Development and validation of a dynamic outcome prediction model for paracetamol-induced acute liver failure: a cohort study

William Bernal, Yanzhong Wang*, James Maggs, Christopher Willars, Elizabeth Sizer, Georg Auzinger, Nicholas Murphy, Damian Harding, Ahmed Elsharkawy, Kenneth Simpson, Fin Stolze Larsen, Nigel Heaton, John O'Grady, Roger Williams, Julia Wendon*

Summary

Background Early, accurate prediction of survival is central to management of patients with paracetamol-induced acute liver failure to identify those needing emergency liver transplantation. Current prognostic tools are confounded by recent improvements in outcome independent of emergency liver transplantation, and constrained by static binary outcome prediction. We aimed to develop a simple prognostic tool to reflect current outcomes and generate a dynamic updated estimation of risk of death.

Methods Patients with paracetamol-induced acute liver failure managed at intensive care units in the UK (London, Birmingham, and Edinburgh) and Denmark (Copenhagen) were studied. We developed prognostic models, excluding patients who underwent transplantation, using Cox proportional hazards in a derivation dataset, and tested in initial and recent external validation datasets. Mortality was estimated in patients who had emergency liver transplantation. Model discrimination was assessed using area under receiver operating characteristic curve (AUROC) and calibration by root mean square error (RMSE). Admission (day 1) variables of age, Glasgow coma scale, arterial pH and lactate, creatinine, international normalised ratio (INR), and cardiovascular failure were used to derive an initial predictive model, with a second (day 2) model including additional changes in INR and lactate.

Findings We developed and validated new high-performance statistical models to support decision making in patients with paracetamol-induced acute liver failure. Applied to the derivation dataset (n=350), the AUROC for 30-day survival was 0·92 (95% CI 0·88–0·96) using the day 1 model and 0·93 (0·88–0·97) using the day 2 model. In the initial validation dataset (n=150), the AUROC for 30-day survival was 0·89 (0·84–0·95) using the day 1 model and 0·90 (0·85–0·95) using the day 2 model. Assessment of calibration using RMSE in prediction of 30-day survival gave values of 0·1642 for the day 1 model and 0·0626 for the day 2 model. In the external validation dataset (n=412), the AUROC for 30-day survival was 0·91 (0·87–0·94) using the day 1 model and 0·91 (0·88–0·95) using the day 2 model, and assessment of calibration using RMSE gave values of 0·079 for the day 1 model and 0·107 for the day 2 model. Applied to patients who underwent emergency liver transplantation (n=116), median predicted 30-day survival was 51% (95% CI 33–85).

Interpretation The models developed here show very good discrimination and calibration, confirmed in independent datasets, and suggest that many patients undergoing transplantation based on existing criteria might have survived with medical management alone. The role and indications for emergency liver transplantation in paracetamol-induced acute liver failure require re-evaluation.

Funding Foundation for Liver Research.

Lancet Gastroenterol Hepatol 2016

Published Online
July 12, 2016
http://dx.doi.org/10.1016/S2468-1253(16)30007-3

See Online/Comment
http://dx.doi.org/10.1016/S2468-1253(16)30021-8

*Authors contributed equally

Institute of Liver Studies, King's College Hospital, London, UK (W Bernal MD, J Maggs MBBS, C Willars FRCA, E Sizer FRCA, G Auzinger MD, Prof N Heaton FRCS, Prof J O'Grady MD, Prof J Wendon MD); **Division of Health and Social Care Research, King's College London, London, UK** (Y Wang PhD); **Departments of Liver Medicine, Anaesthesia and Critical Care, Queen Elizabeth Hospital, Birmingham, UK** (N Murphy FRCA, D Harding FRCAP, A Elsharkawy PhD); **Department of Hepatology, Edinburgh Royal Infirmary, Edinburgh, UK** (K Simpson MD); **Department of Hepatology, Rigshospitalet University Hospital Copenhagen, Copenhagen, Denmark** (F S Larsen MD); **and Foundation for Liver Research, London, UK** (Prof R Williams MD)

Correspondence to:
Dr William Bernal, Liver Intensive Therapy Unit, Institute of Liver Studies, King's College Hospital, London SE5 9RS, UK
william.bernal@kcl.ac.uk

Introduction

Acute liver failure is a rare critical illness, and in many countries, including the UK, paracetamol-induced hepatotoxicity is its most frequent cause.[1,2] The illness may follow a rapidly progressive course with severe hepatic necrosis quickly followed by the development of encephalopathy, multiple extrahepatic organ failures, and death.[3] However, the condition is one in which there is substantial variation in clinical course. In some patients, recovery with medical therapy alone might be possible despite severe multiple organ failure, whereas in others survival might be impossible without emergency liver transplantation.[4,5] The early and accurate evaluation of expected prognosis is key to effective management, to enable successful transplantation in the narrow window of opportunity for those who will benefit from it, and to avoid unnecessary surgery in those who will survive with medical therapy alone. Tools for prognostic evaluation are available, but their limitations are increasingly apparent.[6–8]

Prognosis in paracetamol-induced acute liver failure is most commonly assessed using the King's College Criteria. The criteria were derived from the analysis of patients managed in a single centre between 1973 and 1987 and have been in use to select transplant candidates for more than two decades.[6] Poor prognosis criteria include an arterial pH of less than 7·3 after volume resuscitation, or the combined findings of high-grade

Research in context

Evidence before this study

We searched PubMed and Medline with no date restrictions on June 1, 2015, with the terms "acute liver failure" and "fulminant hepatic failure" to identify studies and publications in addition to those familiar to the authors or cited by guidelines. Reference lists in identified reports were also reviewed to further identify studies of relevance. We identified three meta-analyses of prognostic criteria in acute liver failure, but most studies reporting survival outcomes were case series from single centres with small cohorts and open to reference bias. Available evidence suggests that survival of patients with paracetamol-induced acute liver failure managed with medical care alone has progressively increased over time, and that existing criteria used to select candidates for emergency liver transplantation might no longer be effective.

