http://www.kidney-international.org

Mortality after acute renal failure: Models for prognostic stratification and risk adjustment

GM Chertow[1], SH Soroko[2], EP Paganini[3], KC Cho[1], J Himmelfarb[4], TA Ikizler[5] and RL Mehta[2] for the Program to Improve Care in Acute Renal Disease (PICARD)

[1]Department of Medicine Research, Division of Nephrology, University of California San Francisco, San Francisco, California, USA; [2]Department of Medicine, Division of Nephrology, University of California San Diego, San Diego, California, USA; [3]Department of Medicine, Division of Nephrology, Cleveland Clinic Foundation, Cleveland, Ohio, USA; [4]Department of Medicine, Division of Nephrology, Maine Medical Center, Portland, Maine, USA and [5]Department of Medicine, Division of Nephrology, Vanderbilt University, Nashville, Tennessee, USA

To adjust adequately for comorbidity and severity of illness in quality improvement efforts and prospective clinical trials, predictors of death after acute renal failure (ARF) must be accurately identified. Most epidemiological studies of ARF in the critically ill have been based at single centers, or have examined exposures at single time points using discrete outcomes (e.g., in-hospital mortality). We analyzed data from the Program to Improve Care in Acute Renal Disease (PICARD), a multi-center observational study of ARF. We determined correlates of mortality in 618 patients with ARF in intensive care units using three distinct analytic approaches. The predictive power of models using information obtained on the day of ARF diagnosis was extremely low. At the time of consultation, advanced age, oliguria, hepatic failure, respiratory failure, sepsis, and thrombocytopenia were associated with mortality. Upon initiation of dialysis for ARF, advanced age, hepatic failure, respiratory failure, sepsis, and thrombocytopenia were associated with mortality; higher blood urea nitrogen and lower serum creatinine were also associated with mortality in logistic regression models. Models incorporating time-varying covariates enhanced predictive power by reducing misclassification and incorporating day-to-day changes in extra-renal organ system failure and the provision of dialysis during the course of ARF. Using data from the PICARD multi-center cohort study of ARF in critically ill patients, we developed several predictive models for prognostic stratification and risk-adjustment. By incorporating exposures over time, the discriminatory power of predictive models in ARF can be significantly improved.

Kidney International (2006) **70,** 1120–1126. doi:10.1038/sj.ki.5001579; published online 19 July 2006

KEYWORDS: acute renal failure; mortality; severity of illness; hemodialysis; hemodiafiltration; PICARD

Correspondence: *GM Chertow, Department of Medicine Research, Division of Nephrology, University of California San Francisco, UCSF Laurel Heights Suite 430, 3333 California Street, San Francisco, California 94118-1211, USA. E-mail: chertowg@medicine.ucsf.edu*

Received 19 November 2004; revised 26 July 2005; accepted 11 August 2005; published online 19 July 2006

In critically ill patients, acute renal failure (ARF) is associated with mortality rates in excess of 50%,[1–3] despite increasing availability of sophisticated intensive care, hemodialysis, and hemodiafiltration. Over the past two decades, numerous observational studies have aimed to identify clinical predictors of mortality in ARF. Some have attempted to validate generic or disease-specific predictive instruments;[4–13] others have derived new predictive models;[14–17] and most have been developed from single centers[9,10,14] and have examined risk factors at a single point in time.[8,15,18]

In an effort to develop a registry of critically ill patients with ARF across multiple clinical sites, we created the Program to Improve Care in Acute Renal Disease (PICARD). The major goal of PICARD was to leverage the diversity of several sites and a relatively large sample of patients to better understand the demographic, process, renal, and extra-renal factors associated with relevant clinical outcomes, including mortality, non-recovery of kidney function, and resource utilization. In doing so, our hope was to provide a contemporary view of the ARF disease process and to generate hypotheses that might be eligible for testing in prospective clinical trials (e.g., medication use, dialysis modality and timing, nutritional therapy, etc.).

Herein, we provide a comprehensive analysis of demographic and clinical factors associated with mortality in the 618 patients enrolled in PICARD. To extend previously published work in this area, we aimed to provide predictive models at three key time points during the course of ARF – on the day of ARF diagnosis, on the day of consultation, and in a subcohort, on the day of initiation of dialysis or hemodiafiltration – as well as to provide an integrated approach taking advantage of longitudinal data collection, using models incorporating time-varying covariates. We hypothesized that oliguria, sepsis, respiratory failure, and hepatic failure would be consistently associated with mortality.

RESULTS

Six hundred and eighteen patients were enrolled in PICARD. Table 1 shows a summary of demographic and clinical

characteristics at three key time points during the course of ARF in the intensive care unit (ICU). The cohorts for the day of ARF diagnosis and day of consultation were identical; demographic and historical information were by definition identical; however, dynamic characteristics (e.g., organ failure, vital signs, and laboratory results) changed over time. Three hundred and ninety-eight (64%) patients required initiation of dialysis. Although inference tests were not conducted, one can appreciate that the subcohort requiring dialysis tended to have more significant azotemia, lower urine output, and tended to have more extensive organ system failure, particularly of the lungs and liver, than the full PICARD cohort. Nearly one-third of patients had evidence of significant baseline chronic kidney disease (CKD).

