plantation and response in immunosuppression. World J Surg 1981; 5: 430.
15. Florack G, Sutherland DER, Sibley RK, Najarian JS, Squifflet JP. Combined kidney and segmental pancreas allotransplantation in dogs. Transplant Proc 1985; 17: 374.
16. Kelly WD, Lillehei RC, Merkel FK, Idezuki Y, Goetz FC. Allo transplantation of the pancreas and duodenum along with the kidney in diabetic nephropathy. Surgery 1967; 61: 827.
17. Koyama I, Williams M, Cameron JL, Zuidema GD. Experimental pancreatic allotransplantation in large animals. Transplantation 1986; 42: 333.
18. Nakai I, Kaufman DB, Field MJ, Morel P, Sutherland DER. Differential effects of preexisting uremia and a synchronous kidney graft on pancreas allograft functional survival in rats. Transplantation 1992; 54: 17.
19. Vogt P, Hiller WFA, Steininger B, Klempnauer J. Differential response of kidney and pancreas rejection to cyclosporine immunosuppression. Transplantation 1992; 53: 1269.
20. Westra AL, Petersen AH, Prop J, Wildevuur CRH. The combi-effect—reduced rejection of the heart by combined transplantation with the lung or spleen. Transplantation 1991; 52: 952.
21. Pollack R, Blanchard JM, Lazda VA. The splenic microenvironment and self recognition as factors in allograft rejection in rats—a study using indium-III labeled cells. Transplantation 1986; 42: 528.
22. Dafoe DC, Campbell DA, Marks WH, Borgstrom A, Lloyd RV, Turcotte JG. Association of inclusion of the donor spleen in pancreaticoduodenal transplantation with rejection. Transplantation 1985; 40: 579.
23. Dafoe DC, Rosenberg L, Campbell DA, Merion RM, Lichter AS, Turcotte JG. Clinical pancreaticoduodenal allotransplantation with inclusion of the donor spleen (irradiated and non-irradiated) to prevent thrombosis. Transplant Proc 1988; 20 (suppl. 1): 876.
24. Prieto M, Sutherland DER, Fernandez-cruz L, Heil J, Najarian JS. Urinary amylase monitoring for early diagnosis of pancreas allograft rejection in dogs. J Surg Res 1986; 40: 597.

Received 10 November 1992.
Accepted 30 March 1993.

1607

0041-1337/94/5707-1028$03.00/0
TRANSPLANTATION

Vol. 57, 1028–1036, No. 7, April 1994
Printed in U.S.A.

EARLY DEATH OR RETRANSPLANTATION IN ADULTS AFTER ORTHOTOPIC LIVER TRANSPLANTATION

CAN OUTCOME BE PREDICTED?[1]

HOWARD R. DOYLE,[2] IGNAZIO R. MARINO, NICHOLAS JABBOUR, GIORGIO ZETTI, JOHN MCMICHAEL, SANDI MITCHELL, JOHN FUNG, AND THOMAS E. STARZL

Pittsburgh Transplantation Institute, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

Early, reliable outcome prediction after a liver transplant would help improve organ use by minimizing unnecessary retransplantations. At the same time, early intervention in those cases destined to fail may ameliorate the high morbidity and mortality associated with retransplantation. The purpose of this study was to analyze several parameters that have been identified in the past as being associated with patient and graft outcome, and to try to develop a model that would allow us to make predictions based on data available in the early postoperative period.

A total of 148 patients were followed in a prospective, observational study. Graft failure was defined as patient death or retransplantation within 3 months of surgery. Preoperative variables studied included patient demographics, need for life support, presence of ascites, serum bilirubin, serum albumin, prothrombin time, serum creatinine, and the results of the cytotoxic crossmatch. During the first 5 postoperative days, standard measurements included serum transaminases, serum bilirubin, ketone body ratio, prothrombin time, factor V, and serum lactate. Oxygen consumption was measured shortly after surgery, once the patients had rewarmed to 36°C.

There were 131 successful transplants (88.5%) and 17 failures (11.5%). Most of the variables studied were found to be associated with outcome (by univariate analysis) at different points in the early postoperative period. However, receiver operating characteristic curve analysis showed that the predictive ability of even the best parameter was not adequate to make decisions on individual patients. Multivariate analysis, using stepwise logistic regression, yielded a model with an overall accuracy of 92.7%. Again, receiver operating characteristic curve analysis suggested that this model did not achieve the discriminating power needed for routine clinical use.

We are still not able to accurately predict outcome in the early posttransplant period. We must be very careful when evaluating parameters, or scoring systems, that are said to accomplish this. It is especially important in this era of cost containment, with its renewed pressures to guide therapy based on our perceived understanding of a patient's future clinical course.

This study was supported by Project Grant DK 29961 from the National Institutes of Health, Bethesda, MD.

[2] Address correspondence to: Howard R. Doyle, MD, 3601 Fifth Avenue, Suite 5C, Pittsburgh, PA 15213.

Predicting the outcome soon after liver transplantation remains an elusive goal, although the need to anticipate the result of a therapeutic intervention is particularly compelling in lifesaving organ transplantation. We would like to determine early whether an organ is destined to fail, since early intervention might ameliorate the increased morbidity and mortality that accompany retransplantation. The reverse is also true. Given the donor shortage, we need to be able to determine confidently when it is indicated to manage the patient expectantly, so we will not waste a scarce societal resource.

Several methods have been brought forth, all aimed at helping clinicians make an early determination regarding the ultimate fate of both the graft and patient. Many consist of simply examining individual parameters, such as clotting factors (*1*, *2*), arterial ketone body ratio (*3–6*), lidocaine metabolism (*7*), oxygen consumption (*8–10*), and serum lactate (*11*). Others are reasonably complex scoring systems that, presumably, model biological reality more closely (*12–16*).

At present, no method has been adopted for general clinical use, if by this we mean making difficult individual decisions based upon them. Some are limited by incorporating subjective information that is difficult to standardize and quantify, such as our overall clinical impression regarding the degree of malnutrition in cirrhotic patients (*12*). Others cannot be used in certain subpopulations, such as insulin-dependent diabetics, patients with pancreatitis, or those who are starving (*3*, *4*).

The purpose of our study was to analyze several factors that empirically have been shown to be associated with patient and graft outcome after liver transplantation. We wanted to determine whether recent advances in this field have changed the significance of some of these factors, and whether we can develop a model that will allow early prediction of patient and graft outcome, based on data available in the early postoperative period.