Added value of this study

This study presents novel criteria for the assessment of prognosis in paracetamol-induced acute liver failure, derived and validated in a large multicentre patient cohort. The models developed reflect the current outcomes of the illness and use readily available clinical variables and standardised definitions, with sequential assessment to assess changing prognosis over time. For the first time, the models enable individualised and updated survival predictions to be made and address key practical issues in relation to clinical use.

Implications of all the available evidence

Emergency liver transplantation might be life saving in selected patients with paracetamol-induced acute liver failure; however, many patients who undergo transplantation based on existing criteria might have survived with medical management alone. The use of emergency liver transplantation for this indication requires comprehensive re-evaluation.

encephalopathy, creatinine of more than 300 μmol/L and international normalised ratio (INR) of 6.5 or higher occurring within a narrow timeframe.[6] Early meta-analysis of the diagnostic performance of the King's College Criteria in paracetamol-induced liver failure suggested high specificity but more limited sensitivity; introduction of arterial blood lactate concentration as a supplemental marker was proposed to address this issue.[7–9] However, the increasing success of non-transplantation medical therapies alone is likely to have affected the performance of the King's College Criteria in patients with paracetamol-induced liver injury because substantial improvements in survival with medical care alone have been made for many causes of acute liver failure in the past three decades, particularly cases resulting from paracetamol-induced hepatotoxity.[3] These improvements in non-transplant outcomes have not been reflected in changes in recognised indications for emergency liver transplantation.

Experience has also shown practical issues in the use of the King's College Criteria. The criteria were intended for use in a transplant centre and not early after presentation in the emergency room where intravenous fluid resuscitation has not been undertaken and the effect of high blood levels of paracetamol might contribute to a reversible lactic acidosis.[10,11] The criteria perform less accurately in accidental or intentional staggered overdoses of the drug than after single timepoint deliberate ingestions, and are difficult to interpret because their component variables might be confounded by alteration by medical interventions.[12] Since the criteria provide a binary outcome prediction rather than a continuum of risk, their application can also be difficult, and they do not address a key clinical question in wait listed patients who show signs of improvement—is it better to remove them from the list or to proceed with transplantation?

In exploring the development of a novel model for the prediction of death without transplantation in patients with paracetamol-induced acute liver failure, we sought to address the limitations of the present King's College Criteria. By studying more recent cohorts we hoped to reflect the current outcomes of the illness and develop a model using readily available clinical variables and standardised definitions, with sequential assessment to detect changing prognosis over time.[13] Our objective was to develop a decision support tool that was simple to use and gave a continuous and updated estimation of risk of death rather than a static binary outcome prediction. We hoped to avoid the reference bias that might complicate studies of prognosis in acute liver failure by excluding patients who underwent transplantation, and to assess reproducibility and transportability by the use of several validation cohorts.

Methods

Study design and participants

In this study, we included patients with severe paracetamol-induced hepatotoxicity managed at four specialist liver transplantation intensive care units in the UK and Denmark. In all cases, a history of drug ingestion or detectable blood paracetamol was present, with the exclusion of other causes of acute liver injury. Other inclusion criteria were an INR of 1·5 or higher and absence of a history and clinical or radiological findings of previous liver disease.

The primary derivation and initial internal validation test cohorts were derived from consecutive patients with severe paracetamol-induced hepatotoxicity who had not undergone transplantation and who were admitted

 Published online July 12, 2016 http://dx.doi.org/10.1016/S2468-1253(16)30007-3

during the period 2000–12 to the Liver Intensive Therapy Unit at King's College Hospital, London, UK. The external validation cohorts included patients who did not undergo transplantation admitted to the Rikshopitalet Liver Unit, Copenhagen, Denmark in 2011–13; the Scottish Liver Transplant Unit at Edinburgh Royal Infirmary, Edinburgh, UK in 2008–14; the intensive care unit at Queen Elizabeth Hospital, Birmingham, UK, in 2004–13; and the Liver Intensive Therapy Unit at King's College Hospital in 2012–14. Validating mortality in paracetamol-induced acute liver failure is rarely an issue since almost without exception death in patients who have not undergone transplantation occurs during a single hospital admission, most commonly during the first week in the intensive care unit, with rapid recovery seen in those who survive.[14]

A common approach to clinical management was applied in all units, with emergency liver transplantation considered in patients who fulfilled the standard King's College Criteria. Standard medical care applied has been detailed elsewhere.[3] Briefly, patients who developed encephalopathy at grade 3 or higher were intubated, sedated, and mechanically ventilated. Guided restoration of circulating volume was commenced immediately on admission and used invasive haemodynamic monitoring. Coagulopathy was not supported unless active haemorrhage was present.[15] Norepinephrine was the primary vasopressor used and dobutamine the primary ionotropic agent with adjunctive use of intravenous low-dose hydrocortisone and vasopressin. Renal replacement therapy used continuous veno-venous haemofiltration. Indications for use of renal replacement therapy included not only those standard for patients with acute kidney injury with anuria but also for relative oliguria, metabolic stabilisation, and control of acidosis and hyperammonaemia. Sedation was achieved with fentanyl and propofol infusions, with rare use of paralysis. Treatment for intracranial pressure crises was with bolus intravenous mannitol, hypertonic saline, and increased sedation using thiopentone in refractory cases. Intravenous N-acetyl cysteine was administered to all patients with an infusion of 100 mg/kg every 24 h for a maximum of 5 days or until the INR was lower than 2.

Datasets and statistical methods

The derivation and initial validation patient datasets were taken from the King's College Hospital Liver Intensive Therapy Unit clinical database in which demographic, physiological, and laboratory variables are prospectively collected daily for all patients by specialist audit nurses. These variables included those required for the Sequential Organ Failure Assessment (SOFA) and Acute Physiology and Chronic Health Evaluation scores (appendix p 1). Fewer than 1% of cases had missing values.