Table 1 | Patient characteristics at the day of ARF diagnosis, consultation, and first procedure

Parameter	Day of ARF diagnosis (*n*=618)	Day of consultation (*n*=618)	Day of first procedure (*n*=398)
Mean age (years)	59	59	57
% Female	41	41	42
Race/ethnicity			
Caucasian	80%	80%	78%
African American	8%	8%	7%
Asian	6%	6%	8%
Hispanic	4%	4%	5%
Other	2%	2%	2%
% History of CKD, stage IV or above	32	32	36
% Surgery pre/at ICU admission	38	38	35
% History of hypertension	53	53	48
% History of diabetes mellitus	29	29	27
% History of COPD	16	16	15
% History of heart failure	28	28	26
% History of coronary artery disease	37	37	30
Mean no. of organ systems failed	2.7	2.9	3.1
% Central nervous system failure	21	20	24
% Liver failure	29	31	39
% Hematologic failure	27	25	30
% Cardiovascular failure	48	49	46
% Respiratory failure	42	60	68
% Mechanical ventilation	33	47	36
% Acute lung injury	30	47	45
% ARDS	19	30	30
% Sepsis or septic shock	26	29	37
Mean heart rate (per min)	95	95	96
% Tachycardia	60	39	42
Mean systolic BP (mm Hg)	115	115	114
Mean diastolic BP (mm Hg)	58	56	56
Mean arterial BP (mm Hg)	77	76	75
Mean pulse pressure (mm Hg)	57	59	58
Mean temperature (°C)	37	37	36.9
Median urine output (ml)	1150	1258	800
% Oliguria ($\leqslant$400 ml/day)	27	29	49
Mean respiratory rate	20	20	20
Mean AM weight (kg)	86	87	90
Median total bilirubin (mg/dl)	1.5	1.7	2.4
Mean creatinine (mg/dl)	2.8	3.4	4.5
Mean BUN (mg/dl)	54	65	86
Mean platelets (1000/mm^3)	175	155	146
% Thrombocytopenic ($<150\times10^6$/l)	49	57	62
Mean pH	7.3	7.4	7.3
Mean potassium (mEq/l)	4.6	4.6	4.6
Mean bicarbonate (mEq/l)	21.9	21.4	20.3
Mean leukocyte (1000/mm^3)	14	14	14.9
Mean hemoglobin (g/dl)	10.5	10.2	10.1

AM, adrenomedullin; ARDS, adult respiratory distress syndrome; ARF, acute renal failure; BP, blood pressure; BUN, blood urea nitrogen; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; ICU, intensive care unit.

Predictors at discrete time points

Day of ARF diagnosis. Figure 1 demonstrates in-hospital mortality rates from the day of ARF diagnosis. Tables 2 and 3 show logistic regression and proportional hazards regression models for the outcomes in-hospital death within 60 days, and the time to death within 60 days, respectively, considering the day of ARF diagnosis as the specified time point. Advanced age, liver failure, and higher blood urea nitrogen (BUN) concentrations at the time the creatinine-based ARF criteria were met were associated with mortality using both statistical methods. In the logistic regression model, baseline CKD was associated with lower, and sepsis associated with higher, risks of death. These factors did not reach statistical significance in the proportional-hazards models. Model discrimination was relatively poor; the area under the receiver operating characteristic (ROC) curves for death was 0.62. The model was well calibrated.

Day of consultation. Analogous models considering the day of nephrology consultation as the specified time point are also shown in Tables 2 and 3. Advanced age and liver failure remained significant predictors of death; in addition, parameters of kidney function and extra-renal organ system failure were also associated with the risk of death. When evaluated on the day of consultation, patients with lower urine output, serum creatinine <2.0 mg/dl, and higher BUN were at increased risk of death, confirming findings described in a single-center cohort predating PICARD.[22] Adult respiratory distress syndrome (defined as a ratio of oxygen tension to fractional inspired oxygen concentration (paO_2/FIO_2) <200 mm Hg), sepsis, and relative thrombocytopenia ($<150\times10^6$/l) were also significant predictors of death. The area under the ROC curve for the day of consultation model was considerably higher than for the day of ARF diagnosis models – 0.68 – and the model was well calibrated.

Day of initial dialysis procedure. Predictors of mortality among the subcohort of 398 patients who required dialysis during the ICU stay are also shown in Tables 2 and 3. As with the previous two key time points, advanced age and liver failure were significantly associated with mortality among patients with ARF requiring dialysis. In contrast to the day of consultation models, urine output was not associated with mortality when evaluated on the day of dialysis initiation. The mean serum creatinine and BUN at dialysis initiation were 4.5 mg/dl (10, 90% range 2.3–7.5 mg/dl) and 86 mg/dl (10, 90% range 34–142 mg/dl), respectively. Lower serum creatinine and higher BUN concentrations were associated with an increased risk of mortality. The day of dialysis

initiation ranged from 0 to 43 days after ARF diagnosis and 0 to 31 days after nephrology consultation.

