Kidney International, Vol. 58 (2000), pp. 283–292

CLINICAL NEPHROLOGY – EPIDEMIOLOGY – CLINICAL TRIALS

Predicting patient outcome from acute renal failure comparing three general severity of illness scoring systems

ENRICO FIACCADORI, UMBERTO MAGGIORE, MARIA LOMBARDI, SABINA LEONARDI, CARLO ROTELLI, and ALBERICO BORGHETTI

Reparto Acuti, Dipartimento di Clinica Medica, Nefrologia, and Scienze della Prevenzione, Universitá degli Studi di Parma, Parma, Italy

Predicting patient outcome from acute renal failure comparing three general severity of illness scoring systems.

Background. A major problem of studies on acute renal failure (ARF) arises from a lack of prognostic tools able to express the medical complexity of the syndrome adequately and to predict patient outcome accurately. Our study was thus aimed at evaluating the predictive ability of three general prognostic models [version II of the Acute Physiology and Chronic Health Evaluation (APACHE II), version II of the Simplified Acute Physiology Score (SAPS II), and version II of the Mortality Probability Model at 24 hours (MPM_{24} II)] in a prospective, single-center cohort of patients with ARF in an intermediate nephrology care unit.

Methods. Four hundred twenty-five patients consecutively admitted for ARF to the Nephrology and Internal Medicine Department over a five-year period were studied (272 males and 153 females, median age 71 years, interquartile range 61 to 78, median APACHE II score 23, interquartile range 18 to 28). Acute tubular necrosis (ATN) accounted for 68.7% (292 out of 425) of patients. Renal replacement therapies (hemodialysis or continuous hemofiltration) were used in 64% (272 out of 425) of ARF patients.

Results. Observed mortality was 39.1% (166 out of 425). The mean predicted mortality was 36.2% with APACHE II ($P = 0.571$ vs. observed mortality), 39.3% with SAPS II ($P = 0.232$), and 45.1% with MPM_{24} II ($P < 0.0001$). Lemeshow-Hosmer goodness-of-fit C and H statistics were 15.67 ($P = 0.047$) and 12.05 ($P = 0.15$) with APACHE II, 32.53 ($P = 0.0001$), 39.8 ($P = 0.0001$) with SAPS II, 21.86 ($P = 0.005$), and 20.24 ($P = 0.009$) with MPM_{24} II, respectively. Areas under the receiver operating characteristic (ROC) curve were 0.75, 0.77, and 0.85, respectively.

Conclusions. The APACHE II model was a slightly better calibrated predictor of group outcome in ARF patients, as compared with the SAPS II and MPM_{24} II outcome prediction models. The MPM_{24} II model showed the best discrimination capacity, in comparison with both APACHE II and SAPS II models, but it constantly and significantly overestimated mean predicted mortality in ARF patients. None of the models provided sufficient confidence for the prediction of outcome in individual patients. A high degree of caution must be exerted in the application of existing general prognostic models for outcome prediction in ARF patients.

Key words: scoring systems, acute tubular necrosis, critical care, mortality prediction, prognosis, severity of illness index.

Received for publication June 1, 1999
and in revised form January 13, 2000
Accepted for publication January 30, 2000

Acute renal failure (ARF) is frequently observed in hospitalized patients and is increasingly prevalent in aging populations with multiple chronic comorbidities, in subjects with severe underlying diseases such as neoplasia or AIDS, and in the intensive care unit (ICU) populations [1].

Acute renal failure is associated with a high mortality, and despite the considerable therapeutic progress and technological improvement observed over the past few years, the prognosis of the syndrome remains poor [2–6].

A major problem of studies on ARF arises from a lack of prognostic tools able to adequately express the medical complexity of the syndrome [1]. These tools would be of the utmost importance for an objective risk stratification in interventional studies, quality of care evaluation, and allocation of health care resources. Moreover, no accurate mortality risk estimates are available to the nephrologist as decision-making support in order to avoid unethical and futile care of ARF patients [7, 8].

Acute renal failure-specific prediction models developed in the past were usually derived from single institutions and more often have been validated in specific subsets of ARF patients, such as ARF in the ICU, ARF on hemodialysis or on continuous hemofiltration, ARF following aortic aneurysm or heart surgery, and ARF exclusively caused by acute tubular necrosis (ATN) [9–19]. The generalizability of these ARF- or ATN-specific models to the composite ARF patient cohorts commonly followed in nephrology units or in medical wards is therefore questionable.

General outcome prediction models are widely used in the ICUs [20], but limited information is currently available concerning their application to ARF patients.

Data, in fact, are usually retrospective [21], and in most cases, only first- or second-generation models have been used [12, 16, 22, 23].

Our study was aimed at comparing the ability of three general prognostic models to predict mortality in ARF patients hospitalized in an intermediate nephrology care unit.

To this purpose, we applied three different models to a prospective cohort of ARF patients: two recently developed, easily available, third-generation, prognostic models based on an international cohort of critically ill patients [version II of the Mortality Probability Model (MPM II), version II of the Simplified Acute Physiology Score (SAPS II)], and a standard second-generation model, the Acute Physiology and Chronic Health Evaluation II (APACHE II)] [24–26].

METHODS

Study population

We studied all adults patients admitted with a diagnosis of ARF (community or hospital acquired ARF) to the intermediate nephrology care unit of the Parma Hospital during a five-year period (January 1994 through December 1998). ARF was defined as an abrupt decline in renal function [27] with a recent rise (24 to 48 h) in the plasma creatinine values of more than 50% above the baseline status in the absence of volume-responsive prerenal status; in the case of pre-existing renal disease or known renal insufficiency that had not been dialyzed before and was not considered to be end-stage renal disease (ESRD), patients were required to demonstrate an increase in serum creatinine levels >1 mg/dL from their baseline status (acute-on-chronic renal failure) [28]. The ATN diagnosis was made when no improvement in renal function was observed after correction of possible prerenal causes, as well as by urinary sediment examination, sodium fractional excretion, and renal ultrasonography. In the other ARF patients, etiologic diagnosis was made by angiography or angio-computed tomography scan or by renal biopsy.