We used standardised scores assessing level of consciousness and grade of encephalopathy using the Glasgow coma scale, and cardiovascular dysfunction by the SOFA cardiovascular component score (SOFA CVS; appendix p 2), with a score of 3 or more considered to represent cardiovascular failure.[3,16,17] The Glasgow coma scale score was assessed in patients who had not received sedative agents, and we used the lowest score before their administration, except if signs of intracranial hypertension were present, where a score of 3 was assigned.[3,18] Survival time was from date of admission to date of death. Continuous variables are summarised as mean (SD) or median (IQR) and categorical data as count (percentage). Student's *t* test, Mann-Whitney *U*, and Wilcoxon signed-rank tests were used to test differences in continuous variables where appropriate, and the χ^2 test used for proportions. Multiple survival analyses were undertaken using Cox proportional hazards models to determine the prognostic and predictive value of demographic factors and clinical variables. All-cause mortality was the primary event studied and patients who underwent transplantation were excluded from model development and primary testing. Variable selection and model fitting were conducted through backward stepwise regression based on p values. The model started with a range of clinical variables of important prognostic value suggested by literature review (appendix p 1), and then went through extensive backward stepwise model and variable selection process. The final models consisted of variables most strongly associated with death in the multiple Cox proportional hazards regressions. The proportional hazards assumption for each covariate was tested using the scaled Schoenfeld residuals. The proportionality test showed that all the covariates followed the proportional hazards assumption ($p>0\cdot05$).

To develop and validate the predictive survival models, we divided the King's College Hospital derivation set (n=350) and an initial validation set (n=150) with random case selection. Survival models using day 1 and day 2 were first built from derivation dataset separately and then validated in the validation dataset. The performance and predictive accuracy of the models were assessed using receiver operating characteristic (ROC) analysis. The ROC curves and values were compared among different models and also between derivation and validation stages. Model calibration was assessed comparing observed and predicted survival using root mean square error (RMSE).[19] All tests were two-tailed, and p values lower than $0\cdot05$ were considered statistically significant. ROC values with 95% CIs for all the survival models and hazard ratios (HR) with 95% CIs for the chosen clinical variables were calculated. Statistical analyses were done with statistical software R, version 2.11.1, and SPSS, version 22.0.0. All data were fully de-identified before exchange and analysis, and their use was approved by the Research Ethics Committee of King's College Hospital.

See Online for appendix

Role of the funding source

This study was supported by an unrestricted grant from the Foundation for Liver Research. The funding source had no role in study design, collection, analysis, or

interpretation of the data, or in the writing of the report. Authors had access to the raw data from their individual centres and WB and YW to all de-identified datasets. The corresponding author had final responsibility for the decision to submit for publication.

	Derivation cohort (n=350)	Validation cohort (n=150)
Age (years)	37 (27–48)	36 (26–47)
Sex		
Female	188 (54%)	94 (63%)
Male	162 (46%)	56 (37%)
Died	78 (22%)	35 (23%)
Day 1		
Hepatic encephalopathy grade ≥3	149 (43%)	64 (43%)
Glasgow coma scale score	13 (8–15)	14 (8–15)
Cardiovascular failure*	83 (24%)	35 (23%)
Mean arterial pressure (mm Hg)	74 (63–95)	70 (61–90)
INR	4·0 (2·6–6·2)	3·8 (2·8–6·0)
Bilirubin (μmol/L)	68 (44–97)	72 (44–104)
AST (IU/L)	5776 (2601–9809)	5215 (2089–9315)
Creatinine (μmol/L)	162 (92–267)	172 (93–281)
Arterial pH	7·40 (7·31–7·4)	7·40 (7·31–7·43)
Arterial lactate (mmol/L)	2·8 (1·9–4·5)	2·6 (1·8–4·9)
Day 2		
INR	3·0 (2·1–4·6)	3·0 (2·2–4·9)
Arterial lactate (mmol/L)	2·0 (1·4–3·2)	2·1 (1·4–3·2)

Data are median (IQR) or n (%). INR=international normalised ratio. AST=aspartate aminotransferase. *Sequential Organ failure Assessment (SOFA) cardiovascular score ≥3.

Table 1: **Demographics, clinical and laboratory findings, and outcome of derivation and initial validation cohorts**

	Copenhagen (n=151)	Edinburgh (n=90)	Birmingham (n=72)	King's College Hospital (n=99)	All (n=412)
Age (years)	52 (36–61)	35 (27–46)	40 (31–46)	39 (30–50)	42 (31–53)
Sex					
Female	93 (62%)	49 (54%)	39 (54%)	59 (60%)	240 (58%)
Male	58 (38%)	41 (46%)	33 (46%)	40 (40%)	172 (42%)
INR	3·0 (2·2–4·3)	5·3 (3·9–7·2)	4·9 (3–7·1)	6·2 (3·8–9·7)	3·9 (2·7–6·0)
Arterial pH	7·42 (7·36–7·46)	7·41 (7·31–7·47)	7·29 (7·19–7·38)	7·40 (7·35–7·45)	7·40 (7·30–7·45)
Creatinine (μmol/L)	82 (56–153)	119 (73–216)	186 (96–297)	146 (76–273)	118 (66–236)
Arterial lactate (mmol/L)	2·6 (1·7–4·9)	2·9 (1·8–5·2)	5·6 (3·9–10·9)	3·3 (2·3–7·3)	3·5 (2·1–7·1)
Glasgow coma score	15 (8–15)	15 (14–15)	9 (9–10)	14 (9–15)	14 (9–15)
Cardiovascular failure*	26 (17%)	17 (19%)	57 (79%)	59 (60%)	159 (39%)
Died (%)	23 (15%)	13 (14%)	36 (50%)	27 (27%)	99 (24%)

Data are median (IQR) or n (%). INR=international normalised ratio. *Sequential Organ Failure Assessment (SOFA) cardiovascular score ≥3.

