Modelling survival in acute severe illness: Cox versus accelerated failure time models

John L. Moran MBBS FRACP FJFICM MD,1 Andrew D. Bersten MBBS MD FANZCA FJFICM,2 Patricia J. Solomon BSc PhD,3 Cyrus Edibam MBBS FJFICM,4 Tamara Hunt BN5 and The Australian and New Zealand Intensive Care Society Clinical Trials Group6

1Senior Consultant, Department of Intensive Care Medicine, The Queen Elizabeth Hospital, Woodville, SA, Australia
2Professor, Department of Critical Care Medicine, Flinders Medical Centre, Bedford Park, SA, Australia,
3Associate Professor, School of Mathematical Sciences, University of Adelaide, Adelaide, SA, Australia
4Senior Consultant, Department of Intensive Care Medicine, Royal Perth Hospital, Perth, WA, Australia
5Research Nurse, Department of Critical Care Medicine, Flinders Medical Centre, Bedford Park, SA, Australia
6Australian and New Zealand Intensive Care Society, Carlton, Vic., Australia

Abstract

Background The Cox model has been the mainstay of survival analysis in the critically ill and time-dependent covariates have infrequently been incorporated into survival analysis.

Objectives To model 28-day survival of patients with acute lung injury (ALI) and acute respiratory distress syndrome (ARDS), and compare the utility of Cox and accelerated failure time (AFT) models.

Methods Prospective cohort study of 168 adult patients enrolled at diagnosis of ALI in 21 adult ICUs in three Australian States with measurement of survival time, censored at 28 days. Model performance was assessed as goodness-of-fit [GOF, cross-products of quantiles of risk and time intervals (P ≥ 0.1), Cox model] and explained variation (’R²’, Cox and AFT).

Results Over a 2-month study period (October–November 1999), 168 patients with ALI were identified, with a mean (SD) age of 61.5 (18) years and 30% female. Peak mortality hazard occurred at days 7–8 after onset of ALI/ARDS. In the Cox model, increasing age and female gender, plus interaction, were associated with an increased mortality hazard. Time-varying effects were established for patient severity-of-illness score (decreasing hazard over time) and multiple-organ-dysfunction score (increasing hazard over time). The Cox model was well specified (GOF, P > 0.34) and R² = 0.546, 95% CI: 0.390, 0.781. Both log-normal (R² = 0.451, 95% CI: 0.321, 0.695) and log-logistic (R² 0.470, 95% CI: 0.346, 0.714) AFT models identified the same predictors as the Cox model, but did not demonstrate convincingly superior overall fit.

Conclusions Time dependence of predictors of survival in ALI/ARDS exists and must be appropriately modelled. The Cox model with time-varying covariates remains a flexible model in survival analysis of patients with acute severe illness.

Introduction

The traditional medical survival model has been Cox regression, which has seen many extensions [1]. In the Critical Care environment, patients present with acute severe illness with high attendant mortality and the use of parametric survival models has been suggested as more apposite than the Cox model [2], a sentiment echoed from other domains [3,4]. Acute lung injury (ALI) and the acute respiratory distress syndrome (ARDS) [5] are fulminant respiratory syndromes of diverse aetiology and carry a high short-term mortality; we have recently described [6] the incidence and outcome of a cohort of such patients admitted to adult intensive care units (ICU) in Australia. These patients were followed in each ICU with daily recording of pertinent variables until discharge, death or 28 days following study inclusion. Previous analyses, using both univariate [7] and multivariate approaches [8,9], have identified diverse predictors of patient outcome in ALI and ARDS, related to both the respiratory and non-respiratory systems. The
observations above [6] are now extended to model time to failure (mortality outcome), censored at 28 days after onset of ALI and ARDS. The focus of analysis was twofold: first, a performance comparison of the Cox and parametric survival models, and second, the question of appropriate strategies for handling repeated patient observations over time [10]. Approaches to this issue have been varied: separate models for each day of observation [8], linear regression or analysis of variance [11], trend analysis or weighting of daily scores [12], and Cox model with time-varying covariates [13].