The intermediate nephrology care unit is a closed, six-bed specialty unit for ARF patients that is part of the Internal Medicine and Nephrology Department. It has a bed-to-nurse ratio of 3:1 and is staffed by nephrologists with advanced experience in critical care nephrology. The same nephrologists are also responsible for all consultations in other wards and ICUs in cases of possible ARF, as well as for the renal replacement therapies (RRTs) applied to patients hospitalized in the hospital ICUs. All patients hospitalized with ARF were referred to the unit if they were not on mechanical ventilation at the time of the possible referral. Thus, the only exclusion criterion from the study was ARF on mechanical ventilation during the entire course of hospital stay.

As medical care in Italy is covered by the National Health System, there was no access restriction to treatment. Therapeutic agents, as well as RRTs (particularly for what concurs, such as time of starting, modality, dose and filter membranes) were administered at the discretion of the treating nephrologist. The standard hemodialysis schedule was four hours every other day. Conventional bicarbonate hemodialysis was performed through central venous access by double or triple lumen catheters (internal jugular or subclavian veins) [29], with an ultrafiltration-controlled delivery system, polymethylmetacrylate, ethylvinylalcool or polysulfone hollow fiber filters, and heparin as the anticoagulant. The heparin-free hemodialysis method was used for patients at hemorrhagic risk. Continuous RRTs (CRRTs) were performed as continuous venovenous hemofiltration with prostacyclin as the circuit anti-aggregant and with the same filters as for hemodialysis. Indications for CRRT were represented in most cases by hypotension and/or severe hemodynamic instability and/or severe fluid overload.

Measurements

Data collection. Data collection for each patient included all of the demographic and clinical variables necessary to the outcome prediction models considered, and it was complete in 425 patients for APACHE II and SAPS II and in 410 patients for MPM_{24} II. Survival status was registered up to the hospital discharge.

Predictive models. Predicted risk of death was calculated for APACHE II, SAPS II, and MPM_{24} II using published coefficients and equations on the basis of data obtained during the first 24 hours after admission.

The APACHE scoring system in the version II model consists of three parts: Acute Physiology Score (12 physiologic variables measured within 24 hours of admission, maximum 60 points), Chronic Health Score (premorbid major organ dysfunction, maximum 5 points), and Age Score (maximum 6 points) [26]. The range of the APACHE II score is from 0 to 71. Hospital mortality can be predicted in an equation that includes the APACHE II score, whether the patient had received emergency surgery, and the diagnostic category of the patient.

The SAPS scoring system in the version II model derives the score from 12 physiological variables, age, type of admission (scheduled surgical, unscheduled surgical, or medical), and three underlying disease variables (AIDS, metastatic cancer, and hematologic malignancy) [25]. The resulting SAPS II score is then entered into a published mathematical formula in which the solution gives the numerical value of the predicted hospital mortality.

The MPM version II model uses physiologic variables, chronic diagnosis, acute diagnosis, type of admission, age, use of cardiopulmonary resuscitation, and use of mechanical ventilation to estimate the probability of patient

mortality at sequential time points: admission (MPM_0 II), 24 (MPM_{24} II) [24], 48 (MPM_{48} II), and 72 (MPM_{72} II) hours after admission [30]. The MPM_{24} II model can be compared with both SAPS II and APACHE II. In fact, all three models determine their scores by using the worst level for physiologic variable values measured during the first 24 hours after admission.

Accuracy of outcome prediction. The accuracy of outcome prediction was assessed in terms of discrimination and calibration, according to recent recommendations [20]. Calibration, that is, the accuracy of the risk predictions by the model or ability of a model to describe the mortality pattern in the population, was evaluated by the difference between observed and expected mortality, by calibration curves, and by the Lemeshow-Hosmer C and H statistics [31, 32]. Observed and predicted mortality probabilities were compared by the Wilcoxon signed rank test. Calibration curves are based on linear regression analysis, in which the observed death rates are plotted against predicted death rates stratified by 10% predicted risk ranges. The R^2 value represents the proportion of variation of the dependent variable (observed death rate) that is predicted from the independent variable (predicted death rate). An R^2 value of 1.0 indicates that all plotted points lie on a straight line and that the independent variable is able to predict the dependent variable with 100% certainty. If the predictive model fits the study data well (that is, it calibrates), the plotted points (the predicted and observed death rates) will lie approximately on a 45° line (slope = 1), with intercept = 0. An upward shift of the line implies that the predictive model underestimates the actual rates, while a downward shift represents an overestimation of actual mortality. Chi-square C and H statistics of Lemeshow and Hosmer were also calculated [31, 32]. These tests evaluated the degree of correspondence between a model's estimate probabilities of mortality and the actual mortality experience of patients over groups spanning the entire range of probabilities. The test compared the predicted number of patients who died and who survived with the actual number of patients who died and who survived, within 10 groups defined by the predicted risk of mortality. Patients were stratified by deciles of the estimated probabilities for the *C*-test and in decines for the *H*-test. A high *P* value obtained in the test suggested a good calibration, and a small *P* indicated a poor calibration of the model.

Discrimination, that is, the ability to discriminate between patients who live and those who die, was evaluated by the receiver operating characteristic (ROC) curve and the area under the ROC curve [33]. The area under the ROC curve is a measure of the overall discriminatory power of the prognostic variable, and it measures the probability that in randomly paired dead and surviving patients, the predicted probability of the model will be correctly ranked. The curve is constructed by varying the cut point used to determine the values of the predicted probability to be considered correctly classified, and then plotting the resulting true positive and false positive rates. If a model could perfectly discriminate, the curve would pass through the point (0, 1) on the grid unit. Thus, the closer an ROC curve comes to this ideal point, the better is its discriminant ability. A model with no discriminant ability produces a curve that follows the diagonal of the grid. As a rule, the bigger the area, the better the discriminatory capacity of the model: it is generally thought that an area under the ROC curve of $\geq$0.7 is acceptable, while $\geq$0.8 is good, and $\geq$0.9 is excellent [20]. The areas under the ROC curve and its confidence limits were obtained by the Mann–Whitney two-sample U statistic and its standard error [33]. The method proposed by DeLong, DeLong, and Clarke-Pearson was employed to test whether the difference between the ROC areas was statistically significant [34]. Data were recorded at admission on preprinted forms by experienced medical personnel involved in the study. The Access '97 database (Microsoft, Seattle, WA, USA) was used for final data recording and processing. Data are presented as median and interquartile range, unless otherwise specified. Data analysis and statistics were performed by the Prism software, version 2.01 (GraphPad Software Inc., San Diego, CA, USA), and the STATA statistical software, release 5.0 (StataCorp., College Station, TX, USA).

