2003, **41**:681–686.
77. Liu VM, Kipnis PP, Gould MKM, Escobar GJM: **Length of stay predictions: improvements through the use of automated laboratory and comorbidity variables.** *Med Care* 2010, **48**:739–744.
78. Rosenberg AL, Zimmerman JE, Alzola C, Draper EA, Knaus WA: **Intensive care unit length of stay: recent changes and future challenges.** *Crit Care Med* 2000, **28**:3465–3473.
79. Huntley DA, Cho DW, Christman J, Csernansky JG: **Predicting length of stay in an acute psychiatric hospital.** *Psychiatr Serv* 1998, **49**:1049–1053.
80. Marazzi A, Paccaud F, Ruffieux C, Beguin C: **Fitting the distributions of length of stay by parametric models.** *Med Care* 1998, **36**:915–927.
81. Nathanson BH: **Making progress with the egress. Would P.T. Barnum make a good hospital administrator?** *Crit Care Med* 2011, **39**:1208–1209.
82. Moran JL, Bristow P, Solomon PJ, George C, Hart GK: **for the Australian and New Zealand Intensive Care Society Database Management Committee (ADMC): Mortality and length-of-stay outcomes, 1993–2003, in the binational Australian and New Zealand intensive care adult patient database.** *Crit Care Med* 2008, **36**:46–61.
83. Render ML, Kim HM, Deddens J, Sivaganesin S, Welsh DE, Bickel K, Freyberg R, Timmons S, Johnston J, Connors AF Jr, Wagner D, Hofer TP: **Variation in outcomes in veterans affairs intensive care units with a computerized severity measure.** *Crit Care Med* 2005, **33**:930–939.
84. Cai B, Small DS, Have TRT: **Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias.** *Stat Med* 2011, **30**:1809–1824.
85. Terza JV, Basu A, Rathouz PJ: **Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling.** *J Health Econ* 2008, **27**:531–543.
86. Lee AH, Wang K, Yau KKW, Somerford PJ: **Truncated negative binomial mixed regression modelling of ischaemic stroke hospitalizations.** *Stat Med* 2003, **22**:1129–1139.
87. Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP: **Intensive Care Society's APACHE II study in Britain and Ireland-I: variations in case mix of adult admissions to general intensive care units and impact on outcome.** *BMJ* 1993, **307**:972–977.
88. Rabe-Hesketh S, Skrondal A: **Generalized Linear Mixed Models.** In *International Encyclopedia of Education*. Third edition. Edited by Peterson P, Baker E, McGaw B, Barry M. Oxford: Elsevier; 2010:171–177.

doi:10.1186/1471-2288-12-68

Cite this article as: Moran and Solomon: A review of statistical estimators for risk-adjusted length of stay: analysis of the Australian and new Zealand intensive care adult patient data-base, 2008–2009. *BMC Medical Research Methodology* 2012 **12**:68.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