The provision of dialysis. In Table 4, we show predictors of dialysis from the day of diagnosis and the day of consultation using logistic regression. As expected, lower urine output and higher BUN were associated with the use of dialysis during the ARF episode. Interestingly, older patients were less likely to receive dialysis. Although older age was associated with mortality, early deaths among the elderly (i.e., before dialysis could be provided) did not account for the difference in dialysis practice.

Comparison with generic severity of illness scores. Table 5 shows areas under the ROC curves and likelihood ratios for the models described above (denoted 'PICARD') and 11

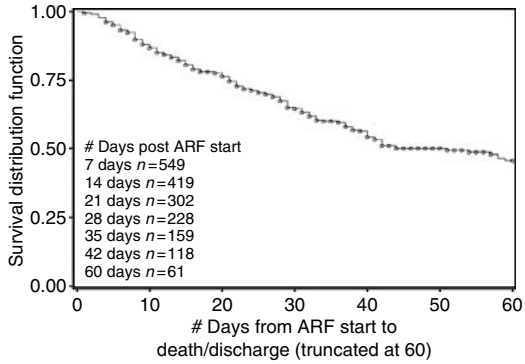

Figure 1 | In-hospital mortality rates within 60 days after ARF diagnosis.

Table 2 | Predictors of mortality using logistic regression

Parameter	Coefficient	OR	95% CI
Day of ARF diagnosis[a]			
Intercept	−1.5023	—	—
Age (per 10 years)	0.1055	1.11	1.00–1.24
CKD stage IV	−0.5555	0.57	0.39–0.84
BUN (per 10 mg/dl)	0.0568	1.06	1.00–1.12
Liver failure	0.4628	1.59	1.03–2.44
Sepsis or septic shock	0.5908	1.81	1.20–2.72
Day of consultation[b]			
Intercept	−1.2563	—	—
Age (per 10 years)	0.1241	1.13	1.01–1.27
Log urine output	−0.2063	0.81	0.72–0.93
Creatinine <2 mg/dl	0.6900	1.99	1.18–3.36
BUN (per 10 mg/dl)	0.0828	1.09	1.03–1.14
Liver failure	0.4811	1.62	1.09–2.41
ARDS	0.5800	1.79	1.23–2.60
Platelets $<150\times10^6$/l	0.5074	1.66	1.17–2.36
Sepsis or septic shock	0.4083	1.50	1.02–2.22
Day of first procedure[c]			
Intercept	−1.9506	—	—
Age (per 10 years)	0.1444	1.16	1.00–1.34
Creatinine (per 1 mg/dl)	−0.2091	0.81	0.72–0.92
BUN (per 10 mg/dl)	0.0860	1.09	1.03–1.15
Liver failure	0.5655	1.76	1.09–2.85
Respiratory failure	0.6100	1.84	1.11–3.04
Platelets $<100\times10^6$/l	0.7436	2.10	1.33–3.33
Sepsis or septic shock	0.5216	1.69	1.04–2.72

ARDS, adult respiratory distress syndrome; ARF, acute renal failure; BUN, blood urea nitrogen; CI, confidence interval; CKD, chronic kidney disease; OR, odds ratio; ROC, receiver operating characteristic.
[a]Area under ROC curve=0.62, Hosmer–Lemeshow χ^2=0.20.
[b]Area under ROC curve=0.68, Hosmer–Lemeshow χ^2=0.67.
[c]Area under ROC curve=0.72, Hosmer–Lemeshow χ^2=0.16.
Note: Different expressions of related variables included in individual models based on model performance and fit (e.g., platelets $<150\times10^6$/l vs $<100\times10^6$/l, discrete vs continuous creatinine in the day of consultation and day of procedure models).

Table 3 | Predictors of mortality using proportional hazards (Cox) regression

Parameter	Coefficient	RR	95% CI
Day of ARF diagnosis[a]			
Age (per 10 years)	0.0942	1.10	1.00–1.20
BUN (per 10 mg)	0.0466	1.05	1.01–1.09
Liver failure	0.6559	1.93	1.41–1.63
Day of consultation[b]			
Age (per 10 years)	0.1124	1.12	1.02–1.22
Log urine output (ml/day)	−0.1176	0.89	0.81–0.97
Creatinine <2 mg/dl	0.4621	1.59	1.09–2.30
BUN (per 10 mg/dl)	0.0560	1.06	1.02–1.09
Liver failure	0.4856	1.63	1.22–2.17
ARDS	0.3413	1.41	1.07–1.85
Platelets $<150\times10^6$/l	0.3217	1.38	1.05–1.81
Day of first procedure[c]			
Age (per 10 years)	0.1250	1.13	1.02–1.26
Liver failure	0.4827	1.62	1.18–2.23
Platelets $<100\times10^6$/l	0.4934	1.64	1.20–2.23
Sepsis or septic shock	0.4823	1.62	1.19–2.20