**Methods**

The methods for data collection and quality control have been previously described [6]; briefly, all ICU admissions between October 1 and November 30 1999 to 21 adult ICU in three Australian states were screened daily for the development of ALI according to the 1994 American-European Consensus Conference [5]. The data collected (at screening, inclusion into the study and daily until discharge, death or 28 days following inclusion) included demographics, co-morbidities, ventilator settings (mechanical tidal volume, mL kg\(^{-1}\)), respiratory and haemodynamic variables [for example, arterial oxygen (PaO\(_2\)) and carbon dioxide (PaCO\(_2\)) tension (mmHg), arterial pH; cardiac index (L min \(^{-1}\) and pulmonary artery pressure (mmHg)), lung injury [14] and chest radiographic scores, adjunctive therapies and individual organ and total sequential organ failure score (SOFA score) [15]. Patient severity of illness, as the Acute Physiology and Chronic Health Evaluation (APACHE) II score [16], was recorded only at study enrolment, at the time of development of ALI development of ALI and its more severe form, ARDS.

**Statistical analysis**

Variables are reported as mean (SD) unless otherwise indicated. Interval data were analysed by t-test, and categorical data by Fisher exact test, where appropriate. Stata statistical software (Version 9.2, Stata Corp 2006, College Station, TX, USA) was used. Binary variables, such as gender, were coded as simple indicator variables (0/1).

Patient variable change over time, by final outcome (survivors and non-survivors), was demonstrated graphically using a non-parametric running line scatter-plot smoother with 95% confidence bands (a ‘nearest-neighbour’ local linear (weighted least squares) smoother [17]) as implemented in the Stata module ‘mrunning’ [18]. Time (on the x-axis) was variably reduced from 28 days when data were sparse to achieve acceptable estimates.

Time to mortality-outcome, censored at 28 days post onset of ALI, was analysed initially using Kaplan–Meier and Cox model estimates. Model selection from candidate variables was accomplished by minimization of the Akaike and Bayesian information criteria (AIC and BIC, respectively, penalized model likelihood criteria [19]), and the predictive effect of additional individual variables (with respect to the final model), reported below in the text, was assessed by the likelihood ratio test. Non-linear covariate effect was adjudged by (martingale) residual-by-time analysis and fractional polynomials [20]. Overall, Cox model fit [21] was assessed by (i) approximation of cumulative Cox–Snell residuals to (−log) Kaplan–Meier estimates, residual plots and testing of the proportional hazards assumption and (ii) specific estimation of an ‘added variables’ [21] goodness-of-fit test, a version of the Moreau, O’Quigley and Lellouch test [22] adapted for multi-record patient data [23]. Time-varying covariates were defined as those having interactions of the (continuously time-varying) covariate with failure-times (time to death) over 28 days, the data set (multiple daily records per patient) being initially expanded to risk-sets. The covariates were parameterized as \( \beta \) coefficients (exponentiated as hazard ratios, HRs), constant-within-time, and \( t \) coefficients, time-varying. The interpretation of the dual constant-within-time and time-varying coefficients is: \[ \log \text{ hazard ratio } \text{LHR} = \beta_0 + \beta_1 \times x + \beta_2 \times t, \]

where \( x_1 \) and \( x_2 \) are observed covariates, \( \beta_0 \) are the coefficients and \( t \) is the time, in appropriate scale. The marginal effect (ME) of \( x_2 \) (derivative of LHR with respect to \( x_2 \)) is: \[ \text{ME}(x_2) = \beta_0 + \beta_1 + \beta_2 \times t, \]

where the effect increases or decreases with time according to the sign of \( \beta_2 \). Smoothed baseline hazards were plotted using non-parametric kernel-density estimates as implemented in the Stata module ‘stkerhaz’ [24].