Table 2: **Demographics, admission (day 1) laboratory findings, and outcome of external validation sets**

Results

We included 500 consecutive patients who had not undergone transplantation with severe paracetamol-induced hepatotoxicity admitted during the period 2000–12 to the Liver Intensive Therapy Unit at King's College Hospital in the primary derivation (n=350) and initial internal validation test cohorts (n=150). External validation cohorts included 151 patients admitted during the period 2011–13 from the Rikshopitalet Liver Unit, 90 patients admitted in 2008–14 to the Scottish Liver Transplant Unit at the Edinburgh Royal Infirmary, 72 patients admitted in 2004–13 to the intensive care unit of Queen Elizabeth Hospital, Birmingham, and 99 patients treated at King's College Hospital Liver Intensive Therapy Unit between 2012 and 2014. Characteristics of the derivation and initial validation sets are shown in table 1, and features on admission of the external validation cohorts are shown in table 2. 78 (22%) of 350 patients in the derivation set died during their hospital stay; median day of death was 9 days (IQR 2–14) after admission. Patients who died were older and, on admission, had evidence of more severe liver dysfunction with higher INR and arterial lactate concentrations, and worse extra-hepatic organ failure with more severe encephalopathy, cardiovascular, and renal dysfunction than did survivors (appendix p 3). Of note, 49 (46%) of 106 patients with a Glasgow coma score of 9 or less on day 1 died versus 29 (12%) of 244 with a Glasgow coma score above this threshold (relative risk 3·9 [2·6–5·8]; p<0·0001).

After extensive model fitting and variable selection, admission (day 1) variables of age, Glasgow coma score, arterial pH and lactate, creatinine, INR, and SOFA cardiovascular failure were identified as the best clinical predictors. Hazard ratios for these component variables on admission are shown in table 3 and the predictive equation in the appendix p 4. Based on the day 1 model, we explored a day 2 model with additional changes in all clinical variables between day 1 and day 2, and, after backward stepwise variable selection, found only changes in blood lactate and INR to be significantly associated with survival. A dynamic day 2 survival model was thus further developed based on clinical variables on day 1 plus changes in blood lactate concentration and INR between day 1 and 2 to reflect the changing patterns in these crucial variables (table 3, appendix p 4). Exploration of models using data up to 7 days after admission did not show significant improvements in model performance above the dynamic day 2 model (data not shown).

When applied to the derivation set, AUROC for prediction of death using the day 1 model was 0·95 (95% CI 0·91–0·99) at 7 days, 0·94 (0·90–0·97) at 15 days, and 0·92 (0·88–0·96) at 30 days, and for the dynamic day 2 model was 0·96 (0·93–1·0) at 7 days, 0·95 (0·91–0·98) at 15 days, and 0·93 (0·88–0·97) at 30 days (figure 1). Assessment of calibration using RMSE in prediction of 30-day survival gave values of 0·1123 for the day 1 model and 0·1317 for the dynamic day 2 model for the derivation set.

35 (23%) of the 150 patients in the initial validation set died during their hospital stay. Using the day 1 model, AUROC for the prediction of death was 0·93 (95% CI 0·86–1·00) at 7 days, 0·91 (0·85–0·96) at 15 days, and 0·89 (0·84–0·95) at 30 days, and with the dynamic day 2 model was 0·94 (0·88–1·0) at 7 days, 0·91 (0·86–0·96) at 15 days, and 0·90 (0·85–0·95) at 30 days (figure 2). Assessment of calibration using RMSE in prediction of 30-day survival gave values of 0·1642 using the day 1 model and 0·0626 for the day 2 model (figure 2). Based on the proposed survival models, individual survival curves were calculated and updated during the first 2 days of intensive care unit admission (figure 3).

Characteristics of the four validation sets differed significantly from one another and from the primary cohorts. Median age of the Danish patients was higher than that of the other cohorts, and the Birmingham cohort had more severe acidosis, lactate, and creatinine, more severe encephalopathy, and higher mortality than the other cohorts (table 2). Overall, the day 1 single component variables were missing in 19 (5%) of 412 cases and day 2 variables in 141 (34%). Where day 2 INR and lactate variables were missing, values were categorised as showing no improvement.

AUROC and RMSE for the individual validation sets are shown in table 4. In the combined external validation sets, AUROC for 30-day survival was 0·91 (95% CI 0·87–0·94) for the day 1 model and 0·91 (0·88–0·95) for the day 2 model and RMSE was 0·079 for the day 1 model and 0·107 for the day 2 model (figure 4).

During the study periods of the derivation and validation sets, 116 patients underwent emergency liver transplantation: 84 at King's College Hospital, 23 at Queen Elizabeth Hospital, and eight at Edinburgh Royal Infirmary. Comparison of admission variables of patients who died with those who underwent transplantation showed significant differences, with patients who died being older and with higher arterial lactate concentrations, and lower Glasgow coma scores and INR values (appendix p 5). Using the day 1 model, median predicted 30-day survival for the cohort who underwent transplantation was 51% (95% CI 33–85), and in patients with severe hepatic encephalopathy (Glasgow coma score ≤9; n=70), predicted survival was 36% (18–73).

Discussion

We present the development and validation of high-performance statistical models to support decision making in the care of patients with paracetamol-induced acute liver failure. The models proposed show very good discrimination and calibration, confirmed on application to several independent external validation datasets with patients with a range of illness severity, and show substantial promise as prognostic tools. We developed models both on (day 1) and after (day 2) admission because their combined use provides practical support for key clinical decisions not addressed by existing single time

	Hazard ratio	95% CI	p value
Day 1 model			
Age (per 5 year increase)	1·07	0·97–1·18	0·168
Cardiovascular failure*	2·41	1·39–4·17	0·002
Glasgow coma scale score	0·90	0·84–0·98	0·009
Arterial pH	0·06	0·01–0·61	0·018
Log (creatinine [per 10 units])	1·74	1·12–2·7	0·013
Log (INR)	1·53	1·04–2·24	0·029
Square root (arterial lactate)	2·01	1·53–2·63	$4·02 \times 10^{-7}$
Day 2 model			
Age (5 years)	1·07	0·98–1·18	0·146
Day 1			
Cardiovascular failure	3·14	1·8–5·49	$5·94 \times 10^{-5}$
Glasgow coma scale score	0·90	0·83–0·97	0·005
Arterial pH	0·09	0·01–0·95	0·045
Log (creatinine [per 10 units])	1·70	1·08–2·67	0·022
Log (INR)	1·97	1·33–2·91	$6·91 \times 10^{-4}$
Square root (arterial lactate)	1·78	1·37–2·32	$1·46 \times 10^{-5}$
Day 2			
Lower arterial lactate	0·31	0·13–0·69	0·0045
Lower INR	0·54	0·29–0·99	0·046

INR=international normalised ratio. *Sequential Organ Failure Assessment (SOFA) cardiovascular score ≥3.