RESULTS

During the study period, ARF diagnosis was made in 461 patients in the hospital. As the study period was a total of five years and our hospital was the ARF referral center for an average population of 450,000 inhabitants, the estimated incidence rate of ARF was 20.48 per 100,000 person-years (95% CI, 18.6 to 22.4). Of the 461 patients with a diagnosis of ARF during the study period, 36 were not included in the study. Twenty-five patients were excluded as they needed mechanical ventilation in the hospital ICUs during the entire course of their ARF or until death. This subgroup of ARF patients was followed by intensivists with the nephrologists as consultants. Eleven patients with ARF who were not on mechanical ventilation could not be transferred to our unit for several reasons (death before referral in 3 cases, bed not available in 2 cases, follow-up in other wards in 5 cases). The study population thus consisted of 425 subjects with ARF of different etiologies who represented 92% (425 out of 461) of all patients hospitalized for ARF during the study period and 97% of the eligible population (425 out of 436 ARF patients not on mechanical ventilation at the time of possible referral to our unit).

Table 1. Etiology of acute renal failure

Acute renal failure	$N = 425$
Acute tubular necrosis	292 (69%)
Glomerulonephritis	17 (4%)
Vasculitis	10 (2.3%)
Acute interstitial nephritis	17 (4%)
Multiple myeloma	17 (4%)
Vascular[a]	30 (7.1%)
Obstructive	42 (9.9%)

[a]Includes atheroembolic disease, renal artery and/or aortic thrombosis or dissection, renal vein thrombosis

Table 2. Demographic and clinical data of the acute renal failure cohort and of the acute tubular necrosis subgroup

	Acute renal failure ($N = 425$)	Acute tubular necrosis ($N = 292$)
Age *years*	71 (61–78)	72 (62–80)
Sex	253 M (65%)	186 M (67%)
	137 F (35%)	96 F (33%)
Location		
Emergency room	18 (4.2%)	12 (4.1%)
Medical wards	260 (61.2%)	169 (57.9%)
Surgical wards	70 (16.5%)	49 (16.8%)
Surgical ICUs[a]	77 (18.1%)	62 (21.1%)
Serum creatinine *mg/dL*	5.3 (3.3–7.1)	4.6 (3.1–6.3)
Blood urea nitrogen *mg/dL*	81 (60–115)	81 (59–115)
APACHE II Score	23 (18–28)	24 (19–31)
Renal replac. therapies	272/425 (64.1%)	189/292 (64.7%)
Hemodialysis only	215/425 (50.6%)	140/292 (47.9%)
CVVH only[b]	44/425 (10.3%)	38/292 (13%)
CVVH + hemodialysis	13/425 (3.1%)	11/292 (3.8%)

Continuous variables are reported as median (interquartile range).
[a]Includes trauma ICU, postoperative ICU, and heart surgery ICU
[b]Continuous venovenous hemofiltration

Table 3. Observed and predicted mortality in the acure renal failure cohort and in the acute tubular necrosis subgroup

	Acute renal failure ($N = 425$)	Acute tubular necrosis ($N = 292$)
Observed mortality	39.1% (166/425)	44.2% (129/292)
Predicted mortality		
APACHE II	36.2%	40.3%
SAPS II	39.3%	43.1%
MPM24II	45.1%[b]	48.4%[a]

[a]$P < 0.001$, [b]$P < 0.0001$ vs. observed mortality, Wilcoxon signed rank test

The etiology of ARF is shown in Table 1. In most cases (292 patients out of 425, 69%), it was represented by ATN. The patients' demographic and clinical data are illustrated in Table 2. Data concerning the ATN subgroup are presented separately. Mortality data (observed vs. predicted mortality) for the ARF cohort and the ATN subgroup are shown in Table 3. MPM_{24} II significantly overestimated the mean predicted mortality in both the ARF cohort and the ATN subgroup. Calibration curves for the three models are illustrated in Figure 1. The curves demonstrated that, in general, the proportion of patients who died increased in accordance with the increase in risk of in-hospital mortality predicted by the three prognostic models. In the case of the ARF cohort, the observed mortality for MPM_{24} II was significantly below the diagonal identity line (predicted = observed) for a large portion of the range of stratified predicted risk. That is, the model significantly overestimated mortality over the entire range of patient risk. The APACHE II calibration curve was closer to the line of perfect predicting ability. For the ARF cohort, Lemeshow-Hosmer C and H statistics are reported in Table 4. ROCs are shown in Figure 2. In the ARF patient, cohort areas under the ROC curve were 0.75, 0.77, and 0.85, respectively ($P = 0.0003$ vs. APACHE II, $P = 0.00008$ vs. SAPS II). In the ATN subgroup, 0.76, 0.76, and 0.84 ($P = 0.01$ vs. APACHE II, $P = 0.008$ vs. SAPS II). Table 5 and Table 6 show the predictive performance of the three prognostic models as applied to the subgroup of ARF patients not transferred to our unit from an ICU setting (348 subjects, 216 males, 132 females, median age 72 years, interquartile range 62.5 to 79.0; median APACHE II score 24, interquartile range 18 to 29; serum creatinine 5.6 mg/dL, interquartile range 3.5 to 7.4; blood urea nitrogen 82.2, interquartile range 60.1 to 118.4), and to the subgroup of ARF patients who needed RRT (272 subjects, 163 males, 109 females, median age 72 years, interquartile range 62.5 to 77.5; median APACHE II score 25, interquartile range 20 to 30; serum creatinine 5.9, interquartile range 4.2 to 7.9; blood urea nitrogen 93.5, interquartile range 68 to 127). Again, the mean predicted mortality was significantly overestimated by MPM_{24} II. APACHE II was less well calibrated in the two subgroups of ARF patients than it was in the entire ARF cohort or in the ATN subgroup, while calibration was slightly better for SAPS II and for MPM_{24} II (in the subgroup of ARF patients who needed RRT), particularly if the C statistic was considered. Discrimination capacity of the three models as applied in the two subgroup of ARF patients was not improved, as compared with that documented in both the entire cohort of ARF patients and in the ATN subgroup. MPM_{24} II had the