Survival time was also modelled by parametric survival analysis [accelerated failure time (AFT) models], where (log) time (\( t \)) is parameterized thus: \[ \ln(t) = x \beta + \ln(\tau), \]

where \( x \) is the covariate vector and \( \beta \) the corresponding coefficient and the random quantity \( \ln(\tau) \) has a specified distribution (exponential, Weibull, log-normal, log-logistic, gamma) [2,4]. The most intuitive manner in which to express AFT model coefficients is in the exponentiated form, as time ratios (\( TR = t_2/t_1 \)) for a unit increment change in the covariate \( (t_1 = \exp(\beta_0 + \ldots + \beta_n), t_2 = \exp(\beta_0 + \ldots + \beta_n)) \) [25]. Thus, \( TR < 1 \) are associated with a decrease in (survival) time and \( TR > 1 \) are associated with a prolonged (survival) time, or, more accurately, a contraction or expansion of time to failure. Adequacy of the AFT models for the data was initially gauged by plotting log(time) against a linear function of the cumulative (Nelson–Aalen) hazard rate: for the log normal, \( \Phi^{-1}[1 - \exp(-H)] \), where \( H \) is the hazard and \( \Phi \) is the standard normal cumulative distribution function, and for the log logistic, \( \ln[\exp(\Phi^{-1}[1 - \exp(-H)])] \) [26]. Competing models (with respect to distribution) were adjudged by approximation of cumulative Cox–Snell residuals to (−log) Kaplan–Meier estimates and minimization of the AIC.

The explained variation [27] of the survival models (adjusted and unadjusted, point estimate and 95% CI over 5000 bootstrap [28] repetitions of the whole data set, with clustering on patient), an analogue of the ‘\( R^2 \)’ measure in linear models, was calculated via the ‘str2d’ Stata module [29] using the method of Royston and Sauerbrei [30]. This measure was used heuristically given that (i) the inferential basis for comparison between non-nested models is lacking and (ii) the extension to time-dependence of \( \beta \) (the covariate and parameter vector) has not been formally established [29].

Heterogeneity of ICU site effects was estimated using frailty models for both the Cox and AFT models; that is, models incorporating latent random effects, in this case gamma distributed with a unit mean and variance estimated from the data and having a multiplicative effect on the hazard [1,21]. Frailty addresses unexplained variability (in time to failure) in terms of omitted covariate(s) or measurement error; thus, if frailty is ignored, an underestimation of covariate effect will be observed [31].
Results

Overview

Over the 2-month time period of the study, 168 of 1977 patients developed ALI. The cohort had a mean (SD) age of 61.5 (18.4) years (30% female). At study enrolment, the APACHE II and SOFA scores were 20 (9) and 8.7 (3.0), respectively. Pertinent respiratory variables were: PaO$_2$/FIO$_2$ ratio 176 (76), PaO$_2$ 103 (49) mmHg, PaCO$_2$ 43 (11) mmHg, arterial pH 7.34 (0.19) and lung injury score 2.0 (0.7). Time to enrolment from ICU admission was 1 day (median, range 0–13 days).

Variable change over time

Time change of key variables, by outcome, over 28 days is shown for (i) PaO$_2$, PaCO$_2$, PaO$_2$/FIO$_2$ ratio, pH, lung injury score and SOFA score in Fig. 1 and (ii) peak and plateau airway pressures, extrinsic positive airway pressure, cardiac index and mean pulmonary artery pressure and mechanical tidal volume in Fig. 2. For variables shown in Fig. 1, separation by final outcome over time was demonstrated for: (i) the SOFA score, (ii) arterial pH from enrolment to day 5 and (iii) PaCO$_2$ from day 9 onwards. For the cardiac and mechanical ventilatory variables (Fig. 2) the 95% CI were wider, due to sparse data, but of note, over time: (i) the cardiac index, as expected, increased in survivors and decreased in non-survivors, (ii) mean pulmonary artery pressure declined, somewhat surprisingly, in both outcome groups, (iii) plateau pressures were consistently below 30 cm H$_2$O in both survivors and non-survivors and (iv) paradoxically, mechanical tidal volume (per predicted kg$^{32}$) tended to decrease in survivors but increased in non-survivors.