Table 3: **Clinical predictors included in admission predictive model in the derivation dataset (n=350)**

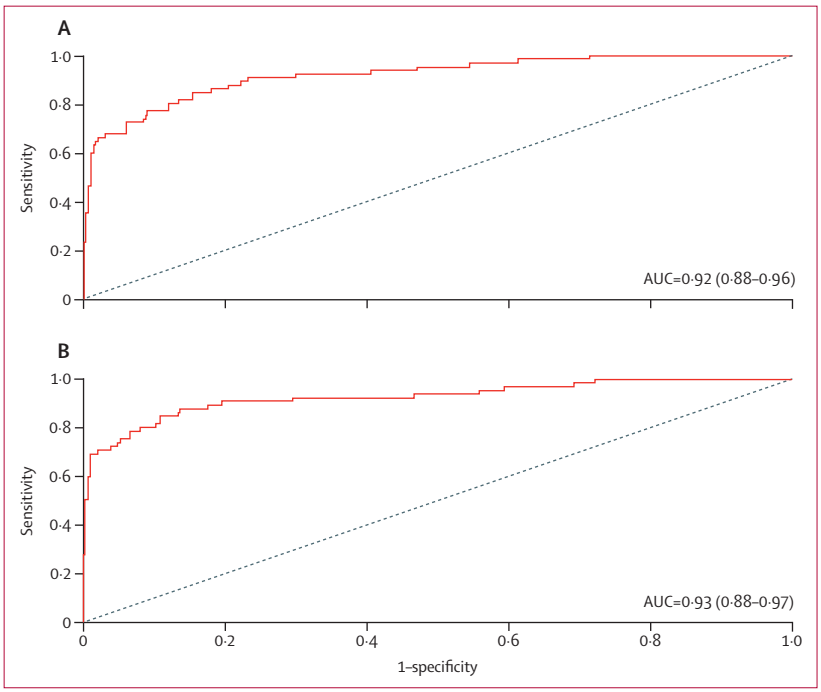

Figure 1: **Area under receiver operating characteristic curve for the derivation dataset**
(A) Day 1 model and (B) day 2 model. Data from the King's College Hospital derivation set (n=350).

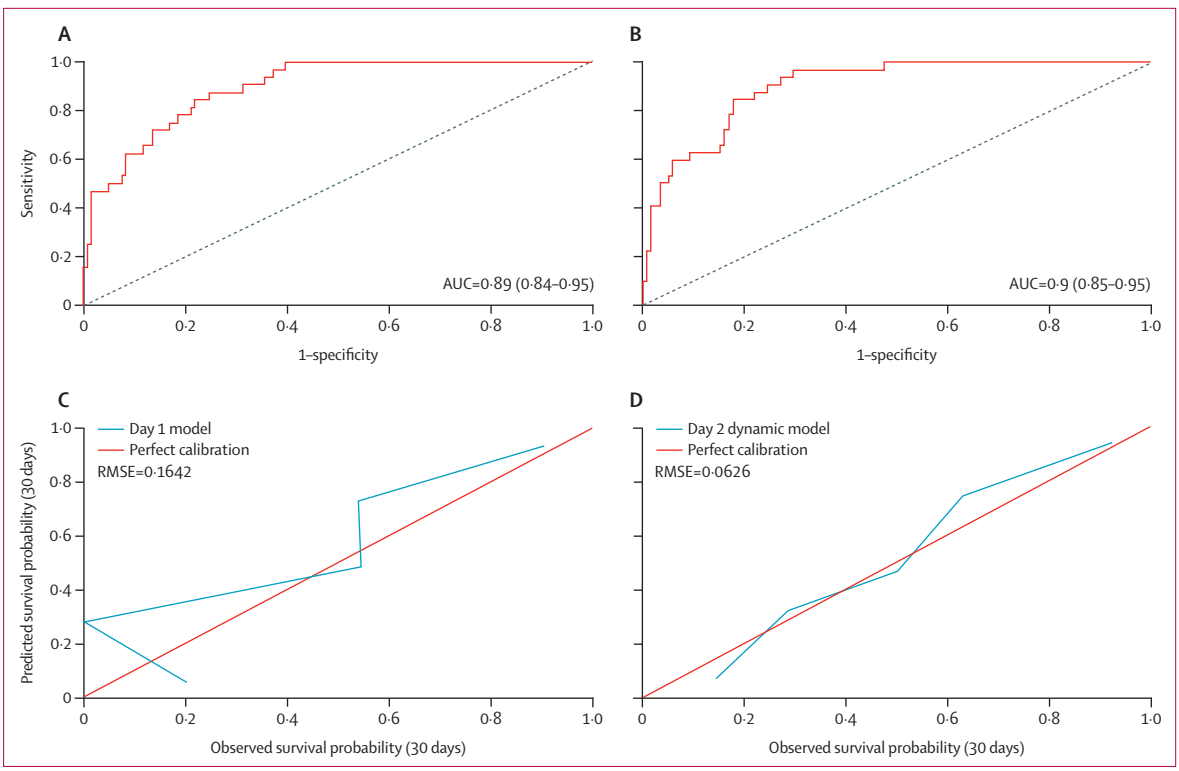

Figure 2: **Discrimination and calibration of day 1 and day 2 models in the initial validation dataset**
(A) Area under receiver operating characteristic curve (AUROC) for the day 1 model. (B) AUROC for day 2 model. (C) Calibration curve for the day 1 model. (D) Calibration curve for day 2 model. Data from the King's College Hospital initial validation cohort (n=150). RMSE=root mean square error.