MATERIALS AND METHODS

Patient Population

From January to August 1992, 149 adult patients undergoing liver transplantation at Presbyterian University Hospital, Pittsburgh, PA, were entered into the study. They were followed prospectively for at least 3 months. Six patients that either died on the operating table or did not survive more than 24 hr were excluded. The study patients underwent a total of 156 transplants, but to keep the observations independent, only the initial transplant is included in our analysis. We also excluded a young woman who died of intractable supraventricular arrhythmias 23 days after the operation. This patient was transplanted for Budd-Chiari syndrome, and required a direct anastomosis between the suprahepatic vena cava and the right atrium. She developed arrhythmias immediately after transplant that were attributed to direct trauma to the conduction system. At the time of her death from a cardiac arrest, she had normal graft function. This left 148 patients, who formed the basis for this report.

Parameters Studied

Preoperative information included age and sex, number of liver transplants (including the index transplant), whether the patient was in the intensive care unit (ICU)*, need for mechanical ventilation immediately before transplantation, presence of ascites, total serum bilirubin, serum albumin, prothrombin time, serum creatinine, and results of the cytotoxic crossmatch.

After surgery, patients were allowed to rewarm to 36°C, at which time the central venous pressure (CVP) and pulmonary wedge pressure were measured by means of a pulmonary artery catheter. The cardiac output (CO) was determined using the thermodilution method, with a cardiac output computer (COM-1, Irvine, CA). The mean arterial pressure (MAP) was measured via an indwelling radial or femoral arterial catheter. Arterial and mixed venous blood samples were obtained for determinations of hemoglobin concentration and gas tensions using a blood gas analyzer (ABL-4, Radiometer, Copenhagen, Denmark). The following values were then calculated using standard formulas:

cardiac index (CI) = CO/body surface area
systemic vascular resistance index (SVRI) = MAP − CVP * 80/CI
arterial oxygen content (CaO_2) = Hb * 1.39 * SaO_2 + 0.003 * PaO_2
venous oxygen content (CvO_2) = Hb * 1.39 * SvO_2 + 0.003 * PvO_2
arteriovenous oxygen difference ($AVDO_2$) = CaO_2 − CvO_2
oxygen delivery (DO_2) = CI * CaO_2 * 10
oxygen consumption (VO_2) = CI * $AVDO_2$

During the first 5 days after surgery, we recorded the total serum bilirubin, serum aspartate aminotransferase, serum alanine aminotransferase, serum lactate, arterial ketone body ratio (AKBR), factor V, and prothrombin time. These measurements were all part of our routine postoperative protocol. AKBR was determined as described previously (*4*). If the team managing the patient ordered several measurements in a given day, we recorded only those values that were obtained close to, and were available for, morning rounds.

The only intraoperative parameter analyzed was the peak serum lactate. Other variables associated with the surgery, or the donated organ, have been identified as having prognostic significance. We postulated that their influence would translate into a recognizable postoperative state, and did not include them. The decision to retransplant was made by the clinical team caring for the patient, in consultation with the most experienced members of the transplant division. Since the study was strictly observational, and patients were treated according to our established clinical protocols, institutional review board approval was not necessary.

Definitions. Graft failure: patient death or retransplantation within 3 months of surgery in patients who survived at least 24 hr after the operation. *Patient groups*: group I, patients who underwent a successful transplant, that is, who survived more than 90 days without need for retransplantion; group II, patients who had graft failure, as defined above. *Primary nonfunction*: lacking a technical complication, a graft that never demonstrates evidence of initial function, such that retransplantation has to be carried out within 2 weeks of the original operation, or the patient succumbs to liver failure before a suitable graft can be obtained. *Sepsis*: systemic response to infection, manifested by 2 or more of the following—(1) temperature > 38°C or < 36°C, (2) heart rate > 90 beats/min, (3) respiratory rate > 20 breaths/min or $PaCO_2$ < 32 torr, (4) white blood cells > 12,000 cell/mm^3, < 4,000 cells/mm^3, or > 10% immature (band) forms (*17*). *Severe sepsis*: sepsis associated with organ dysfunction, hypoperfusion, or hypotension (*17*).

Statistical Analysis

Continuous data are presented as the mean ± SEM. Categorical data are presented as frequencies. For continuous data, independent two-tailed *t* tests were used to determine whether there was a difference between groups, with the level of significance set at 0.05. The Levene test was used to determine whether the population variances were unequal, in which case a separate-variance *t* test was used instead. In cases in which the distribution of the data was highly skewed, a log transformation was also performed, and both transformed and untransformed data were tested for significance. Pear-

* Abbreviations: AKBR, arterial ketone body ratio; AUC, area under the receiver operating characteristic curve; ICU, intensive care unit; POD, postoperative day; ROC, receiver operating characteristic; SGPT, serum glutamic pyruvic transaminase; SGOT, serum glutamic oxaloacetic transaminase.

son's chi-square statistic was used to test for differences in data expressed as frequencies, with the level of significance set at 0.05.

Stepwise logistic regression analysis was used to derive an expression to calculate the probability of graft failure in a given patient. Several models were constructed, using both forward stepwise selection and backward elimination, with a likelihood ratio test to determine which variables should be removed from the model. The final model was adopted based on general accuracy, interpretability, and receiver operating characteristic (ROC) curve analysis.

To evaluate the predictive power of individual tests, and that of the logistic regression models, ROC curve analysis was performed using software (Labroc1) kindly provided by Dr. Charles Metz, Department of Radiology, University of Chicago, Chicago, IL. All other calculations were done using SPSS for Windows.

RESULTS

Of 148 patients analyzed, 131 (88.5%) were successful according to our definition (group I) and 17 (11.5%) failed (group II).

Cause of Graft Failure

Severe sepsis: Seven graft failures were due to patient death from severe sepsis, occurring 2–64 days after transplantation. *Pancreatitis*: One patient died of acute necrotizing pancreatitis 30 days after being retransplanted for recurrent hepatitis B. *Arrhythmias*: One patient died after developing ventricular tachycardia and subsequent fibrillation 3 days after surgery. He had no predisposing cardiac risk factors. *Primary nonfunction*: There were 3 instances of primary nonfunction. *Ischemic injury*: Four grafts were lost due to severe ischemic injury 16–90 days after transplantation. *Rejection*: Only 1 graft was lost to uncontrolled acute rejection, with severe central venulitis, 33 days after transplantation.