Survival over time

Kaplan–Meier estimates, with 95% CI and patient number at risk over 28 days, and covariate-unadjusted smoothed hazard are shown in Fig. 3. Peak hazard was apparent at days 7–8 after onset of ALI/ARDS.

Cox model

Cox model predictors were: age, gender, enrolment APACHE II score and PaO$_2$/FIO$_2$ ratio and SOFA score (both modelled using

Figure 1 Time change of patient respiratory and organ system variables over 28 days. Change over time (days) post onset of acute respiratory distress syndrome (ARDS) of patient variables, by outcome: running line smooth (solid lines) with 95% confidence bands (dashed lines).
Modelling survival in acute severe illness

Figure 2  Time change of cardiac and mechanical ventilatory variables over 28 days. Change over time (days) post onset of acute respiratory distress syndrome (ARDS) of patient variables, by outcome: running line smooth (solid lines) with 95% confidence bands (dashed lines).

Figure 3  Survival over 28 days post onset of acute respiratory distress syndrome (ARDS): Kaplan–Meier estimates and smoothed hazard. Left panel: time to failure (28 day mortality outcome) of ARDS cohort: Kaplan–Meier estimates: point estimates (solid line with patients at risk beneath) and 95% CI (dashed lines). Right panel: unadjusted smoothed mortality hazard (solid line) with 95% CI (dashed lines).
and a significant interaction between age and gender was also noted. For the Cox model, goodness-of-fit was acceptable ($P > 0.34$) across six cross-products of quantiles of risk and time intervals. Predictor multi-collinearity and non-linear covariate effect, including a quadratic effect of age, were not evident. Of interest, aetiology of lung injury (direct vs indirect lung injury and the subsets within these two broad categories), time to enrolment from ICU admission, source (emergency service, ward location or postoperative) and prime cardiovascular variables, in particular mean pulmonary artery pressure, were not predictive. Plots of the Cox model parameter estimates (that is, the $\beta$ coefficients, with 95% CI) over time [33] are seen in Fig. 4; most marked time change of parameter estimate occurred with enrolment APACHE II score and SOFA score. The final Cox model included time-varying covariates for the SOFA score and PaO$_2$/FIO$_2$ ratio (estimated on day-by-day values) and the enrolment APACHE II score, as seen in Table 1.

### Parametric survival models

With respect to adequacy of the model to the data [approximation of the linear function of the hazard to log(time)], and by residual analysis, the best fitting parametric model appeared to be the log-normal, which was parameterized in the AFT metric, although the log-logistic also showed satisfactory performance. The same time invariant variables were identified as in the Cox model, but only APACHE II score demonstrated a time-varying component (Table 1).

### Survival and hazard plots

Patient survival curves for men and women, for the Cox (log-normal) and Cox regression models, are seen in the top two panels of Fig. 5; the curves were adjusted for PaO$_2$/FIO$_2$ ratios of 107, 146, 200, 262 and 300 (corresponding to the 10th, 25th, 50th, 75th and 90th percentiles of PaO$_2$/FIO$_2$ ratio) and mean values of age (63 years), APACHE II score (21) and SOFA score (7). There was good survival probability for PaO$_2$/FIO$_2$ ratios of 300, 262 and 200, being at least 85% at 28 days. Survival estimates for the log-normal AFT models were worse than those for the corresponding Cox model. The modest effect of the interaction between age and gender is seen in the bottom panels (for the Cox model) for the above percentiles of age, at 35, 52, 63, 76 and 80 years, respectively, with mean values of APACHE II and SOFA scores and PaO$_2$/FIO$_2$ ratio at 146. The survival curves of women (left panel) with the age interaction are displaced upwards (‘better’ survival) and those of men with the age interaction are displaced downwards (‘worse’ survival), with respect to the line of effect for PaO$_2$/FIO$_2$ ratio at 146 (medium-dash line in the middle panels).