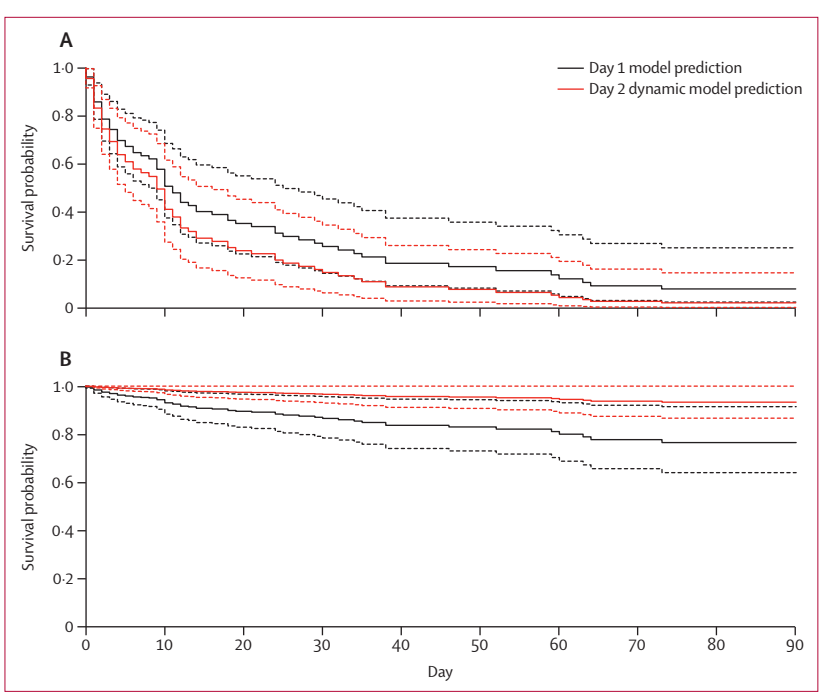

Figure 3: Predicted survival curves by day 1 and day 2 models for King's College Hospital patients
(A) Died at 23 days. (B) Patient survived to hospital discharge. Broken lines represent 95% CI.

point prognostic criteria. In management of paracetamol-induced acute liver failure the first 2 days are crucially important for the selection of transplant candidates; early prediction is required and is directly clinically relevant.[14] Experience has shown that the clinical condition of patients with paracetamol-induced liver failure might change rapidly after admission and initial therapy—with improvement in some patients and deterioration in others. The two-stage assessment of prognosis enables quantitation of this change, objectively reassessing prognosis in patients who show signs of improvement. Assessment of test performance showed both models to have high discrimination but importantly calibration was improved in the day 2 model. This sequential approach, derivation from recent patient cohorts, and use of novel important prognostic variables represent clear advances beyond the original King's College Criteria.

There has been criticism in the statistical literature of the use of stepwise regression for model fitting relating to its biases and suboptimal results. However, we used stepwise regression for its simplicity and in conjunction with expert opinion to decide which variables to include in the model. Our predictive models were robustly trained and validated on both internal and external datasets, which resulted in excellent model performance. Nevertheless, there are alternatives to stepwise regression, which include partial least squares regression

and least absolute shrinkage and selection operator. These alternatives can overcome some of the shortcomings of stepwise regression, although they have their own limitations.[20] We will explore these novel methods in future research to further enhance the performance of our predictive models.

We chose to exclude patients who underwent emergency liver transplantation from model derivation and initial validation. Valid criticism can be made both of the inclusion or exclusion of patients who underwent transplantation in this process, and both have been used in the scientific literature. Our main concern was risk of reference bias in view of the drastic improvements in survival with medical management alone in patients with paracetamol-induced liver failure, and the possibility that detail of transplantation practice varied between centres.[3,5] Although this approach could introduce changes in cohort composition, it is unlikely to have yielded a less unwell sample since, in key respects, those who later died were more severely ill. Predictions of a survival model derived in a population that did not undergo transplantation applied to one in which patients underwent transplantation should be interpreted with caution. However, this survival estimate is derived from the analysis of a population of patients with illness of the same cause, including cases that had similar acuity of illness and with a model that uses variables derived solely from the pre-transplant phase of illness.

When applying the day 1 model to patients who underwent transplantation during the study period, the median predicted survival was slightly more than 50% but fell to 36% when only those with severe encephalopathy were considered. These observations reinforce the key prognostic importance of development

	n	Day 1 model		Day 2 model	
		AUROC (95% CI)	RMSE	AUROC (95% CI) 95% CI	RMSE
Copenhagen	151	0·93 (0·88–0·98)	0·140	0·94 (0·87–1·00)	0·208
Edinburgh	90	0·88 (0·79–0·96)	0·281	0·89 (0·81–0·97)	0·239
Birmingham	72	0·84 (0·74–0·94)	0·053	0·83 (0·73–0·93)	0·111
Kings	99	0·92 (0·85–0·99)	0·117	0·93 (0·86–0·99)	0·185
Combined	412	0·91 (0·87–0·94)	0·079	0·91 (0·88–0·95)	0·107

AUROC=area under receiver operating characteristic curve; RMSE=root mean square error.