Preoperative Clinical Variables

Table 1 shows the preoperative clinical characteristics of both groups. The only preoperative factors that were statistically associated with outcome were serum creatinine and whether the patient was undergoing a retransplantation. The difference in total serum bilirubin was only of marginal significance ($P=0.05$), and this was also the case in the subgroup of patients whose preoperative serum bilirubin was greater than 3 mg/dl (46.4% of group I and 70.6% of group II, $P=0.06$). There were no differences in terms of age and sex, presence of ascites, degree of hypoalbuminemia, prolongation of prothrombin time, or whether the patient was ICU bound or required mechanical ventilation before the transplant. There were no patients in group II who had a positive cytotoxic crossmatch, whereas 14.3% of those in group I had a positive crossmatch (7.6% strong positive, 5% positive, and 1.7% weakly or doubtful positive). This difference, however, was not significant ($P=0.1$).

TABLE 1. Demographics and selected preoperative parameters for patients with successful (group I) and failed (group II) grafts

	Group I (n=131)	Group II (n=17)	
Age (yr)	49.5±1.2	53.1±2.9	NS
Sex (M/F)	76/55	12/5	NS
ICU bound	32%	53%	NS
Mechanical ventilation	15.6%	23.5%	NS
Retransplantation	11.2%	35.3%	$P<0.01$
Total bilirubin (mg/dl)	8.2±0.97	17.5±4.3	$P=0.05$
Prothrombin time (sec)	14.6±0.23	14.5±0.53	NS
Albumin (g/dl)	3.22±0.07	2.9±0.17	NS
Creatinine (mg/dl)	1.42±0.11	2.1±0.32	$P<0.05$
Ascites	69.8%	75%	NS
Positive crossmatch	14.3%	0%	NS

Postoperative Clinical Variables

Figure 1 shows the postoperative behavior of serum bilirubin, creatinine, lactate, and AKBR. The differences in the serum bilirubin became more pronounced during the early postoperative period. Mean serum bilirubin on postoperative day (POD) 1 was 7.6±0.47 mg/dl in group I and 12.1±1.7 mg/dl in group II ($P=0.02$). By POD 5, the mean value in group I had decreased to 5.9±0.47, but in group II it had increased to 15.4±2.4 ($P<0.01$). This was not due to preoperative differences that were carried into the postoperative period. If we set to 1 the value of serum bilirubin on POD 1, in group I it declined to 0.83±0.04 by POD 2, while in group II it increased to 1.24±0.13 ($P<0.01$). By POD 5, the bilirubin ratio in group I was 0.91±0.06, whereas in group II it had increased to 1.74±0.32 ($P=0.02$). So, on average, serum bilirubin decreases slightly in the early postoperative period in successful grafts, while it increases significantly in failed grafts.

Serum creatinine increased in both groups after the operation, but did so at a much higher rate in group II. Whereas the serum creatinine on POD 1 was 1.26±0.07 in group I and 1.64±0.23 in group II (NS), it had gone up by POD 5 to 1.59±0.09 in group I and 3.1±0.46 in group II ($P<0.01$).

The AKBR tended to be higher in group I, but the difference between the 2 groups was significant only on POD 1 (1.02±0.04 vs. 0.8±0.08, $P<0.05$). Peak lactate during surgery was 9.9±0.37 mmol/L in group I and 10.5±1.2 in group II (NS). Postoperative serum lactate was higher in group II, although it tended to normalize during the early postoperative period in both groups. Differences were significant on POD 2 (2.2±0.1 for group I vs. 3.3±0.53 for group II, $P<0.05$) and POD 5 (1.8±0.1 for group I vs. 2.9±0.4 for group II, $P<0.01$).

Figure 2 shows the behavior of serum transaminases, factor V, and prothrombin time in the early postoperative period. The serum glutamic oxaloacetic transaminase (SGOT) decreased steadily in both groups, from 1458±153 (group I) and 2424±753 (group II) on POD 1 down to 135±14 (group I) and 241±61 (group II) by POD 5. These differences were significant only on POD 3. The serum glutamic pyruvic transaminase (SGPT) values on POD 1 were 924±84 (group I) and 1256±323 (group II) and rose transiently on POD 2 before starting to decline. However, due to the small size of group II and the large variance, this observation is probably of little significance. By POD 5, SGPT had decreased to 394±33 (group I) and 572±133 (group II). None of these differences reached statistical significance. Factor V rose steadily and linearly throughout the early postoperative period. It was significantly higher in group I on POD 2 (0.56±0.03 vs. 0.39±0.08, $P<0.05$), POD 3 (0.8±0.04 vs. 0.47±0.1, $P<0.01$), and POD 5 (1.09±0.04 vs 0.75±0.1, $P<0.01$). Prothrombin time on PODs 2–4 also correlated well with outcome, although the trend was the same in both groups. Group I patients had a mean prothrombin time of 14.7±0.27 sec on POD 2, while in group II patients, it was 19.6±1.56 sec

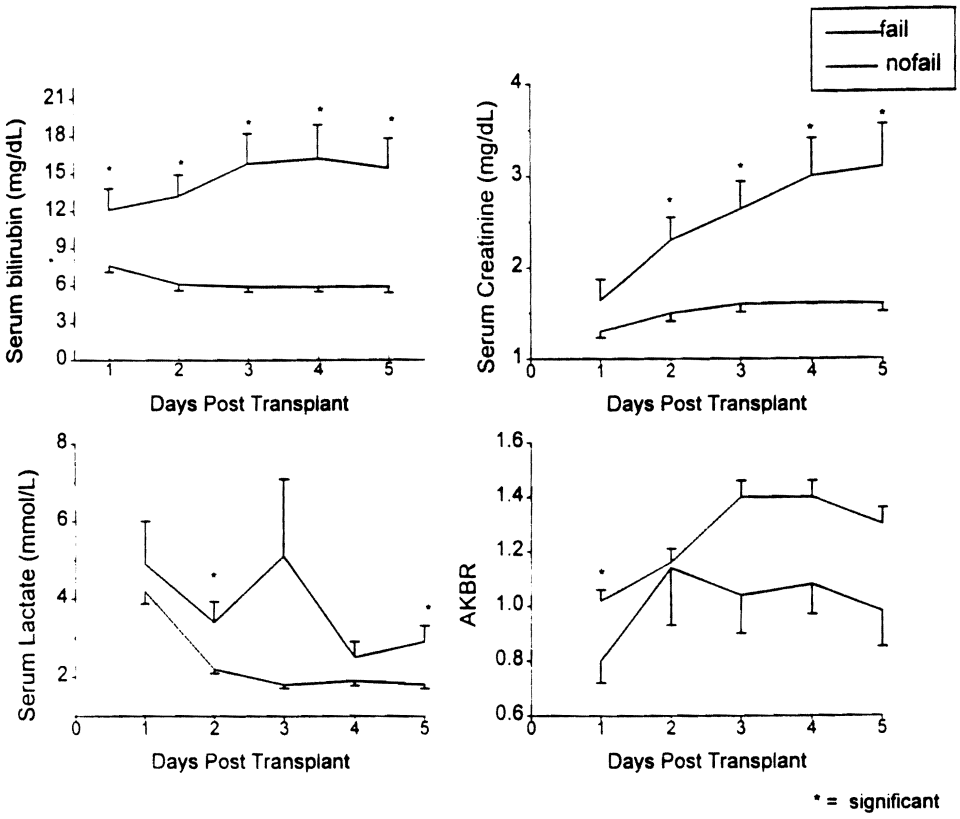

FIGURE 1. Serum bilirubin, creatinine, lactate, and AKBR for the first 5 postoperative days in patients with successful (nofail) and unsuccessful (fail) grafts.