This can be understood in that, without the age interaction, the HR for men was 0.397 (95% CI: 0.178–0.885, $P = 0.02$) with respect to women, the HR for the gender–age interaction was 0.986 (Table 1) and the overall HR for the linear combination (age + gender-age) was 1.018. That is, as age increased across gender (men vs women), survival worsened.

Baseline hazard plots for the 5 percentiles of PaO$_2$/FIO$_2$ ratio, men and women, for the Cox (top panels) and AFT (log-normal and log-logistic, middle and bottom panels respectively) models are seen in Fig. 6. The Cox model showed uniformly declining hazard from variably initial high values. Both the AFT distribution models demonstrated progressive increase in hazard over time with a minimal to modest decline at higher hazard levels after 8–12 days for log-normal and 12–16 days for log-logistic.

### Model comparison

On comparison of the Cox and AFT survival models: none of the hazard profiles of the models (Fig. 6) reflected the unadjusted non-parametric estimate which peaked at 6–7 days and declined thereafter (Fig. 3, right panel); residual analysis in both the Cox and AFT models demonstrated a tendency to over-predict at short survival times (<10 days) and under-predict beyond this time period; stratifying by gender, all models showed relatively poor approximation to Kaplan–Meier estimates, although such may be generally expected, as Kaplan–Meier are marginal estimates and the Cox/AFT are conditional (on specified covariates); estimates of global fit, using cumulative Cox–Snell residuals, suggested some superiority of the log normal AFT model.

Explained variation (Table 1) was relatively good for all models, although the 95% CI were wide. Point estimates for $R^2$ of the Cox model were consistently larger than those of the AFT models, both for adjusted and unadjusted ‘$R^2$’; and when the $\beta$ parameter dimension of the models was equalized by ‘forcing’ the time-varying SOFA and PaO$_2$/FIO$_2$ covariates into the AFT models. Superiority of either AFT model for explained variation was not demonstrated as 95% CI for the difference spanned zero (Table 1: Delta: 95% CI for log-logistic–log-normal).

---

**Figure 4** Time change of parameter estimates (as hazard ratio, HR), Cox model. Panels demonstrate change in Cox regression parameter estimates (that is, the $\beta$ coefficients, exponentiated as HR ± 95% CI) of variables over 28-day period. Vertical axis: Parameter estimates ($\beta$ coefficient); Horizontal line = HR of 1. Horizontal axis: Survival time (days): post ARDS diagnosis.
### Table 1 Cox and AFT model specifications

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Cox</th>
<th>AFT log-normal</th>
<th>AFT log-logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>CI_L</td>
<td>CI_U</td>
</tr>
<tr>
<td>Time-invariant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOFA score</td>
<td>0.854</td>
<td>0.724</td>
<td>1.007</td>
</tr>
<tr>
<td>APACHE II score</td>
<td>1.075</td>
<td>1.003</td>
<td>1.153</td>
</tr>
<tr>
<td>PaO&lt;sub&gt;2&lt;/sub&gt;/FIO&lt;sub&gt;2&lt;/sub&gt; ratio</td>
<td>0.988</td>
<td>0.98</td>
<td>0.995</td>
</tr>
<tr>
<td>Age</td>
<td>1.003</td>
<td>1.009</td>
<td>1.057</td>
</tr>
<tr>
<td>Age-gender interaction</td>
<td>0.986</td>
<td>0.974</td>
<td>0.998</td>
</tr>
<tr>
<td>Time-varying</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APACHE II score</td>
<td>0.995</td>
<td>0.991</td>
<td>1</td>
</tr>
<tr>
<td>PaO&lt;sub&gt;2&lt;/sub&gt;/FIO&lt;sub&gt;2&lt;/sub&gt; ratio</td>
<td>1.001</td>
<td>1</td>
<td>1.002</td>
</tr>
<tr>
<td>SOFA score</td>
<td>1.036</td>
<td>1.021</td>
<td>1.052</td>
</tr>
<tr>
<td>AIC</td>
<td>178.290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>221.710</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&lt;sup&gt;2&lt;/sup&gt; (95% CI)</td>
<td>0.546 (0.390, 0.781)</td>
<td>0.451 (0.321, 0.695)</td>
<td>0.470 (0.346, 0.714)</td>
</tr>
<tr>
<td>R&lt;sup&gt;2&lt;/sup&gt; (95% CI) adjusted</td>
<td>0.400 (0.188, 0.684)</td>
<td>0.295 (0.06, 0.645)</td>
<td>0.384 (0.24, 0.644)</td>
</tr>
<tr>
<td>Delta: 95% CI for log-logistic–log-normal</td>
<td>R&lt;sup&gt;2&lt;/sup&gt; adjusted</td>
<td>−0.251, 0.289</td>
<td></td>
</tr>
<tr>
<td>Equivalent β dimension</td>
<td>AIC</td>
<td>174.020</td>
<td>175.670</td>
</tr>
<tr>
<td>BIC</td>
<td>228.300</td>
<td>229.850</td>
<td></td>
</tr>
<tr>
<td>R&lt;sup&gt;2&lt;/sup&gt; (95% CI)</td>
<td>0.546 (0.390, 0.781)</td>
<td>0.487 (0.401, 0.715)</td>
<td>0.490 (0.403, 0.748)</td>
</tr>
</tbody>
</table>