Table 4: **Diagnostic test performance of both models for predicting 30-day survival applied to individual and combined external validation sets**

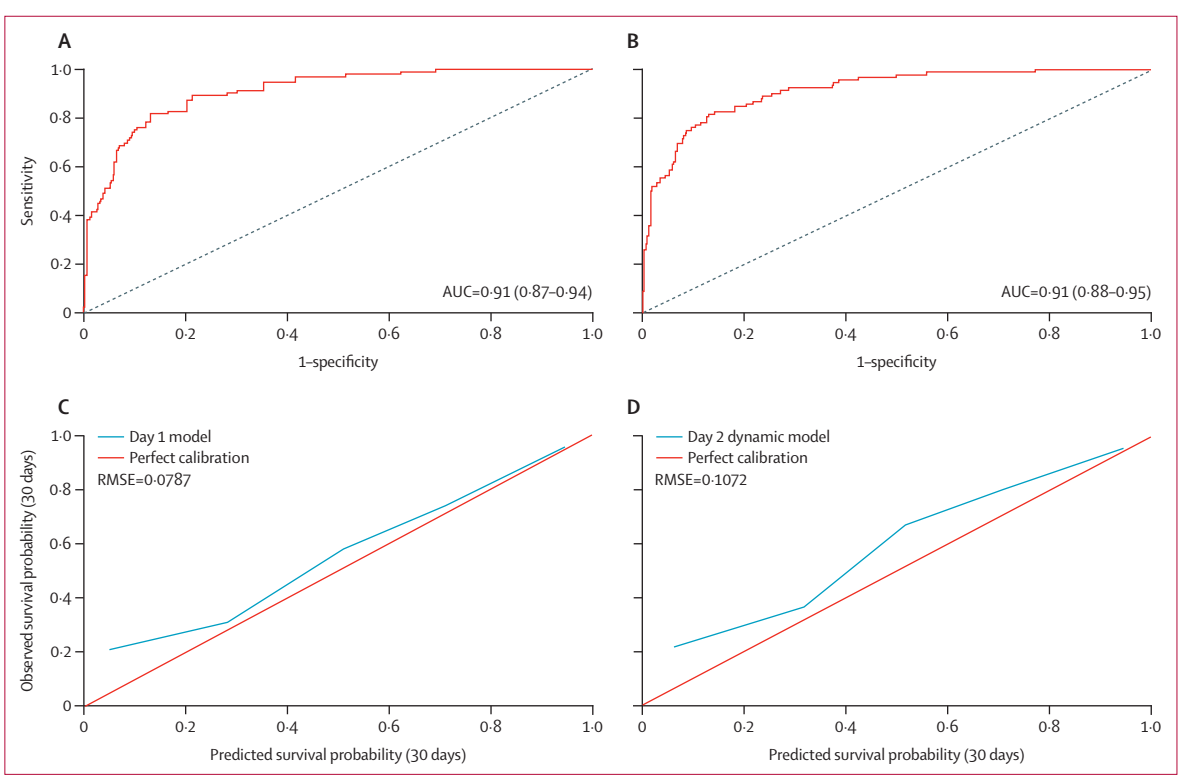

Figure 4: **Discrimination and calibration of day 1 and day 2 models in combined external validation sets**
(A) Area under receiver operating characteristic curve (AUROC) for day 1 model. (b) AUROC for day 2 model. (C) Calibration curve for day 1 model. (D) Calibration curve for day 2 model. Data from combined external validation datasets (n=412).

of hepatic encephalopathy and that transplantation should not be considered in the absence of high-grade encephalopathy. Importantly, a substantial proportion of patients who underwent transplantation might have survived with medical management alone. These findings quantify the substantial improvements in survival independent of liver transplantation that have occurred over time, and suggest that transplantation as an intervention in paracetamol-induced acute liver failure requires comprehensive re-evaluation.[3,21]

In considering the application of these models in a clinical setting, limitations to their potential practical use should be assessed. In other critical illness scoring systems, use of individual patient outcome prediction has historically been with caution, because the primary purpose of most such scores was for group outcome assessment and quality assurance.[22–24] However, there is clear precedent within hepatology for use of scores, including Model for End-Stage Liver Disease (MELD), for individual decision making in relation to transplantation—and in the widely accepted use of the King's College Criteria and other poor prognosis criteria for selection of patients with acute liver failure for emergency liver transplantation.[24,25]

We see these models as tools to quantify the risk of death and support expert clinical judgment and decision making; experience in other areas of acute and critical care medicine suggests that combining an objective prognosis measure with a physician's clinical estimate results in the most accurate assessment of actual prognosis.[23,26,27] Further, the sequential assessment of our models provide is likely to be of benefit because, in critically ill patients with and without liver disease, trends in illness severity provide additional prognostic value over single static determinations.[23,28–30] Rather than considering a single timepoint survival estimate, transplantation wait listing decisions might best be made from observations of the dynamic course of the illness. An obvious issue is that of the threshold of estimated survival that should trigger addition of a patient to the waiting list; a figure of 25% demonstrated in recent reports of early transplantation for acute alcoholic hepatitis provides a useful parallel.[31] A website has been developed to use the new paracetamol prediction models in a form that is accessible and easy to use, and whose output will provide real-time estimates of expected survival, while accumulating a further prospective confirmatory validation set.

For the **website** see www.kingsalfpredictor.org

In developing these models we deliberately chose not to rely on the reported timing of drug ingestion in our patient selection and survival modelling. In practice, this information is often inconsistent or unavailable and, in the case of overdoses, staggered over days, making definition of a specific timepoint of ingestion impossible. Even without reliance on this information, the model functioned well. However, it is important to recognise that these patients were assessed in liver transplantation centres usually days after drug ingestion, with established and clinically significant hepatic necrosis and after receiving initial resuscitative measures at their receiving hospitals before transfer. Use of the models to predict survival in patients soon after overdose or at early after first presentation has not been assessed.

These models were primarily designed to identify patients whose survival would be enhanced by emergency liver transplantation. However, identification as having a poor prognosis does not necessarily mean that survival will be improved by liver transplantation. The models are designed to predict survival without transplantation but not survival after surgery, when factors not considered in our model are of prognostic importance.[32] In many cases contraindications to emergency liver transplantation might be present, and here the models might rather serve to guide patients and family members in the expected outcome of illness.

Contributors

The study was conceived by WB and JW. YW and WB developed the predictive model and WB wrote the first draft of the paper. WB, JM, NM, AE, DH, KS, and FSL contributed data, and all authors contributed in detail to the writing of the final version of the report.

Declaration of interests

We declare no competing interests.

Acknowledgments

This study was funded by the Foundation for Liver Research.