Fig. 1. Calibration curves for (*A*) APACHE II (△), (*B*) SAPS II (○), and (*C*) MPM_{24} II (◇), constructed by plotting observed death rate against predicted death rate stratified by 10% risk ranges. For any decision criterion, the true positive rate is the percentage of patients predicted to die from those who actually died. The false positive rate is the percentage of patients predicted to die from those who actually survived. ARF cohort data are in the left column, ATN subgroup in the right column.

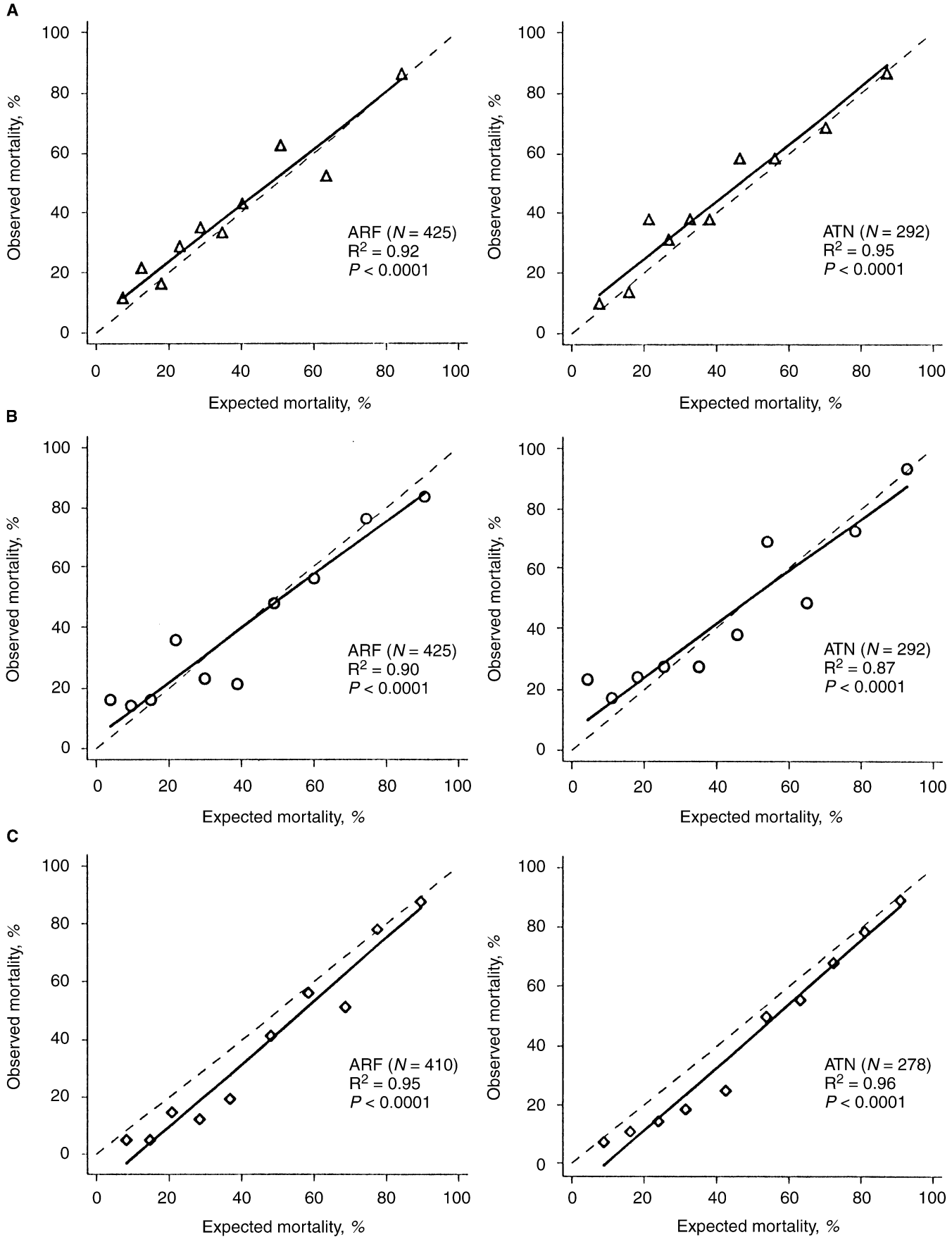
A
Observed mortality, %
Expected mortality, %
ARF (N = 425)
R2 = 0.92
P < 0.0001
ATN (N = 292)
R2 = 0.95
P < 0.0001
B
ARF (N = 425)
R2 = 0.90
P < 0.0001
ATN (N = 292)
R2 = 0.87
P < 0.0001
C
ARF (N = 410)
R2 = 0.95
P < 0.0001
ATN (N = 278)
R2 = 0.96
P < 0.0001

Table 4. Calibration indices (Lemeshow-Hosmer C and H statistics) of the ARF patient cohort and the ATN patient subgroup

	ARF Cohort (N = 425)		ATN Subgroup (N = 292)	
	C test (p value)	H test (P Value)	C test (P value)	H test (P value)
APACHE II	15.67 (0.047)	12.05 (0.15)	4.13 (0.84)	9.66 (0.29)
SAPS II	32.53 (0.0001)	39.8 (0.0001)	36.71 (0.0001)	31.2 (0.0001)
MPM_{24} II	21.86 (0.005)	20.24 (0.009)	9.6 (0.29)	11.56 (0.17)

Table 5. Observed and predicted mortality in the ARF subgroup of patients not referred from an ICU setting, and in the ARF subgroup of patients who needed renal replacement therapies

	ARF patients not referred from an ICU setting (N = 348)	ARF patients who needed renal replacement therapy (N = 272)
Observed mortality	40.8% (142/348)	49.3% (134/272)
Predicted mortality		
APACHE II	36.4%	40.5%[b]
SAPS II	42.3%	47.1%
MPM_{24} II	47.3%[c]	53.6%[a]

[a]$P < 0.05$, [b]$P < 0.01$, [c]$P < 0.001$ vs. observed mortality, Wilcoxon signed rank test

higher ROC area value, that is, a better discriminating capacity.