ARDS, adult respiratory distress syndrome; ARF, acute renal failure; BUN, blood urea nitrogen; CI, confidence interval; RR, risk ratio.
[a]*Note*: The baseline survival function at 30 days was 0.711 and at 60 days was 0.518. The mean values of the non-discrete variables were age 54.4 years and BUN 54.0 mg/dl. Likelihood ratio=23.3.
[b]*Note*: The baseline survival function at 30 days was 0.771 and at 60 days was 0.619. The mean values of the non-discrete variables were age 52.9 years, and BUN 67.7 mg/dl and log urine output 6.13 ml/day. Likelihood ratio=47.1.
[c]*Note*: The baseline survival function at 30 days was 0.782 and at 60 days was 0.621. The mean value of the non-discrete variable was age 52.9 years. Likelihood ratio=33.4.

Table 4 | Predictors of the need for dialysis

Parameter	Coefficient	OR	95% CI
Day of ARF diagnosis[a]			
Intercept	4.0685	—	—
Age (per 10 years)	−0.1978	0.82	0.73–0.92
Log urine output	−0.4028	0.67	0.55–0.81
BUN (per 10 mg/dl)	0.0609	1.06	1.00–1.13
Liver failure	0.6012	1.82	1.11–3.00
Day of consultation[b]			
Intercept	7.1449	—	—
Age (per 10 years)	−0.2344	0.79	0.70–0.89
Log urine output	−0.8709	0.42	0.34–0.51
BUN (per 10 mg/dl)	0.1012	1.11	1.05–1.17
Liver failure	0.5311	1.70	1.08–2.67

ARF, acute renal failure; BUN, blood urea nitrogen; CI, confidence interval; OR, odds ratio.
[a]Area under ROC curve=0.68, Hosmer–Lemeshow χ^2=0.75.
[b]Area under ROC curve=0.77, Hosmer–Lemeshow χ^2=0.46.
Note: The areas under the ROC curve were significantly different ($P<0.05$) for the day of ARF diagnosis and day of consultation.

Table 5 | (a) Areas under ROC curve for selected generic and disease-specific models[a] and (b) likelihood ratios for selected generic and disease-specific models[b]

Model	Day of ARF diagnosis	Day of consultation	Day of first procedure
(a)			
PICARD	0.62	0.68*	0.72†
CCF	0.58	0.65	0.68
APACHE II	0.63	0.66*	0.67†
APACHE III	0.66	0.70*	0.67
Brussels	0.55	0.51	0.48
Liaño	0.53	0.56	0.55
LOD	0.64	0.66	0.65†
MOD	0.62	0.67*	0.70†
MPM	0.56	0.60*	0.59
SAPS II	0.69	0.70	0.71†
SHARF	0.56	0.57	0.60†
SOFA	0.64	0.70*	0.73†
SOFA (NR)	0.62	0.65*	0.68†
(b)			
PICARD	23.3	47.1	33.4
CCF	1.9	14.0	20.0
APACHE II	6.8	21.5	9.6
APACHE III	13.2	32.3	13.6
Brussels	3.5	0.6	0.6
Liaño	2.1	5.2	3.0
LOD	10.6	16.7	4.1
MOD	5.8	17.5	8.3
MPM	2.6	13.6	6.3
SAPS II	24.4	29.5	18.4
SHARF	3.7	4.8	4.6
SOFA	10.2	38.7	30.1
SOFA (NR)	6.2	19.2	11.6

APACHE, acute physiology and chronic health evaluation; ARF, acute renal failure; CCF, Cleveland Clinic Foundation; LOD, logistic organ dysfunction score; MOD, multiple organ dysfunction score; MPM, mortality probability model; PICARD, Program to Improve Care in Acute Renal Disease; ROC, receiver operating characteristic; SAPS, simplified acute physiology score; SHARF, Stuivenberg Hospital Acute Renal Failure; SOFA, sequential organ failure assessment; SOFA (NR), SOFA with no renal points.
[a]*Note*: Comparisons within models by time frame.
*Day of consultation significantly higher than the day of ARF diagnosis.
†Day of first procedure significantly higher than the day of consultation for the subset of patients who were dialyzed.
[b]*Note*: Likelihood ratios are derived from proportional hazards models.
A higher likelihood ratio indicates better model discrimination.

other predictive equations – eight generic and three ARF specific (Cleveland Clinic Foundation (CCF), Stuivenberg Hospital Acute Renal Failure, and Liaño equations). It should be noted that several of the predictive equations were specifically developed at discrete time points (e.g., acute physiology and chronic health evaluation (APACHE) II upon ICU admission, CCF upon initiation of dialysis in the ICU), and the relative performance of the models vs the PICARD model may reflect the timing of model derivation. For example, the CCF score was developed among patients with ARF requiring dialysis. The predictive power of the CCF model was poor when applied on the day of ARF diagnosis, but very good when applied on the day of dialysis initiation. Among existing generic predictive models, the sequential organ failure assessment (SOFA) score (with the 'renal' SOFA organ system included) performed most favorably, particularly at the two later time points.