HR, hazard ratio estimate from Cox model. CI<sub>U</sub>, 95% upper confidence interval for parameter. CI<sub>U</sub>, 95% lower confidence interval for parameter. P<sub>HR</sub>, P-value for hazard or time ratio. TR, time ratio derived from log-normal accelerated failure time model, with 95% lower (CI<sub>L</sub>) and upper (CI<sub>U</sub>) confidence limits. P<sub>TR</sub>, P-value for TR. TR < 1 are associated with a decrease in (survival) time and TR > 1 are associated with a prolonged (survival) time.

AIC, Akaike information criterion (−2L + 2(c + p + 1)), where L is the log-likelihood, c is the number of model covariates and p is the number of model-specific ancillary parameters. BIC, Bayesian information criterion (−2L + log(n) × (c + p + 1)), where L is the log-likelihood, c is the number of model covariates and p is the number of model-specific ancillary parameters and n is the number of observations). R<sub>2</sub>, explained variation (R<sup>2</sup>) [29]. R<sub>2</sub> adjusted, dimension of β subtracted from R<sub>2</sub>. Equivalent β dimension, AFT models computed with same number of covariates as Cox model.
No evidence of significant ICU site heterogeneity ($P \geq 0.5$) was demonstrable for any of the survival models.

**Discussion**

**Appropriate models**

The approach adopted in the current paper was to develop a parsimonious overall model which would both ascertain predictor variables and encompass repeated covariate measurement, but, at the same time, avoid an *ad hoc* multiple test approach to establishing the covariate framework [34]. The Cox model displayed non-proportionality and to overcome this, time-varying covariates were incorporated into the analysis. The best fitting AFT model was the log-normal which also demonstrated a time-varying covariate effect [35,36], but only for the enrolment APACHE score. Such ‘differences’ in the number of significant variables between non-nested multivariable models has been previously noted [37]. The Cox model expresses the multiplicative effect of covariates on the hazard (force of mortality) via conditioning on failure times and the parametric log-normal model, the multiplicative effect of covariates on time to failure, by expanding or contracting it. To this extent the AFT model may have explanatory advantage in that covariates have a direct effect on survival time (‘a direct physical
interpretation’ [38]), not on a conditional probability [39]. Comparison of the coefficients from both the Cox and AFT models and the implied directional change in the hazard and (survival) time (Table 1), serves to establish a comparability of interpretation. This may be more formally expressed in terms of a general regression model:

$$\lambda(t|Z) = \lambda_0(t \exp(\beta_1 Z))\exp(\beta_2 Z),$$

where $\lambda(t|Z)$ is the hazard function at time $t$, $Z$ is the p-vector covariate and $\lambda_0$ is the baseline hazard function. The model becomes proportional hazards if $\beta_1 = 0$, AFT when $\beta_1 = \beta_2$ and accelerated hazards (scale change in the hazard function without altering the overall shape of the baseline hazard function) when $\beta_2 = 0$ [40]. Thus for age, the Cox model showed an increase in HR per unit (year) increase in age and the AFT model a decrease of survival time per unit increase of age (TR < 1, deceleration of the ME of age). Other covariate effects must be understood in the context of the dual coefficients ($rh$ and $t$) expressing the time-varying change.