DISCUSSION

In this study, to our knowledge for the first time, the predictive accuracy of two third-generation models for outcome prediction applied to a representative and consistent population of ARF patients was prospectively compared with that of the currently most widely used model, APACHE II. Moreover, our study allowed the performance of those same models to be analyzed in clinically relevant subgroups of patients with ARF.

Our study population can be considered representative of the ARF patient populations not on mechanical ventilation who are usually followed in nephrology units. The incidence rate data for ARF in the present study (20.49 per 100,000 person-years; 95% CI, 18.6 to 22.4) are quite similar and are even higher compared with those presented in other studies from Europe: 17.2 per 100,000 person-years [35], 18.5 per 100,000 person-years (abstract; McGregor et al, XXIX Congr EDTA-ERA, 1992, p 54), and particularly to the incidence rate of 16.3 per 100,000 person-years documented in a recent study from Spain [36]. The latter study was based on similar diagnostic criteria for ARF and was done within a similar health system. Moreover, since the vast majority of eligible patients were included in the study and since demographic and clinical characteristics of patients eligible but not included in our study were not different, it is very unlikely that the selection of the patients could have seriously biased our results.

The overall predictive ability of the models we have considered in the present study was not very satisfying, and moreover, it was not uniform over the several subgroups of ARF patients considered. In fact, while APACHE II had an acceptable calibration in the entire ARF patient cohort and in the ATN subgroup, the model-predictive capacity worsened when it was applied to the subgroup of patients who needed RRT. In most cases, SAPS II did not calibrate nor discriminate well, even despite the fact that the predicted mortality by the model was the closest to the observed value. Finally, despite having an acceptable receiver ROC area, MPM_{24} II did not calibrate well, as it constantly and significantly overestimated the in-hospital mortality rate in the ARF patient cohort and in the other subgroups of ARF patients.

Many factors could have introduced limitations in the applicability of general models to ARF patients considered in the present study. These factors are linked to both the model's characteristics and particularity of the syndrome.

First, there are well-known factors not accounted for in the prognostic models that can have some impact on patient's outcome even in general ICUs, such as number and type of comorbidities, disparities in technical and therapeutic resources available in the ICU, the type of ICU (open vs. closed), organization (bed-to-nurse ratio), and differences in medical staffing level [20].

Second, as many patients in our study (77 out of 425) were referred to our nephrology unit from an ICU setting, a possible limitation could result from a failure in applying the scoring systems to these patients in the first 24 hours after hospitalization in the ICU. Thus, at least from a theoretical point of view, in patients transferred from an ICU setting, the calculations should have been done on the basis of data collected on the day of admission to that ICU rather than to the nephrology unit. In fact, the predictive models we have used were designed for an ICU population with measurements to be derived during the first 24 hours of admission to the ICU. However, the aim of this study was to demonstrate the usefulness of the scoring systems at the time the nephrologist sees ARF patients in the nephrology ward, and at that point must decide, for instance, whether to start the