Table 6 | Predictors of mortality using time-dependent covariates

Cox model	Parameter RR	95% CI
Age (per decade)	1.13	1.01–1.26
Sepsis[a]	1.87	1.33–2.63
CNS failure[a]	4.58	3.30–6.35
Cardiovascular failure[a]	4.50	3.11–6.51
Liver failure[a]	1.90	1.34–2.71
Hematologic failure[a]	1.46	1.01–2.10
Dialysis[b]	1.79	1.21–2.66

CI, confidence interval; CNS, central nervous system; RR, relative risk.
[a]Sepsis status and organ system failure updated daily, last value carried forward where missing.
[b]Dialysis status carried forward after initiation.

Time-dependent covariate analyses. Table 6 shows the results of proportional hazards regression models exploring time to death (with censoring at discharge or hospital day 60). Using integrated exposure data, the risk of mortality was associated with older age, sepsis, and the presence of central nervous system, cardiovascular, liver, and hematologic failure. The provision of dialysis was associated with an increased risk of death. In companion models where parameters of kidney function (rather than dialysis) were included as candidate variables, higher serum creatinine concentrations and oliguria were significantly associated with mortality; in these models, parameter estimates for the extrarenal covariates were virtually identical (data not shown).

DISCUSSION

Herein, we provide a comprehensive evaluation of risk factors for mortality in the PICARD multi-center cohort study of critically ill patients with ARF. In contrast to other studies, in which a single discrete time point during the course of ARF was examined, we examined three key time points – the day of ARF diagnosis, the day of consultation, and in the subset of patients who required dialysis, the day of first procedure. All generic and disease-specific models performed relatively poorly early in the course of ARF. With established ARF (on the day of consultation), or ARF requiring dialysis, many of the models' predictive power improved significantly.

The ability to predict accurately in-hospital mortality in patients with ARF is important for several reasons. First, objective data can be employed in shared medical decision-making, to the benefit of practitioners, patients, and their families. Second, policy makers and practitioners can use prognostic models for quality improvement by objectively evaluating relative performance among ICUs, hospitals, or larger health-care delivery systems. Finally, knowledge of the key factors associated with mortality in critically ill patients with ARF may inform the design of prospective clinical trials, improving sample size or effect size estimates and potentially guiding a stratified randomization.

As expected, the models derived from the PICARD data themselves demonstrated marginally superior performance characteristics when compared with other models. Of the

generic severity scores, the SOFA score performed best in this cohort. Of the disease-specific severity scores, the CCF score performed best, particularly among patients requiring dialysis, the subtype of patients in whom the CCF score was derived.[9] Other frequently cited severity scores for ARF in the ICU, including those derived by Liaño[12] and others, performed poorly with attempted cross-validation in the PICARD cohort. To determine whether the models derived here are 'overfit' (i.e., uniquely predictive in the current data set), the PICARD models will require cross-validation in other populations. As PICARD included patients from five medical centers and encompassed considerable heterogeneity in patient characteristics,[19] models derived from PICARD may be less prone to overfitting than similar models derived from individual centers.[5,7,14]

There are several important theoretical advantages to disease-specific predictive models in critical illness. Virtually all generic severity scores include 'points' for ARF; in an ARF cohort, these points are assigned equivalently to all persons, limiting the value of the information. Moreover, clinical or laboratory indicators of injury may reflect different complications in cohorts depending on disease stage. For example, in several studies of ARF in the ICU,[3,15,20–22] a lower serum creatinine concentration has been associated with increased mortality, possibly reflecting reduced creatinine generation owing to diminished muscle mass (associated with advanced age, wasting, or other acute or chronic disease) or possibly reflecting hemodilution owing to volume overload.[23] Generic severity scores would classify patients with higher serum creatinine as higher risk.

Prior studies on risk assessment in ARF have yielded conflicting results.[8,10,11,20,24] In part, these differences reflect the inclusion populations and the time point selected for the evaluation of risk factors (typically the day of nephrology consultation or first procedure). Some prior studies have demonstrated an increased risk in older vs younger individuals, and in men relative to women; these associations have not been shown consistently.[16] Most studies have demonstrated a strong link between extra-renal organ system failure and the risk of death with ARF in the ICU, propagating the notion that persons die with, rather than from ARF. However, studies in which patients were followed relatively early in the disease course suggest an association between the severity of ARF (indicated by the provision of dialysis, often with oliguria) and mortality. These and other epidemiological data strongly suggest some persons indeed die of, not only with, ARF.[25–27]