($P<0.01$). By POD 4, the prothrombin time had decreased to 12.7±0.12 in group I and 15.3±1.1 in group II ($P<0.05$).

To compensate for the unequal variances, and the fact that most of these data had a highly skewed distribution, a log transformation of selected variables was done. This revealed that the differences between groups were significant during much of the observation period for both SGOT ($P<0.01$ on days 2 and 3, and $P<0.05$ on days 4 and 5) and SGPT ($P<0.05$ on days 2–4). Similar results were obtained when transforming AKBR ($P<0.05$ on days 1, 3, and 5), serum lactate ($P<0.05$ on days 2 and 3, and $P<0.01$ on day 5), and preoperative serum bilirubin ($P=0.03$).

Table 2 shows the area under the ROC curve (AUC) for variables measured on POD 2 and POD 3, as well as multivariate models generated for those days using logistic regression analysis (see below). An ROC curve is obtained by plotting the true-positive fraction against the false-positive fraction. It describes the ability of a test to discriminate between disease and nondisease, and is independent of prevalence (*18*). In our particular case, we are measuring their ability to correctly classify patients as either belonging to the successful or failed graft groups. An AUC of 1 means the test has perfect discrimination, whereas an AUC of 0.5 means it is no better than chance. We can see that, individually, serum bilirubin performs better than any of the other biochemical or hematological parameters measured on those days. Serum creatinine and prothrombin time follow. There also is a tendency toward improvement in the discriminatory ability of some parameters, as we go from POD 2 to POD 3. Tables 3 and 4, however, give us a better idea as to how these tests translate in actual practice. If we take the individual parameter with the best overall performance, serum bilirubin on POD 3, we can see what happens as we vary the cutoff value that we use to decide whether the test is "abnormal" or not, that is, the point at which we would consider the test indicative of eventual graft failure. If we set this point at 23.5 mg/dl, we obtain a high degree of specificity (the false-positive fraction is only 2%), but at the cost of a very low sensitivity (true-positive fraction of only 16%). As we lower the cutoff value, the sensitivity increases, but so does the false-positive fraction. To obtain a sensitivity of 75%, we need to be willing to accept a false-positive fraction of 15%.

Oxygen delivery (DO_2/m^2) in the early postoperative period was 727±18 ml/m² in group I, while in group II, it was 625±49 ml/m² (NS). Oxygen consumption (VO_2/m^2) was 124±2 ml/m² in group I and 150±23 ml/m² in group II (NS). There was no difference in the core temperature between the 2 groups at the time these measurements were made (36.4±0.1°C vs. 36.2±0.1°C).

Several models were generated, using logistic regression analysis and information available on either POD 2 or POD 3, in an attempt to increase the predictive power by combining several parameters. The choice of POD 2 or POD 3 was made because it allows the washout effect of massive intraoperative transfusions and fluid resuscitation to fade, while the differences in parameter values are still at a maximum. The models were then further evaluated, considering their overall accuracy, goodness of fit, and AUC. The best overall model, generated using information from POD 3, had a predictive accuracy of 92.7%, and is described by the following expression:

$$p = \frac{e^{(-7.65 + 0.15 * \text{Bili} + 0.16 * \text{PT} + 0.58 * \text{Cr})}}{1 + e^{(-7.65 + 0.15 * \text{Bili} + 0.16 * \text{PT} + 0.58 * \text{Cr})}}$$

where P is the probability that a graft will fail, Bili is the

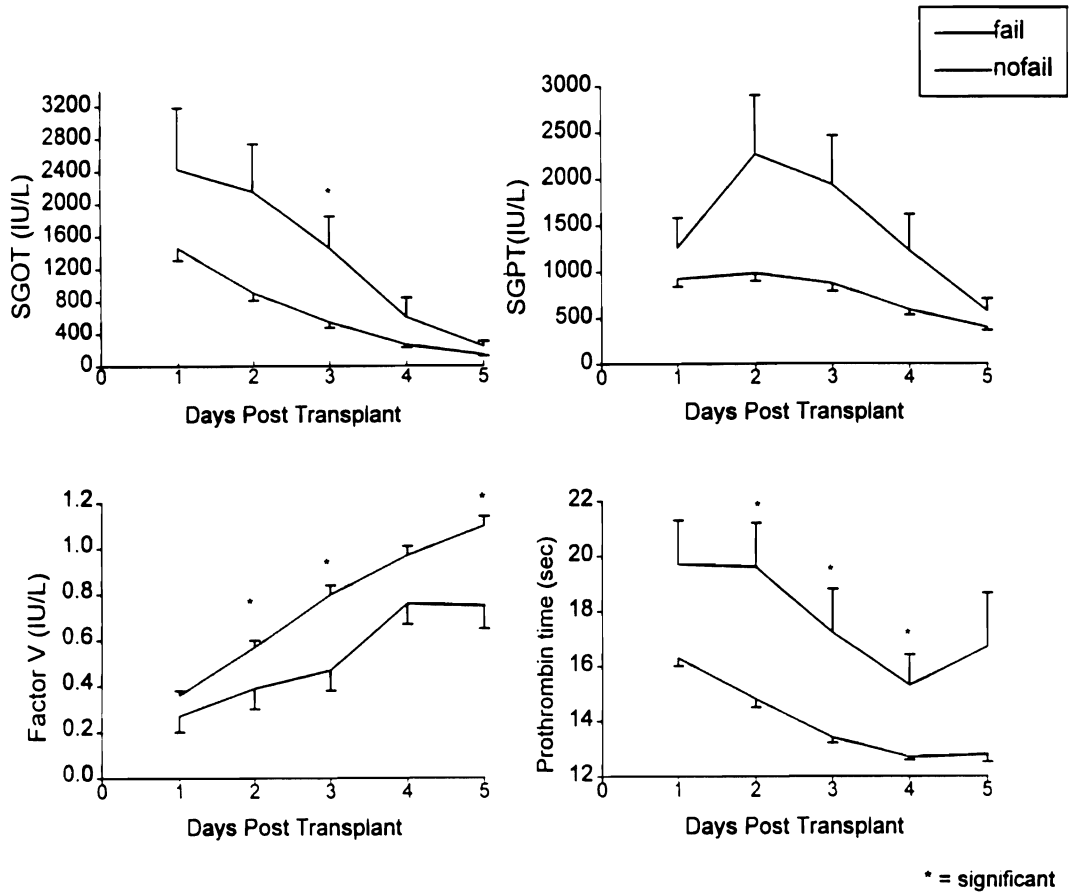

FIGURE 2. Serum transaminases, factor V, and prothrombin time for the first 5 postoperative days in patients with successful (nofail) and unsuccessful (fail) grafts.