The AFT model did not demonstrate a convincingly superior fit compared with the Cox model. This was in contrast to Knaus et al. [2,41] who investigated the performance of Cox versus AFT models on survival to 28 days in sepsis patients using similar analytic methods, although importantly, they did not utilize time-varying covariates (an ‘invariably difficult’ concept [42]). In a retrospective study of 1195 septic patients, Knaus et al. [2] assessed the overall utility of the model(s) for predicting the 28 day mortality (at 36%) and the best fit of the time distribution of death. The observed progressive decrease over 28 days of the (mortality) impact of the initial APACHE III acute physiology score violated proportional hazards assumptions and a log-normal AFT model provided the best fit. In a second study of patients enrolled into a phase III sepsis trial [41], the log-normal AFT model was again found to be optimal and the (mortality) hazard was noted to be maximal between trial enrolment (day 0) and day 2. The implica-

Figure 6 Smoothed hazards for Cox and accelerated failure time (AFT) models adjusted for (n = 5) percentiles of PaO$_2$/FIO$_2$ and mean values of other significant covariates. Long dash line: PaO$_2$/FIO$_2$ ratio 300, SOFA score 6.9, age 63 years, APACHE II score 21. Long dash-dot line: PaO$_2$/FIO$_2$ ratio 262, SOFA score 6.9, age 63 years, APACHE II score 21. Short dash line: PaO$_2$/FIO$_2$ ratio 200, SOFA score 6.9, age 63 years, APACHE II score 21. Dashed line: PaO$_2$/FIO$_2$ ratio 146, SOFA score 6.9, age 63 years, APACHE II score 21. Solid line: PaO$_2$/FIO$_2$ ratio 107, SOFA score 6.9, age 63 years, APACHE II score 21.
tions of these and the current findings are (i) the proportional hazards assumption of the Cox model requires that covariate effect be constant on the relative hazard scale, which may be unrealistic; (ii) in pathophysiological states where initial (mortality) event rates are high, AFT models may possess advantage, as recently reported in non-ICU settings [43]; and (iii) the choice of a statistical model cannot be indifferent to the underlying biological process.

**Predictor variables**

The continuous variables were modelled linearly, there being no convincing evidence of non-linearity in effect. An alternative, ‘traditional’ approach [8] would have been that of ‘cut-point’ variable analysis, usually at the median or quartile values of the particular variable. Little justification is given in the literature for particular cut-points and appropriate P-value adjustment for ‘optimal’ cut-points is rarely undertaken [44]. Moreover, from the view-point of pathophysiology, it would appear unrealistic to suggest that covariate effects change substantially at empiric cut-points. Analytic consequences, which have bearing upon our understanding, also flow from such cut-point analysis: inflated estimates of effect and increased variance of these estimates, increased Type I error and decreased efficiency of analysis [45]. Of interest, using such a cut-point approach, an ‘optimal’ (for effect) PaO\textsubscript{2}/FIO\textsubscript{2} ratio was found at 87 (adjusted P-value, 0.02, [46]). In the full model, the HR associated with PaO\textsubscript{2}/FIO\textsubscript{2} ratio <87 was 4.77 (P = 0.001), with no time-varying component, but there was no overall advantage for the model incorporating this categorical variable versus PaO\textsubscript{2}/FIO\textsubscript{2} ratio (considered as a continuous variable), as assessed by AIC. Similarly, no advantage was found in the full model for a categorical variable partitioning the PaO\textsubscript{2}/FIO\textsubscript{2} ratio <200 and >200 (corresponding to the ALI/ALI with ARDS clinical division, P = 0.51).