There are several strengths to the analyses presented here. The PICARD study employed multiple sites and a relatively large number of patients were studied. Longitudinal data were collected, affording the investigators the ability to examine multiple time points. The study design allowed the determination of a predictive model at the time of ARF diagnosis. Although this point coincided with the day of hospitalization, ICU admission, and consultation in some patients, it preceded consultation in two-thirds of the cases. We demonstrate that models derived very early in the course of ARF do not discriminate mortality well, possibly reflecting the lack of specificity of current definitions of ARF. Nevertheless, providing a uniform time zero to track the course of ARF independent of the timing of consultation diminishes bias considerably, and permits evaluation of time-varying exposures, the latter allowing risk stratification with less misclassification. For example, among all patients with sepsis and ARF, certain individuals will resolve sepsis over several days with antibiotics and supportive therapy. Others will develop progressive septic shock and other end-organ complications refractory to conventional therapy. Use of static exposure data would misclassify the former patients, and correctly classify the latter. If disseminated intravascular coagulation were to develop in the wake of sepsis, it would not be identified as a risk factor with traditional analysis, and sepsis or oliguria or another factor associated with disseminated intravascular coagulation might be 'credited' with increased risk using the static models. This process of integrating exposure data over the course of ARF should improve prognostic stratification in subsequent cohort studies, and may be useful in clinical trial design, depending on the timing of the intended intervention.

There are also several important limitations. While serving diverse and distinct patient populations, all PICARD sites are tertiary-care academic medical centers, and the average level of acuity among ICU patients may be higher than in non-tertiary care settings. Conversely, as we required informed consent from all patients (or surrogates) and a significant fraction of patients died before consent could be obtained,[21] some of the sickest patients with ARF in the ICU may have not been included in the PICARD cohort. Cross-validation of the PICARD models will be required to assess their relative performance and generalizability. Many other variables not presented in the models reviewed here were collected in PICARD and examined for their relation with mortality. Selected data elements were missing in a majority of patients (e.g., central pressures from pulmonary artery catheters). Thus, many variables could not be included in population-wide predictive models because of power considerations and to abrogate bias. The models presented here may be the most generalizable of potential models, but other clinical and laboratory variables may be important in settings in which they are available. For example, impaired nutritional status is known to be associated with mortality and morbidity in acute and chronic kidney disease, and we were able to capture relatively little information on protein and energy intake or resting energy expenditure. Additional subgroup analyses (e.g., among patients with pulmonary artery catheters or indirect calorimetry) may be of interest, but are beyond the scope of this report.

In summary, we provide a series of three static models predicting mortality after ARF. These models can be used for risk adjustment when evaluating other patient characteristics or treatment strategies at specific key time points (e.g., upon initiation of dialysis). In addition, we provide an integrated model examining risk factors for mortality, which included

advanced age, sepsis, central nervous system, cardiovascular, hematological, and hepatic organ system failure as well as the severity of ARF (as indicated by the need for dialysis, or rising serum creatinine and reduced urine output). These findings should be validated in other large, diverse patient populations, in conjunction with the evaluation of new biomarkers, to help inform the design and implementation of randomized clinical trials in ARF.

MATERIALS AND METHODS

Study participants

The PICARD network is comprised of five academic medical centers in the United States: University of California San Diego (Coordinating Center), CCF, Maine Medical Center, Vanderbilt University, and University of California San Francisco). Over a 31-month period (February 1999 to August 2001), all patients consulted for ARF in the ICU were evaluated by PICARD study personnel for potential study participation. Given the large number of ICU beds at CCF, one in six ARF patients were randomized for possible study inclusion, to avoid single center over-representation. ARF was defined as an increase in serum creatinine $\geqslant 0.5$ mg/dl with baseline serum creatinine <1.5 mg/dl, or an increase in serum creatinine $\geqslant 1.0$ mg/dl with baseline serum creatinine $\geqslant 1.5$ mg/dl and <5.0 mg/dl. Patients with a baseline serum creatinine $\geqslant 5.0$ mg/dl were not considered for study inclusion. Baseline CKD was defined as an estimated glomerular filtration rate <30 ml/min/1.73 m^2 (corresponding to National Kidney Foundation Kidney Disease Quality Outcomes Initiative (K/DOQI) stage IV CKD). For each patient, the day on which ARF was identified and diagnosed by the above criteria was designated the ARF diagnosis date. The ARF diagnosis date may or may not have coincided with the nephrology consultation date, or with the date of first procedure (dialysis or hemodiafiltration) among individuals who required dialysis. Vital signs, hemodynamic, and laboratory data were recorded on or before the ARF diagnosis date, up to 3 days before nephrology consultation, and daily thereafter. Multiple generic and ARF-specific severity scores were calculated, also longitudinally. Organ failure and related parameters were defined using validated published criteria.[19] The timing, modality, and intensity of dialysis were determined by the treating physician with no influence from study personnel.