TABLE 2. AUC for individual parameters and the logistic regression (LR) model on PODs 2 and 3

	AUC day 2	AUC day 3
LR model	0.84±0.05	0.90±0.04
Bilirubin	0.84±0.04	0.89±0.03
Creatinine	0.76±0.05	0.79±0.05
Prothrombin time	0.72±0.087	0.75±0.08
SGOT	0.70±0.07	0.71±0.07
Lactate	0.68±0.07	0.75±0.08
Factor V	0.67±0.08	0.74±0.08
SGPT	0.66±0.08	0.67±0.08
AKBR	0.60±0.08	0.67±0.08

serum bilirubin, PT is prothrombin time, Cr is serum creatinine, and e is the base of the natural logarithms. The best model for POD 2 had an overall accuracy of 87.8%. Their ROC curves, constructed using the individual probabilities estimated by the models, had an AUC of 0.90±0.04 for POD 3 and 0.84±0.05 for POD 2. These areas are no different from those of serum bilirubin for the same days (Table 2). If we look at the individual performance of the 2 models, we can appreciate that if we set the cutoff value for the POD 3 model (Table 4) at a probability of failure of 0.73, we achieve perfect specificity, but the true-positive fraction is only 17%. A cutoff value of 0.11 gives us a true-positive fraction of 75%, but the false-positive fraction increases to 13%.

Analyses performed using log-transformed data, or the ratios of different parameters with respect to their initial values, did not improve their predictive ability.

TABLE 3. ROC curve analysis showing the performance at various critical result values for individual parameters and the logistic regression (LR) model on POD 2

	Cutoff value	FPF[a]	TPF[a]
LR model	0.65	0.02	0.24
	0.41	0.06	0.38
	0.15	0.21	0.70
	0.02	0.85	0.99
Bilirubin (mg/dl)	21.2	0.03	0.16
	10.75	0.13	0.52
	6.35	0.32	0.86
	2.4	0.81	0.99
Creatinine (mg/dl)	3.45	0.04	0.14
	1.85	0.26	0.62
	1.25	0.49	0.87
	0.75	0.82	0.99
Prothrombin time (sec)	26.35	0.01	0.18
	16.55	0.19	0.55
	13.95	0.51	0.76
	12.35	0.86	0.92
SGOT (IU/L)	4,391	0.01	0.13
	1,315	0.19	0.49
	472	0.53	0.76
	211	0.81	0.91

[a] FPF, false-positive fraction; TPF, true-positive fraction.

DISCUSSION

Given the lack of uniform definitions in this field, the first decision to make when studying outcome after liver transplantation is how to define graft failure. In clinical practice, we would like to differentiate between 3 broad categories:

TABLE 4. ROC curve analysis showing the performance at various critical result values for individual parameters and the logistic regression (LR) model on POD 3

	Cutoff value	FPF[a]	TPF[a]
LR model	0.73	0.00	0.17
	0.47	0.02	0.40
	0.11	0.13	0.75
	0.03	0.43	0.96
Bilirubin (mg/dl)	23.5	0.02	0.16
	9.40	0.15	0.75
	6.70	0.26	0.90
	2.25	0.80	1.00
Creatinine (mg/dl)	3.75	0.06	0.12
	1.85	0.29	0.73
	1.35	0.46	0.92
	0.75	0.80	0.99
Prothrombin time (sec)	23.00	0.01	0.15
	14.90	0.13	0.51
	13.00	0.42	0.75
	11.85	0.82	0.93
Lactate (mmol/L)	5.50	0.01	0.16
	3.30	0.06	0.40
	2.25	0.21	0.60
	1.00	0.81	0.93

[a] FPF, false-positive fraction; TPF, true-positive fraction.

those patients who will survive with a well-functioning graft, those who will survive but require retransplantation, and those who will die in the early posttransplant period (with or without retransplantation). Our definition was deliberately broad, and was that of either patient death or retransplantation within 3 months of surgery. Patients who did not survive longer than 24 hr were excluded because there is not enough information available for analysis, and these cases seldom constitute a prognostic dilemma. We elected to use 3 months as our final cutoff point because most of the deaths or graft losses that occur during the first year take place within 90 days. For those patients transplanted at our institution in 1991, patient survival was 81% at the end of 1 year, with a graft survival of 72%. Sixty-seven percent of the deaths and 72% of the graft losses occurred within 90 days (unpublished data). These figures include those losses that took place within 24 hr.

We did not attempt to make a distinction between grafts that failed due to need for retransplantation and those due to patient death, since the small number of cases expected in the graft failure group would have precluded any meaningful statistical analysis. Another issue was whether to separate deaths that were seemingly unrelated to graft failure into a distinct category. Shaw et al. (*15*) have suggested a stratification of posttransplantation deaths to study failures in a more systematic way. They divided patient deaths, 44 in their series of 180 liver transplants, into 4 categories, depending on whether they were directly related to pretransplantation morbidity, surgical complications, failures of immunosuppression, or unusual circumstances. Although this classification provides a framework for further study, and is conceptually attractive in that it focuses our attention on different patient categories, we decided not to use it. It is often hard to determine why a graft failed or a patient died. Should a ruptured mycotic aneurysm be included among the technical failures or the unexpected deaths? If the rupture follows acute pancreatitis in a patient with hepatitis B, should that be grouped with pretransplant health conditions? Is sepsis the primary cause of death in a patient with marginal graft function, or just the terminal event, in the presence of liver failure? On the other hand, and especially if we are trying to study those factors that determine posttransplant morbidity and mortality, we need to exclude at least those unusual cases where we can be reasonably certain that the morbid event was unrelated to the procedure. At this point, we can give no strict rules. Each case should be considered on its own merits, and one should err on the side of over inclusion, rather than high selection.