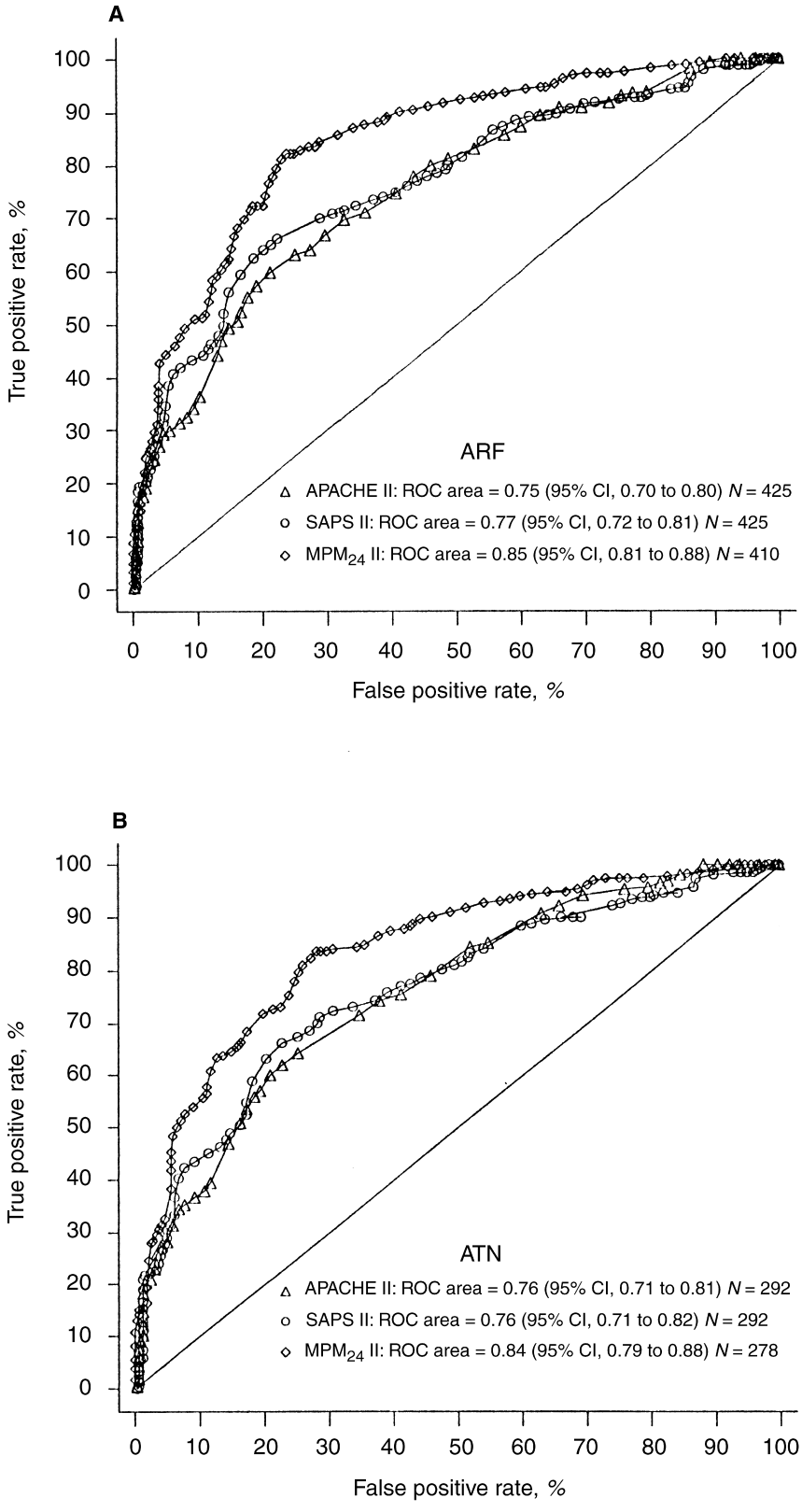

Fig. 2. Receiver operator characteristic (ROC) curves and areas for APACHE II (▲), SAPS II (○), and MPM_{24} II (◇). (*A*) Acute renal failure (ARF) cohort data. APACHE II ROC area = 0.75 (95% CI, 0.70 to 0.80), N = 425; SAPS II ROC area = 0.77 (95% CI, 0.72 to 0.81), N = 425; MPM_{24} II ROC area = 0.85 (95% CI, 0.81 to 0.88), N = 410. (*B*) Acute tubular necrosis (ATN) subgroup cohort data. APACHE II ROC area = 0.76 (95% CI, 0.71 to 0.81), N = 292; SAPS II ROC area = 0.76 (95% CI, 0.71 to 0.82), N = 292; MPM_{24} II ROC area = 0.84 (95% CI, 0.79 to 0.88), N = 278.

Table 6. Calibration and discrimination indexes in the ARF subgroup of patients not referred from an ICU setting, and in the ARF subgroup of patients who needed renal replacement therapies

	ARF patients not referred from an ICU setting (N = 348)			ARF patients who needed renal replacement therapies (N = 272)		
	C test (P value)	H test (P value)	ROC area (95% C.I.)	C test (P value)	H test (P value)	ROC area (95% C.I.)
APACHE II	9.66 (0.29)	17.15 (0.03)	0.75 (0.70–0.80)	14.49 (0.07)	21.55 (0.006)	0.72 (0.66–0.78)
SAPS II	6.2 (0.62)	33.14 (0.0001)	0.75 (0.70–0.81)	5.28 (0.73)	43.35 (0.0001)	0.73 (0.67–0.78)
MPM_{24} II	14.97 (0.059)	19.42 (0.013)	0.85 (0.80–0.89)	7.12 (0.52)	11.76 (0.16)	0.82 (0.78–0.87)

patient on RRT. Under such clinical circumstances, mortality probability obtained on the day of admission to the ICU rather than to the nephrology unit (which in some cases is many days before ARF diagnosis) indeed should not be very useful. However, in our study, no relevant improvement was obtained, especially in the discrimination capacity, when the models were applied to the ARF patient subgroup for which they were specifically designed (that is, after patients transferred from an ICU setting were excluded). Thus, it is unlikely that our results are seriously biased because of patient transferal from the hospital ICUs to our unit.

Third, the prognostic models applied by us were originally developed for use in the ICUs, so they should not be used outside of the ICUs. However, our ARF patients had a median APACHE II score of 23 (24 in the ATN subgroup), so they can be considered a group of critically ill patients independently from the localization in the hospital. Moreover, there is evidence in the literature that prognostic models are reliable instruments even in intermediate-care unit settings [37].

Fourth, prognostic systems considered in the present study were developed in general ICUs, where the percentage of patients with ARF is relatively low in comparison with the entire cohort considered for development and validation. The concept of a critical percentage for each model variable or condition, above which the predictive model deteriorates, was recently stressed in an elegant model simulation with the MPM II system [38].

Fifth, it is well known that a possible cause of poor performance of a general predictive model can be represented by the presence of subgroups of patients in the target population for whom the model does not perform well [39, 40]. This can be very relevant when the model is applied in ICUs that have unique patient characteristics, such as in specialty ICUs.

Sixth, the presumed cause of ARF in the original development cohorts from the two third-generation models was ATN, which usually represents more than 90% of parenchymal ARF in the ICUs, but not in other clinical settings, such as in nephrology wards. Anyway, in our study, in which ATN represented 69% of the cause of ARF, no major differences were observed in calibration or discrimination of the models when the ATN subgroup was considered separately. Similar results were obtained even in another very important subgroup of ARF patients, that is, those who needed RRT.

Finally, other important factors known to significantly influence ARF patient outcome could have impaired the performance of the models considered, such as: timing of ARF, time to diagnosis of ARF, time of referral or lead-time bias, true differences in case mix, pre-existing nutritional status, systematic differences in the effectiveness of treatment, and, finally, modality, dose, and membrane used for RRT [1, 41, 42].

In our study, the best compromise between calibration and discrimination in the ARF cohort and in the ATN subgroup was obtained with the APACHE II model.

The APACHE II scoring system was developed in 1985 and is based on data from over 5000 patients [26]. This model is an accurate measure of a patient's severity of illness. It strongly correlates with patients' outcome and has been validated in several countries. Moreover, it has been extensively used in the analysis of ARF patients' prognosis or in the assessment of severity of illness in this clinical condition [12, 17, 19, 21, 23, 43–45]. However, utilization of the APACHE II model in ARF patients has several shortcomings, which have been thoroughly reviewed elsewhere [3].