References

1 Bernal W, Auzinger G, Dhawan A, Wendon J. Acute liver failure. *Lancet* 2010; **376:** 190–201.

2 Larson AM, Polson J, Fontana RJ, et al. Acetaminophen-induced acute liver failure: results of a United States multicenter, prospective study. *Hepatology* 2005; **42:** 1364–72.

3 Bernal W, Hyyrylainen A, Gera A, et al. Lessons from look-back in acute liver failure? A single centre experience of 3300 patients. *J Hepatol* 2013; **59:** 74–80.

4 Ichai P, Samuel D. Etiology and prognosis of fulminant hepatitis in adults. *Liver Transpl* 2008; **14** (suppl 2): 67–79.

5 Ichai P, Legeai C, Francoz C, et al. Patients with acute liver failure listed for superurgent liver transplantation in France: reevaluation of the Clichy-Villejuif criteria. *Liver Transpl* 2015; **21:** 512–23.

6 O'Grady JG, Alexander GJ, Hayllar KM, Williams R. Early indicators of prognosis in fulminant hepatic failure. *Gastroenterology* 1989; **97:** 439–45.

7 Bailey B, Amre DK, Gaudreault P. Fulminant hepatic failure secondary to acetaminophen poisoning: a systematic review and meta-analysis of prognostic criteria determining the need for liver transplantation. *Crit Care Med* 2003; **31:** 299–305.

8 Craig DGN, Ford AC, Hayes PC, Simpson KJ. Systematic review: prognostic tests of paracetamol-induced acute liver failure. *Aliment Pharmacol Ther* 2010; **31:** 1064–76.

9 Bernal W, Donaldson N, Wyncoll D, Wendon J. Blood lactate as an early predictor of outcome in paracetamol-induced acute liver failure: a cohort study. *Lancet* 2002; **359:** 558–63.

10 Esterline RL, Ray SD, Ji S. Reversible and irreversible inhibition of hepatic mitochondrial respiration by acetaminophen and its toxic metabolite, N-acetyl-p-benzoquinoneimine (NAPQI). *Biochem Pharmacol* 1989; **38:** 2387–90.

11 Zein JG, Wallace DJ, Kinasewitz G, Toubia N, Kakoulas C. Early anion gap metabolic acidosis in acetaminophen overdose. *Am J Emerg Med* 2010; **28:** 798–802.

12 Craig DGN, Bates CM, Davidson JS, Martin KG, Hayes PC, Simpson KJ. Overdose pattern and outcome in paracetamol-induced acute severe hepatotoxicity. *Br J Clin Pharmacol* 2011; **71:** 273–82.

13 Kumar R, Shalimar, Sharma H, et al. Persistent hyperammonemia is associated with complications and poor outcomes in patients with acute liver failure. *Clin Gastroenterol Hepatol* 2012; **10:** 925–31.

14 Reddy KR, Ellerbe C, Schilsky M, et al. Determinants of outcome among patients with acute liver failure listed for liver transplantation in the United States. *Liver Transpl* 2016; **22:** 505–15.

15 Roberts LN, Bernal W. Management of bleeding and thrombosis in critically ill patients with liver disease. *Semin Thromb Hemost* 2015; **41:** 520–26.

16 Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/ failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; **22:** 707–10.

17 Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974; **2:** 81–84.

18 Bernal W, Hall C, Karvellas CJ, Auzinger G, Sizer E, Wendon J. Arterial ammonia and clinical risk factors for encephalopathy and intracranial hypertension in acute liver failure. *Hepatology* 2007; **46:** 1844–52.

19 Perkins W, Tygert M, Ward E. Chi-square and classical exact tests often wildly misreport significance; the remedy lies in computers. Sept 15, 2011. http://arxiv.org/abs/1108.4126v2 (accessed July 11, 2016).

20 Harrell FE Jr. Regression modeling strategies with applications to linear models, logistic regression, and survival analysis. New York: Springer New York, 2001.

21 O'Grady J. Timing and benefit of liver transplantation in acute liver failure. *J Hepatol* 2014; **60:** 663–70.

22 Wong DT, Barrow PM, Gomez M, McGuire GP. A comparison of the acute physiology and chronic health evaluation (APACHE) II score and the trauma-injury severity score (TRISS) for outcome assessment in intensive care unit trauma patients. *Crit Care Med* 1996; **24:** 1642–48.

23 Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. *Crit Care Med* 2011; **39:** 163–69.

24 Lemeshow S, Klar J, Teres D. Outcome prediction for individual intensive care patients: useful, misused, or abused? *Intensive Care Med* 1995; **21:** 770–76.

25 Barber K, Madden S, Allen J, et al. Elective liver transplant list mortality: development of a United Kingdom end-stage liver disease score. *Transplantation* 2011; **92:** 469–76.

26 Knaus WA, Harrell FE Jr, Lynn J, et al. The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. *Ann Intern Med* 1995; **122:** 191–203.

27 Sniederman A, D'Agostino R, Pencina M. The Role of physicians in the era of predictive analytics. *JAMA* 2015; **314:** 25–26.

28 McPhail MJ, Shawcross DL, Abeles RD, et al. increased survival for patients with cirrhosis and organ failure in liver intensive care and validation of the chronic liver failure-sequential organ failure scoring system. *Clin Gastroenterol Hepatol* 2015; **13:** 1353–60.e8.

29 Ferreira F, Bota D, Bross A, Melot C, Vincent J. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* 2001; **286**: 1754–58.

30 Goettler CE, Waibel BH, Goodwin J, et al. Trauma intensive care unit survival: how good is an educated guess? *J Trauma Inj Infect Crit Care* 2010; **68:** 1279–87.

31 Mathurin P, Moreno C, Samuel D, et al. Early liver transplantation for severe alcoholic hepatitis. *New Engl J Med* 2011; **365:** 1790–800.

32 Germani G, Theocharidou E, Adam R, et al. Liver transplantation for acute liver failure in Europe: outcomes over 20 years from the ELTR database. *J Hepatol* 2012; **57:** 288–96.