Seven patients died of documented bacterial sepsis. There were 3 instances of primary nonfunction, and an additional 4 grafts were lost to severe ischemic injury. One patient died of pancreatitis, and another lost the liver to severe ongoing rejection. One patient died of cardiac arrhythmias 3 days after surgery, with no predisposing factors identified. In 41.3%, graft failure was due to patient death from sepsis, another 41.3% were lost to either primary nonfunction or severe ischemic injury, which most of the time reflects the same pathologic process, and 17.4% were lost to miscellaneous causes, such as pancreatitis, arrhythmias, and rejection. Sepsis continues to be the leading cause of death in liver transplant patients. In a study performed at our institution in 1986 (*19*), 52.5% of early deaths were due to infectious complications, and 71% of ICU deaths in the Cambridge program had the same etiology (*20*). In contrast to an earlier series from our institution (*19*), no patients died of uncontrollable rejection.

We did not specifically analyze the effect of Child-Pugh scores (*21*), a measure that, in the past, has not been useful in this setting (*22*). Our findings, when we consider independently 4 of the 5 parameters used to calculate it (see Table 1), are in keeping with this. There were no differences in preoperative prothrombin time, serum albumin, or presence of ascites. The difference in preoperative serum bilirubin was only marginally significant ($P=0.05$), although this might have been a function of the large unequal variances and the relatively small size of group II, rather than biological unimportance. There was also no difference in the fraction of patients whose bilirubin was greater than 3 mg/dl (see *Results*). We find the Child-Pugh score less than useful for other reasons as well. Cirrhotic patients routinely receive infusions of albumin solutions, either after large volume paracentesis or to correct some arbitrarily defined level of hypoalbuminemia. Although the latter practice is of dubious value, it seems to have become firmly entrenched, and it is now very difficult to determine what the baseline serum albumin was in a given patient. We also find encephalopathy to be difficult to assess in mechanically ventilated patients, who very often receive some form of sedation.

Two preoperative factors that we found associated with outcome, namely serum creatinine and whether the patient was undergoing a retransplantation, are well-known risk factors. Serum creatinine has been identified previously by our group as the best single preoperative indicator of short-term prognosis after liver transplantation (*13*), and retransplantation continues to carry a significantly higher mortality (*23*).

The proportion of patients that were either ICU bound or on mechanical ventilation was higher in group II, although the difference was not significant. Thirty-two percent of group I patients were ICU bound, as opposed to 53% for group II ($P=0.08$), and 15.6% in group I were on mechanical

ventilation, compared with 23.5% in group II ($P=0.4$). This is in contrast to a larger group of patients from our institution who were analyzed recently, in whom a higher urgency status and need for life support correlated with higher mortality (*24*). The patients in our current study are different in at least 2 respects: we excluded those who did not survive longer than 24 hr or who were retransplanted after having been entered in the study. This probably included some of our sickest patients, and explains the different results.

An interesting finding was the fact that a positive cytotoxic crossmatch had apparently no deleterious effects. All the positive cytotoxic crossmatches were found in group I (14.3%). Takaya et al. (*25*) have previously described the adverse effects of a positive cytotoxic crossmatch on patients treated with FK506 and low-dose steroids. Those with a positive crossmatch had 1-year graft and patient survivals of 56% and 68%, respectively, as opposed to 82% and 86% for those with a negative crossmatch. In positive crossmatch patients, intraoperative blood usage was higher, similar to a previous observation made in high panel-reactive antibody patients by Marino et al. (*26*), and postoperative liver function took longer to normalize. In a small series reported later, the authors found that, in the case of a positive crossmatch, adding a higher steroid regimen and PGE1 infusion allowed them to obtain results that were comparable to those seen with a negative crossmatch (*27*). This prompted a change in policy at our institution whereby, in the face of a positive crossmatch, patients receive a traditional methylprednisolone burst and subsequent taper. All patients who can tolerate it are given a PGE1 infusion during the first posttransplant week. Although we were not specifically studying this point, our data are consistent with the notion that the use of high-dose steroids, and probably PGE1, ameliorates the untoward effects of a positive crossmatch (*28*).

Several postoperative factors were associated with graft failure. Serum bilirubin was higher in group II throughout the early postoperative period (Fig. 1). This difference was due not only to higher preoperative levels, but also to the fact that bilirubin tended to rise even more in cases of graft failure, while it decreased slightly in patients with successful grafts. Renal function was also worse in group II patients, as reflected by a much higher serum creatinine throughout most of the early postoperative period (Fig. 1). McCauley et al. have studied renal function after liver transplantation both in patients receiving CsA (*29*) and those receiving FK506 (*30*). They found a 39.4% incidence of posttransplant renal failure (serum creatinine greater than 3 mg/dl), which correlated strongly with mortality (*29*). Our initial experience showed that renal function also tends to deteriorate after liver transplantation under FK506 therapy (*30*), and the present data indicate that the association between renal dysfunction and liver allograft failure still holds. The etiology of this renal dysfunction is certainly multifactorial, with drug toxicity, extent of pre-existing liver disease, intraoperative hemodynamic instability, and presence of sepsis playing a role.

Abnormalities in blood coagulation are a hallmark of end-stage liver disease, and have been used for many years to gauge allograft function (*31*). In this series, prothrombin times on days 2 through 4 were significantly higher, and factor V levels on days 2, 3, and 5 were significantly lower in patients who ultimately failed (Fig. 2). Factor V levels reach a nadir shortly after revascularization of the graft, and return to normal over the next several days (*31*). It has been suggested that a greatly depressed factor V level in the early postoperative period is a very good predictor of outcome, both in terms of the graft (*1*) and overall mortality (*14*). Our results certainly support the idea that coagulation abnormalities parallel graft function and eventual outcome. However, the discriminating power of the tests, especially factor V, does not seem as good as initially thought. One problem with these 2 measurements (factor V and prothrombin time), and many others we rely upon, is that they can be affected significantly by replacement therapy that is far from standardized, even within the same institution. Different patients with the same degree of coagulopathy, as measured by standard testing, can receive significantly different amounts of replacement products depending on other factors, some of which can be largely subjective. Although we did not quantify the volume of replacement products that were administered, the number of patients who received one or more infusions of fresh-frozen plasma during a given 24-hr period was significantly higher in group II on days 3 and 4, and of marginal significance on day 2 (data not shown, $P<0.01$ for days 3 and 4, and $P=0.055$ for day 2). It is unlikely, then, that the lesser derangements seen in group I were due to a better correction of their coagulopathy, if one was present. On the other hand, the differences between the groups would probably have been more pronounced if replacement therapy had been the same for both.