When there is evidence that a given general model is not fully appropriate for outcome prediction in a particular clinical setting, apart from the development of new models or customization of the existing general models [38, 39, 46], a possible solution can be represented by the use of specialized models. Currently available ARF- or ATN-specific models have usually been derived from single institutions and have been more often used in specific subsets of ARF patients [9–19], but their validation in ARF cohorts from other institutions has given controversial results (abstracts; Fernandez N et al, *J Am Soc Nephrol* 10:A720, 1999; Fiaccadori E et al, *J Am Soc Nephrol* 10:A724, 1999).

Thus, generalizability of these models to the composite ARF patient cohorts commonly followed in nephrology units or in medical wards is still open to discussion.

In conclusion, we have demonstrated that the APACHE II model was a slightly better calibrated predictor of group outcome in both the ARF patient cohort

and in the ATN subgroup, but this model did not provide sufficient confidence for prediction of outcome in individual patients. MPM_{24} II model showed the best discrimination capacity, as compared with both APACHE II and SAPS II models, but it constantly and significantly overestimated the mean predicted mortality in ARF patients.

We truly need accurate prognostic models in critically ill and particularly in ARF patients, because "human's ability to make assessment of uncertainty in complicated situations is poor" [47]. However, it should be equally stressed that, as recommended in a recent Consensus Conference [48], a high degree of caution is needed in the utilization of prognostic models for clinical purposes. The severity of illness scores, in fact, are not appropriate for routine use in clinical decision making nor for making triage decisions. Moreover, they do not take into account other important factors than mortality, for example, morbidity, disability, and quality of life after discharge.

ACKNOWLEDGMENTS

This work was supported in part by a grant from Parma University (Progetti Scientifici di Ateneo, 1997). It was presented in part at the 31st Annual Meeting, American Society of Nephrology, October 26–29, 1998 (*J Am Soc Nephrol* 9:A670, 1998), and at the European Society of Intensive Care Medicine, Stockholm, September 6–9, 1998 [*Intensive Care Med* 24(Suppl 1):S68, 1998]. This study would not have been possible without the invaluable contribution and commitment of the Nurses of the Nephrology Intermediate Care Unit. The authors thank Mrs. Nancy Birch-Podini for reviewing the manuscript.

Reprint requests to Enrico Fiaccadori, M.D., Ph.D., Dipartimento di Clinica Medica, Nefrologia and Scienze Della Prevenzione, Universitá degli Studi di Parma, Via Gramsci 14-43100 Parma, Parma, Italy.
E-mail: fiaccado@ipruniv.cce.unipr.it

REFERENCES

1. Nolan CR, Anderson RJ: Hospital-acquired acute renal failure. *J Am Soc Nephrol* 710–718, 1998
2. Druml W: Prognosis of acute renal failure 1975–1995. *Nephron* 73:8–15, 1996
3. McCarthy JT: Prognosis of patients with acute renal failure in the intensive care unit: A tale of two eras. *Mayo Clin Proc* 71:117–126, 1996
4. Levy EM, Viscoli CM, Horwitz RI: The effect of acute renal failure on mortality. *JAMA* 275:1489–1494, 1996
5. Mangano Mora C, Diamondstone LS, Ramsay JG, Aggarwal A, Herskowitz A, Mangano DT, Multicenter Study of Perioperative Ischemia Research Group: Renal dysfunction after myocardial revascularization: Risk factors, adverse outcomes, and hospital resource utilization. *Ann Int Med* 128:194–203, 1998
6. Brivet FG, Kleinknecht DJ, Loirat P, Landais PJM, French Study Group on Acute Renal Failure: Acute renal failure in intensive care units: Causes, outcome and prognostic factors of hospital mortality: A prospective, multicenter study. *Crit Care Med* 24:192–198, 1996
7. Hamel MB, Phillips RS, Davis RS, Desbiens N, Connors AF, Teno JM, Wenger N, Lynn J, Wu AW, Fulkerson W, Tsevat J, SUPPORT Investigators: Outcomes and cost-effectiveness of initiating dialysis and continuing aggressive care in seriously ill hospitalized patients. *Ann Int Med* 127:195–202, 1997
8. MacKay K, Moss AH: To dialyze or not dialyze: An ethical and evidence-based approach to the patient with acute renal failure in the intensive care unit. *Adv Ren Replace Ther* 4:288–296, 1997
9. Rasmussen H, Pitt E, Ibels L, McNeil D: Prediction of outcome in acute renal failure by discriminant analysis of clinical variables. *Arch Intern Med* 145:2015–2018, 1985
10. Lohr J, MacFarlane M, Grantham J: A clinical index to predict survival in acute renal failure patients requiring dialysis. *Am J Kidney Dis* 11:254–259, 1988
11. Berisa F, Beaman M, Adu D, McGonigle RJ, Michael J, Downing R, Fielding JW, Dunn J: Prognostic factors in acute renal failure following aortic aneurysm surgery. *Q J Med* 76:689–698, 1990
12. Schaefer J, Jochimsen F, Keller F, Wegscheider K, Distler A: Outcome prediction in acute renal failure in medical intensive care. *Intensive Care Med* 17:19–24, 1991
13. Groeneweld A, Tran D, van der Meulen J, Nauta J, Thijs L: Acute renal failure in the medical intensive care unit: Predisposing, complicating factors and outcome. *Nephron* 59:602–610, 1991
14. Barton I, Hilton P, Taub N, Warburton FG, Swan AV, Dwight J, Mason JC: Acute renal failure treated by haemofiltration: Factors affecting outcome. *Q J Med* 86:81–90, 1993
15. Zanardo G, Michielon P, Paccagnella A, Rosi P, Caló M, Salandin V, Da Ros A, Michieletto F, Simini G: Acute renal failure in the patient undergoing cardiac operation: prevalence, mortality rate, and main risk factors. *J Thorac Cardiovasc Surg* 107:1489–1495, 1994
16. Chertow GM, Christiansen CL, Cleary PD, Munro C, Lazarus M: Prognostic stratification in critically ill patients with acute renal failure requiring dialysis. *Arch Intern Med* 1505–1511, 1995
17. Neveu H, Kleinknecht D, Brivet F, Loirat P, Landais P, French Study Group on Acute Renal Failure: Prognostic factors in acute renal failure due to sepsis: Results of a prospective multicentre study. *Nephrol Dial Transplant* 11:293–299, 1996
18. Paganini EP, Haltenberg WK, Goormastic M: Risk modeling in acute renal failure requiring dialysis: The introduction of a new model. *Clin Nephrol* 46:206–211, 1996
19. Chertow GM, Lazarus JM, Paganini EP, Allgren RL, Lafayette RA, Sayegh MH, Auriculin Anaritide Acute Renal Failure Study Group: Predictors of mortality and the provision of dialysis in patients with acute tubular necrosis. *J Am Soc Nephrol* 9:692–698, 1998
20. Lemeshow S, Le Gall R: Modeling the severity of illness of ICU patients: A systems update. *JAMA* 272:1049–1055, 1994
21. Douma CCE, Redekop WK, Van der Meulen JHP, Van Olden RF, Haeck J, Struijk DG, Krediet RT: Predicting mortality in intensive care patients with acute renal failure treated with dialysis. *J Am Soc Nephrol* 8:111–117, 1997
22. Maher E, Robinson K, Scoble J, Farrimond JG, Browne DRG, Sweny P, Moorhead JF: Prognosis of critically ill patients with acute renal failure: APACHE II score and other predictive factors. *Q J Med* 72:857–866, 1989
23. Parker RA, Himmelfarb J, Tolkoff-Rubin N, Chandran P, Wingard RL, Hakim RM: Prognosis of patients with acute renal failure (ARF) requiring dialysis: Results of a multicenter study. *Am J Kidney Dis* 32:432–443, 1998
24. Lemeshow S, Teres D, Klar J, Spitz Avrunin J, Gehlbach SH, Rapoport J: Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 270:2478–2486, 1993
25. Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 270:2957–2963, 1993
26. Knaus WA, Draper EA, Wagner DP, Zimmerman JE: APACHE II: A severity of disease classification system. *Crit Care Med* 13:818–829, 1985
27. Brady HR, Brenner BM, Lieberthal W: Acute renal failure, in *The Kidney* (5th ed), edited by Brenner BM, Philadelphia, W.B. Saunders, 1996, pp 1200–1252
28. Rose BD: Acute renal failure: Prerenal versus acute tubular necrosis, in *Pathophysiology of Renal Disease* (2nd ed), edited by Rose BD, New York, McGraw-Hill, 1987, pp 65–119
29. Fiaccadori E, Gonzi GL, Zambrelli P, Tortorella G: Cardiac arrhythmias during central venous catheter procedures in acute renal failure: A prospective study. *J Am Soc Nephrol* 7:1079–1084, 1996