A detailed description of PICARD inclusion and exclusion criteria, data elements, data collection, and management strategies are described elsewhere.[28] Patients who were contacted by study personnel and who signed (or whose proxy signed) informed consents were enrolled in the study cohort. The reason for non-enrollment was determined for patients who did not sign informed consent,[29] although no additional data were collected for privacy considerations. The Committees on Human Research at each participating clinical site approved the study protocol and informed consent. In-hospital mortality within 60 days after ARF diagnosis was the principal outcome of interest.

Statistical analysis

Continuous variables were expressed as mean $\pm$ s.d. or median and compared using analysis of variance (general linear models with adjustment for multiple comparisons) or the Kruskal–Wallis test where appropriate. Categorical variables were expressed as proportions and compared with the Cochran–Mantel–Haenszel χ^2 test or Fisher's exact test. Logistic regression was employed to determine the odds of in-hospital mortality within 60 days after ARF diagnosis, incorporating exposure data on the day of ARF diagnosis, the day of nephrology consultation, and among patients requiring dialysis, the day of first procedure. Proportional hazards ('Cox') regression was used to determine the association of baseline exposures with time to death within 60 days. Finally, proportional hazards regression was extended to incorporate time-varying covariates, to take advantage of serially collected data, including vital signs, laboratory tests, organ system failure, and the provision of dialysis.

For continuous variables, values were initially categorized into quintiles to evaluate for linear vs non-linear associations with mortality. Where associations were not linear, clinically meaningful cutoffs were used to dichotomize the data, and indicator variables were included in regression models. Discrete data were coded as present or not present, and in time-dependent models, variables could change over time (e.g., with the development or recovery of organ system failure).

Multivariable logistic and proportional hazards regression models were constructed with backward variable selection, using $P<0.05$ for variable retention. Effect modification was evaluated by including multiplicative interaction terms for selected variables. Factors not included in multivariable models were re-entered individually to evaluate for residual confounding. For time-varying covariate models, we used the last value carried forward approach to handle missing data elements. For patients who required dialysis, moving average values for urine output, serum creatinine, and BUN were used to avoid confounding by the direct effects of dialysis on parameters of kidney function. Once dialysis was initiated, an 'on-dialysis' status was carried forward regardless of the frequency or intensity of dialysis therapy.

In logistic regression models, discrimination was assessed using the area under the ROC curve. Calibration was assessed using the Hosmer–Lemeshow goodness-of-fit test. The Hosmer–Lemeshow test compares model performance (observed vs expected) across deciles of risk, to test whether the model is biased (i.e., performs differentially at the extremes of risk). A nonsignificant value for the Hosmer–Lemeshow χ^2 suggests an absence of such bias. In proportional hazards regression models, the relative discrimination of alternative models was assessed using the likelihood ratio test. Plots of log (−log (survival rate)) against log (survival time) were performed to establish the validity of the proportionality assumption.

Two-tailed P-values <0.05 were considered significant. Statistical analyses were conducted using SAS 8.2 (SAS Institute, Cary, NC, USA).

ACKNOWLEDGMENTS

We are grateful to the study coordinators and other personnel who were critical to the PICARD project's implementation. These include: Tracy Seifert RN, Michelle Garcia RN, Lydia Sweeney RN, Tom Greene PhD, and Brett Larive MS (CCF), Stephanie Freedman RN, and Rebecca McClellan RN (Maine Medical Center), Pamela Kent RD, M. Tugrul Sezer, MD and Cathi Martin RD (Vanderbilt University), Carmencita Gruta RN, Maria Pascual, RN, and Rachel Manaster (University of California San Diego), and Susan Robertson NP and Jennifer Luan (University of California San Francisco). We appreciate the assistance of Professor David DeLong (Duke University) who provided guidance in the comparison of ROC curves. The study was supported by the following research grants: NIH-NIDDK RO1-DK53412, RO1-DK53411 and RO1-DK53413, and R33-DK67645.

REFERENCES

1. Silvester W, Bellomo R, Cole L. Epidemiology, management, and outcome of severe acute renal failure of critical illness in Australia. *Crit Care Med* 2001; **29**: 1910–1915.