The metabolic state of the patient, as reflected by the AKBR, oxygen consumption, and serum lactate are also considered important correlates of outcome. AKBR (Fig. 2) on POD 1 was significantly higher in group I, and after a log transformation of the data, the differences were also found to be significant on PODs 3 and 5. The discriminating power of AKBR in this series of patients was much lower than that quoted previously in the literature (*3, 5*), including reports from our institution (*4*). Factors that may explain this include the use of a different definition of graft failure, and the fact that we did not exclude from the analysis patients who were diabetic, had pancreatitis, or showed biochemical evidence of starvation. Under these conditions, if we use the previously suggested AKBR value of less than 0.7 by POD 2 as our cutoff point, ROC curve analysis shows the false-positive fraction would be 16%, while the true-positive fraction would be only 35%.

Oxygen consumption (VO_2) is decreased in cirrhotic patients (*32*), perhaps reflecting an abnormal supply dependency in this population (*33*). As might be expected, this falls even further during the anhepatic phase (*8, 34*). There have been attempts at correlating perioperative changes in oxygen consumption with graft outcome, but some series are limited by the small number of patients (*8, 35*). In a study of 125 liver recipients who had VO_2 measured at several points during transplantation, it was found that, except for 4 patients who had primary nonfunction, there was a marked increase in VO_2 after reperfusion (*36*). In another report, based on 100 consecutive patients, it was stated that failure of VO_2 to increase above pre-anhepatic levels after reperfusion correctly identified patients with subsequent graft failure 70.6% of the time. When used with postreperfusion serum glucose, to arrive at a "metabolic index," correct classification increased to 92.4% of cases (*9*). The authors subdivided the

patients into 2 groups: those who survived with normal liver function for more than 3 months (group 1, n=90) and those who had primary nonfunction requiring retransplantation within 24 hr (group 2, n=10). It is unusual that a transplant program would encounter 100 consecutive patients who cleanly segregate into either one of these 2 categories, and it is not clear to us how applicable these findings are in more typical liver transplant populations.

Information regarding oxygen consumption in the posttransplant period is meager at best (*37, 38*), and there is only one report we are aware of that has looked at its relationship with early allograft failure (*10*). The authors reviewed retrospectively 86 patients who underwent primary liver transplantation, and in whom complete hemodynamic profiles were obtained within 48 hr of admission to the ICU. They found that VO_2 was significantly lower in the group of patients whose graft failed within the first month, while there were no differences in oxygen delivery (*10*). We found no difference in either DO_2 or VO_2 in this series. Our results may be due to a type II error, since the groups, especially the failure group, were relatively small. However, we believe that they are representative because these data were obtained prospectively, shortly after the patients had returned to the ICU and had been allowed to rewarm to 36°C. This reduced the confounding effects of hypo- or hyperthermia. It also allowed us to carry out the measurements while most patients were still recovering from the anesthetic, reducing the variability introduced by patient activity.

Lactic acidosis is a common finding in fulminant liver failure, and has a negative correlation with survival (*39*). During liver transplantation, there is a progressive hyperlactatemia, which usually peaks at the end of the anhepatic phase (*11*). It is likely due to diminished clearance by the liver, although tissue hypoxia may also play a role. We found that serum lactate was significantly higher in group II on both POD 2 and POD 5. In contrast to other reports (*11*), we did not find any difference between groups in terms of the intraoperative peak serum lactate level. By itself, serum lactate is a poor discriminator.

In fact, there is enough overlap between groups that no single parameter studied is very helpful when it comes to an individual patient, especially considering the decisions that might be contemplated. The same can be said for most of the trends we observe every day at the bedside. We all take comfort in the observation that "the patient's numbers are getting better," as if this were a reliable indicator of clinical improvement, when in reality this is not necessarily so. Clearly, if all liver function tests are unrelentingly deteriorating, the prognosis is poor, but these are not the problematic cases we would like to sort out quickly. As we can see in Figures 1 and 2, liver function test results tend to move in the same direction whatever the outcome. However, there is one important exception: serum bilirubin, which has been regarded as the most important indicator of clinical progress since the beginning of liver transplantation. However, although we can say that a continued rise in bilirubin portends a poor outcome, the individual variability is still such that it precludes us from formulating a consistent plan of action based solely on it. Serum creatinine stabilizes shortly after the transplant in the successful patients, while it continues to rise in the failure group. Renal dysfunction, however, can be caused by many factors, not the least important being the very medications we use routinely in these patients, and this complicates its use as a clinical predictor.

Even when we combine several parameters, we gain relatively little in discriminating power. The best multivariate models we have generated, with information available on the second and third postoperative days, had an overall accuracy of 87.8% and 92.7%, respectively. As good as these numbers may sound, we must be very careful when we interpret them, especially since they are greatly affected by the prevalence of the condition we are trying to predict. ROC curve analysis provides us with a more realistic picture of the performance we can expect from a test, or multiple regression model, and it is independent of prevalence (*18*). Tables 3 and 4 show the individual performances as we vary the critical test result, that is, as we change what we consider to be the cutoff point that we will use to assign a patient to one group or another. Even with multivariate models, to detect most patients who are destined to fail we need to be willing to accept onerous false-positive fractions. It is instructive to see that even after extensive multivariate analyses, we can do no better than we would by just asking how jaundiced is the patient that day.

We may be able to improve our models by gathering a larger sample, and then separating the presumably different categories that we elected to treat as one. It is also possible that we simply do not have enough information, because we are measuring the wrong things, or not measuring them often enough. It may be necessary to incorporate intraoperative and donor variables, which we chose to leave out under the premise that their influence should translate into a recognizable postoperative state. Whatever we do, we should take a critical look at the results, and remember that a very small *P*-value cannot be equated to predictive ability. We can no longer accept misleading measurements, such as overall accuracy, that are greatly influenced by the prevalence of the condition we are trying to predict. It is especially important in the current fiscal climate, where we will be pressured into making allocation decisions based on a given predictive model.

There is a growing feeling among us that the final answer is not going to be a simple matter of more data collection, so we can continue to take increasingly more elaborate snapshots. We are trying to describe complex, nonlinear systems that will very likely continue to resist our more traditional approaches. It may be time to borrow new tools.

Acknowledgments. The authors thank Derek Angus, MD, for reviewing the manuscript and providing insightful comments.