30. Lemeshow S, Klar J, Teres D, Avrunin JS, Gehlbach SH, Rapoport J, Rue M: Mortality probability models for patients in intensive care unit for 48 or 72 hours: A prospective, multicenter study. *Crit Care Med* 22:1351–1358, 1994
31. Hosmer DW, Lemeshow S: *Applied Logistic Regression.* New York, John Wiley & Sons, 1989
32. Lemeshow S, Hosmer DW: A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 115:92–106, 1982
33. Hanley J, McNeil B: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36, 1982
34. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 44:837–845, 1988
35. Feest TG, Round A, Hamad S: Incidence of severe acute renal failure in adults: Results of a community based study. *Br Med J* 306:481–483, 1993
36. Llano F, Pascual J, Madrid Acute Renal Failure Study Group: Epidemiology of acute renal failure: A prospective, multicenter, community-based study. *Kidney Int* 50:811–818, 1996
37. Auriant I, Vinatier I, Thaler F, Tourneur M, Loirat P: Simplified acute physiology score II for measuring severity of illness in intermediate care units. *Crit Care Med* 26:1368–1371, 1998
38. Murphy-Filkins RL, Teres D, Lemeshow S, Hosmer DW: Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: How to distinguish a general from a specialty intensive care unit. *Crit Care Med* 24:1968–1973, 1996
39. Moreno R, Apolone G: Impact of different customization strategies on the performance of a general severity score. *Crit Care Med* 25:2001–2008, 1997
40. Beck DH, Taylor BL, Millar B, Smith GB: Prediction of outcome from intensive care: A prospective cohort study comparing Acute Physiology and Chronic Health Evaluation II and III prognostic systems in a United Kingdom intensive care unit. *Crit Care Med* 25:9–15, 1997
41. Fiaccadori E, Lombardi M, Leonardi S, Rotelli CF, Tortorella G, Borghetti A: Outcome of pre-existing malnutrition in acute renal failure: A prospective study. *J Am Soc Nephrol* 10:581–593, 1999
42. Star RA: Treatment of acute renal failure. *Kidney Int* 54:1817–1831, 1998
43. Schiffl H, Lang SM, Konig A, Strasser T, Haider MC, Held E: Biocompatible membranes in acute renal failure: Prospective, case-controlled study. *Lancet* 344:570–572, 1994
44. Bellomo R, Parkin G, Boyce N: Acute renal failure in the critically ill: Management by continuous veno-venous hemodiafiltration. *J Crit Care Med* 8:140–144, 1993
45. Hakim RM, Wingard RL, Parker RA: Effect of the dialysis membrane in the treatment of patients with acute renal failure. *N Engl J Med* 331:1338–1342, 1994
46. Zhu BP, Lemeshow S, Hosmer DW, Klar J, Avrunin J, Teres D: Factors affecting the performance of the models in the mortality probability model II system and strategies of customization: A simulation study. *Crit Care Med* 24:57–63, 1996
47. Silverstein MD: Predictive instruments and clinical judgment in critical care. *JAMA* 26:1758–1760, 1988
48. ESICM-SRLF Consensus Conference: Predicting outcome in ICU patients. *Intensive Care Med* 20:390–397, 1994