2. Bates DW, Su L, Yu DT *et al.* Correlates of acute renal failure in patients receiving parenteral amphotericin B. *Kidney Int* 2001; **60**: 1452–1459.
3. Liaño F, Junco E, Pascual J *et al.* The spectrum of acute renal failure in the intensive care unit compared with that seen in other settings. The Madrid Acute Renal Failure Study Group. *Kidney Int Suppl* 1998; **66**: S16–S24.
4. van Bommel EF, Bouvy ND, Hop WC *et al.* Use of APACHE II classification to evaluate outcome and response to therapy in acute renal failure patients in a surgical intensive care unit. *Renal Fail* 1995; **17**: 731–742.
5. Shusterman N, Strom BL, Murray TG *et al.* Risk factors and outcome of hospital-acquired acute renal failure. Clinical epidemiologic study. *Am J Med* 1987; **83**: 65–71.
6. Schaefer JH, Jochimsen F, Keller F *et al.* Outcome prediction of acute renal failure in medical intensive care. *Intensive Care Med* 1991; **17**: 19–24.
7. Rasmussen HH, Pitt EA, Ibels LS, McNeil DR. Prediction of outcome in acute renal failure by discriminant analysis of clinical variables. *Arch Intern Med* 1985; **145**: 2015–2018.
8. Fiaccadori E, Maggiore U, Lombardi M *et al.* Predicting patient outcome from acute renal failure comparing three general severity of illness scoring systems. *Kidney Int* 2000; **58**: 283–292.
9. Halstenberg WK, Goormastic M, Paganini EP. Validity of four models for predicting outcome in critically ill acute renal failure patients. *Clin Nephrol* 1997; **47**: 81–86.
10. Douma CE, Redekop WK, van der Meulen JH *et al.* Predicting mortality in intensive care patients with acute renal failure treated with dialysis. *J Am Soc Nephrol* 1997; **8**: 111–117.
11. de Mendonca A, Vincent JL, Suter PM *et al.* Acute renal failure in the ICU: risk factors and outcome evaluated by the SOFA score. *Intens Care Med* 2000; **26**: 915–921.
12. Liaño F, Gallego A, Pascual J *et al.* Prognosis of acute tubular necrosis: an extended prospectively contrasted study. *Nephron* 1993; **63**: 21–31.
13. Liaño F, Garcia-Martin F, Gallego A *et al.* Easy and early prognosis in acute tubular necrosis: a forward analysis of 228 cases. *Nephron* 1989; **51**: 307–313.
14. Lins RL, Elseviers M, Daelemans R *et al.* Prognostic value of a new scoring system for hospital mortality in acute renal failure. *Clin Nephrol* 2000; **53**: 10–17.
15. Paganini EP, Halstenberg WK, Goormastic M. Risk modeling in acute renal failure requiring dialysis: the introduction of a new model. *Clin Nephrol* 1996; **46**: 206–211.
16. Chertow GM, Lazarus JM, Paganini EP *et al.* Predictors of mortality and the provision of dialysis in patients with acute tubular necrosis. The auriculin anaritide acute renal failure study group. *J Am Soc Nephrol* 1998; **9**: 692–698.
17. Clermont G, Acker CG, Angus DC *et al.* Renal failure in the ICU: comparison of the impact of acute renal failure and end-stage renal disease on ICU outcomes. *Kidney Int* 2002; **62**: 986–996.
18. Fortescue EB, Bates DW, Chertow GM. Predicting acute renal failure after coronary bypass surgery: cross-validation of two risk-stratification algorithms. *Kidney Int* 2000; **57**: 2594–2602.
19. Chang RW, Jacobs S, Lee B, Pace N. Predicting deaths among intensive care unit patients. *Crit Care Med* 1988; **16**: 34–42.
20. Mehta RL, Pascual MT, Gruta CG *et al.* Refining predictive models in critically ill patients with acute renal failure. *J Am Soc Nephrol* 2002; **13**: 1350–1357.
21. Paganini EP, Larive B, Kanagasundaram NS. Severity scores and outcomes with acute renal failure in the ICU setting. *Contrib Nephrol* 2001; **57**: 181–195.
22. Chen YC, Tsai MH, Hsu CW *et al.* Role of serum creatinine and prognostic scoring systems in assessing hospital mortality in critically ill cirrhotic patients with upper gastrointestinal bleeding. *J Nephrol* 2003; **16**: 558–565.
23. Mehta RL, McDonald B, Gabbai F *et al.* Nephrology consultation in acute renal failure: does timing matter? *Am J Med* 2002; **113**: 456–461.
24. El-Shahawy MA, Agbing LU, Badillo E. Severity of illness scores and the outcome of acute tubular necrosis. *Int Urol Nephrol* 2000; **32**: 185–191.
25. Levy EM, Viscoli CM, Horwitz RI. The effect of acute renal failure on mortality. A cohort analysis. *JAMA* 1996; **275**: 1489–1494.
26. Chertow GM, Levy EM, Hammermeister KE *et al.* Independent association between acute renal failure and mortality following cardiac surgery. *Am J Med* 1998; **104**: 343–348.
27. Bates DW, Su L, Yu DT *et al.* The mortality and costs of acute renal failure associated with amphotericin B therapy. *Clin Infect Dis* 2001; **32**: 686–693.
28. Mehta RL, Pascual MT, Soroko S *et al.* Spectrum of acute renal failure in the intensive care unit: the PICARD experience. *Kidney Int* 2004; **66**: 1613–1621.
29. Chertow GM, Pascual MT, Soroko S *et al.* Reasons for non-enrollment in a cohort study of ARF: the Program to Improve Care in Acute Renal Disease (PICARD) experience and implications for a clinical trials network. *Am J Kidney Dis* 2003; **42**: 507–512.