REFERENCES

1. Bilik R, Superina RA, Poon AO. Coagulation plasma factor levels are early indicators of graft nonfunction following liver transplantation in children. J Pediatr Surg 1992; 27(3): 302.
2. Moia M, Martinelli I, Gridelli B, et al. Prognostic value of hemostatic parameters after liver transplantation. J Hepatol 1992; 15: 125.
3. Taki Y, Gubernatis G, Yamaoka Y, et al. Significance of arterial ketone body ratio measurement in human liver transplantation. Transplantation 1990; 49(3): 535.
4. Asonuma K, Takaya S, Selby R, et al. The clinical significance of the arterial ketone body ratio as an early indicator of graft viability in human liver transplantation. Transplantation 1991; 51(1): 164.
5. Takada Y, Ozaki N, Ringe B, et al. Receiver operating charac-

teristics (ROC) analysis of the ability of arterial ketone body ratio to predict graft outcome after liver transplantation—its sensitivity and specificity. Transplant Int 1992; 5: 23.
6. Egawa H, Shaked A, Konishi Y, et al. Arterial ketone body ratio in pediatric liver transplantation. Transplantation 1993; 55(3): 522.
7. Oellerich M, Burdelski M, Ringe B, et al. Lignocaine metabolite formation as a measure of pre-transplant liver function. Lancet 1989; 1: 640.
8. Svensson KL, Persson H, Henriksson A, et al. Whole body gas exchange: amino acid and lactate clearance as indicators of initial and early allograft viability in liver transplantation. Surgery 1989; 105: 472.
9. Steltzer H, Tuchy G, Hiesmayr M, et al. Peri-operative liver graft function: monitoring using the relationship between blood glucose and oxygen consumption during anesthesia. Anaesthesia 1992; 47: 955.
10. Takaya S, Nonami T, Selby R, et al. The relationship of systemic hemodynamics and oxygen consumption to early allograft failure after liver transplantation. Transplant Int 1992; 6: 73.
11. Steltzer H, Tuchy G, Hiesmayr M, Zimpfer M. Oxygen kinetics during liver transplantation: the relationship between delivery and consumption. J Crit Care 1993; 8(1): 12.
12. Shaw BW, Wood RP, Gordon RD, et al. Influence of selected patient variables and operative blood loss on six-month survival following liver transplantation. Semin Liver Dis 1985; 5(4): 385.
13. Cuervas-Mons V, Millan I, Gavaler JS, et al. Prognostic value of preoperatively obtained clinical and laboratory data in predicting survival following orthotopic liver transplantation. Hepatology 1986; 6(5): 922.
14. Stock PG, Estrin JA, Fryd DS, et al. Factors influencing early survival after liver transplantation. Am J Surg 1989; 157: 215.
15. Shaw BW, Wood RP, Stratta RJ, et al. Stratifying the causes of death in liver transplant recipients. Arch Surg 1989; 124: 895.
16. Baliga P, Merion RM, Turcotte JG, et al. Preoperative risk factor assessment in liver transplantation. Surgery 1992; 112: 704.
17. Members of the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Crit Care Med 1992; 20: 864.
18. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978; 8: 283.
19. Cuervas-Mons V, Martinez AJ, Dekker A, et al. Adult liver transplantation: an analysis of the early causes of death in 40 consecutive cases. Hepatology 1986; 6(3): 495.
20. Park GR, Gomez Arnau M, Lindop JR, et al. Mortality during intensive care after orthotopic liver transplantation. Anaesthesia 1989; 44: 959.
21. Pugh RNH, Murray Lyon IM, Dawson JL, et al. Transection of the oesophagus for bleeding oesophageal varices. Br J Surg 1973; 60: 646.
22. Shaw BW. Exclusion criteria for liver transplantation. Transplant Proc 1989; 21(3): 3484.
23. Starzl TE, Demetris AJ. Liver transplantation: a 31-year perspective. Part III. Curr Probl Surg 1990; 27: 181.
24. Bronsther O, Fung JJ, Tzakis A, Starzl TE. Prioritization and organ distribution for liver transplantation. JAMA 1994; 271(2): 140.
25. Takaya S, Bronsther O, Iwaki Y, et al. The adverse impact on liver transplantation of using positive cytotoxic crossmatch donors. Transplantation 1992; 53: 400.
26. Marino IR, Weber T, Kang YG, et al. HLA alloimmunization and blood requirements in orthotopic liver transplantation. Transplant Proc 1989; 21(1): 789.
27. Takaya S, Iwaki Y, Starzl TE. Liver transplantation in positive cytotoxic crossmatch cases using FK506, high-dose steroids, and prostaglandin E1. Transplantation 1992; 54: 927.
28. Marino IR, Doyle HR, Starzl TE. Conventional immunosuppressive drugs. In: Thomson AW, Starzl TE, eds. Immunosuppressive drugs: advances in anti-rejection therapy. London: Edward Arnold, 1994.
29. McCauley J, Van Thiel DH, Starzl TE, Puschett JB. Acute and chronic renal failure in liver transplantation. Nephron 1990; 55: 121.
30. McCauley J, Fung JJ, Todo S, et al. Changes in renal function after liver transplantation under FK506. Transplant Proc 1991; 23(6): 3143.
31. Groth CG. Changes in coagulation. In: Starzl TE, ed. Experience in hepatic transplantation. Philadelphia: W.B. Saunders, 1969: 159.
32. Moreau R, Lee SS, Soupison T, et al. Abnormal tissue oxygenation in patients with cirrhosis and liver failure. J Hepatol 1988; 7: 98.
33. Moreau R, Lee SS, Hadengue A, et al. Relationship between oxygen transport and oxygen uptake in patients with cirrhosis: effects of vasoactive drugs. Hepatology 1989; 9(3): 427.
34. Steib A, Freys G, Gohard R, et al. Tissue oxygenation during liver transplantation. Crit Care Med 1992; 20(7): 977.
35. Groh J, Welte B, Gerhard M, et al. Value of total body oxygen consumption as a parameter of graft function after liver transplantation. Transplant Proc 1992; 24(6): 2696.
36. Gubernatis G, Bornscheuer A, Taki Y, et al. Total oxygen consumption, ketone body ratio and a special score as early indicators of irreversible liver allograft dysfunction. Transplant Proc 1989; 21(1): 2279.
37. Shanbhogue RL, Bistrian BR, Jenkins RL, et al. Increased protein catabolism without hypermetabolism after human orthotopic liver transplantation. Surgery 1987; 101(2): 146.
38. Delafosse B, Faure JL, Bouffard Y, et al. Liver transplantation-energy expenditure, nitrogen loss, and substrate oxidation rate in the first two postoperative days. Transplant Proc 1989; 21(1): 2453.
39. Bihari D, Gimson AE, Waterson M, Williams R. Tissue hypoxia during fulminant hepatic failure. Crit Care Med 1985; 13: 1034.

Received 1 July 1993.
Accepted 14 September 1993.