The linear modelling approach had advantage, interpretation wise, with respect to the time-varying coefficients. In a Cox model without time-varying coefficients, the HR computed for the enrolment APACHE II score and SOFA score were 1.02 and 1.13, respectively, indicating an increase in HR with unit increase of either score. With the time-varying nature of effect taken into account (Table 1), the hazard for enrolment APACHE II score, not unreasonably, decreased in intensity over time (similar to that found for the APACHE III acute physiology score by Knaus et al. [2]) and that for the SOFA score progressively increased over time (Fig. 4). The effect of PaO\textsubscript{2}/FIO\textsubscript{2} ratio on ALI/ARDS outcome in the literature has been contradictory [47], depending upon definitions, analytic cut-off values and time-course. The HR for the PaO\textsubscript{2}/FIO\textsubscript{2} ratio in a Cox model without time-varying coefficients was 0.994, suggesting a progressive decrease in hazard per unit increase in PaO\textsubscript{2}/FIO\textsubscript{2} ratio. Again, such an effect was seen to be modified when the time-varying coefficient was computed, and the hazard was not demonstrated to decrease over time (time-varying coefficient = 1.001, see Table 1). Thus ‘recovery’ of lung function over time, as indicated by an improvement in PaO\textsubscript{2}/FIO\textsubscript{2} ratio, belies the effect of other concomitant processes which militate against eventual recovery.

Multiple non-pulmonary organ system dysfunction, however defined, occurring either before ICU admission [48–50] or subsequent to the development of ALI/ARDS [7,51,52] has been associ-
is no (formal) coincidence of the concepts of statistical significance of variables, adequacy of model fit and predictive power (explained variation): ‘The fact that a model is correctly specified for a set of prognostic variables . . . implies that on average predictions will be correct . . . ’ (original emphasis) [27].

One particular problem of comparison of non-nested models is the inability to compare, say, AIC values, of the Cox and log-normal models due to the different estimation of the two models (partial versus maximum likelihood respectively), a difference which would appear to vitiate the comparisons of Chapman et al. who ignored such differences in computing the ‘relative likelihood’ (ratio of Cox versus AFT likelihoods) as a model selection measure [57]. The divergence between AIC and BIC as model selection measures is also illustrated by the change in these indices when the β dimension of the three models was equalized (Table 1, Equivalent β dimension) – AIC decreased, with the expected increment of \( R^2 \), but BIC increased, corresponding to the well-described aims of these two measures – expected prediction of new data (AIC) and identification of the models with the highest probabilities of being the ‘true’ model for the data (BIC) [19].

Given the modest size of the data set and the relatively small number of events, ‘validation’ of the models by, for instance, split sample techniques was impracticable. This question was addressed by an intensive (5000 repetitions) bootstrap of the explained variation measure. The 95% CIs of \( R^2 \) of all models were relatively wide, but the lower limits were quite acceptable, given the known relatively low scalar values of \( R^2 \) in non-linear models compared with linear [58].

With AFT models, in the presence of time-varying covariates, the prediction and interpretation of time is a nontrivial exercise [25]. Similarly, the interpretation of both the underlying hazard [59], unspecified in the Cox model, and covariate treatment effects [13] are problematic under time-varying covariates. As pointed out by Kalbfleisch and Prentice, such ‘internal’ covariates require the survival of the patient for their existence and only determine failure ‘information’ for time less than or equal to their instantaneous presence. Thus the conditional analysis gives information on instantaneous failure rates and the hazard is not directly related to the overall survivor function [13,60].

Conclusions

Survival in acute severe illness (ALI/ARDS) was dependent upon patient characteristics of age and gender, and underlying severity of illness as reflected by multiple organ dysfunction, which displayed a time dependence which must be accounted for in analysis [38]. Peak mortality hazard occurred at days 7–8 after ALI/ARDS onset, suggesting a therapeutic window of opportunity. The Cox model with time-varying covariates exhibited considerable flexibility and parametric survival models were not demonstrated to be analytically superior.

References


