http://www.kidney-international.org

A clinical scoring system highly predictive of long-term kidney graft survival

Yohann Foucher[1,2], Pascal Daguin[1,2], Ahmed Akl[1,2], Michèle Kessler[3], Marc Ladrière[3], Christophe Legendre[4], Henri Kreis[4], Lionel Rostaing[5], Nassim Kamar[5], Georges Mourad[6], Valérie Garrigue[6], François Bayle[7], Bruno H. de Ligny[8], Mathias Büchler[9], Carole Meier[10], Jean P. Daurès[11], Jean-Paul Soulillou[1,2] and Magali Giral[1,2]

[1]Institut de Transplantation et de Recherche en Transplantation-Urologie-Néphrologie, CHU Nantes, IUN and INSERM U643, Nantes, France; [2]Faculté de Médecine, Université de Nantes, Nantes, France; [3]Service de néphrologie et transplantation rénale, CHU Brabois, Nancy, France; [4]Service de transplantation, Hôpital Necker, Paris, France; [5]Unité de transplantation d'Organes, CHU Toulouse, Hôpital Rangueuil, Toulouse, France; [6]Service de néphrologie et transplantation, CHU Montpellier, Montpellier, France; [7]Service de néphrologie et transplantation rénale, CHU Grenoble, Grenoble, France; [8]Service de néphrologie et transplantation rénale, CHU Clémenceau, Caen, France; [9]Service de néphrologie et transplantation rénale, CHU Bretonneau, Tours, France; [10]Service de néphrologie et transplantation rénale, CHU Colmar, Colmar, France and [11]Unité de biostatistique, Institut universitaire de recherche clinique, Montpellier, France

Determining early surrogate markers of long-term graft outcome is important for optimal medical management. In order to identify such markers, we used clinical information from a cross-validated French database (Données Informatisées et VAlidées en Transplantation) of 2169 kidney transplant recipients to construct a composite score 1 year after transplantation. This Kidney Transplant Failure Score took into account a series of eight accepted pre- and post-transplant risk factors of graft loss, and was subsequently evaluated for its ability to predict graft failure at 8 years. This algorithm outperformed the traditional surrogates of serum creatinine and the estimated graft filtration rate, with an area under the receiver–operator characteristic curve of 0.78. Validation on an independent database of 317 graft recipients had the same predictive capacity. Our algorithm was also able to stratify patients into two groups according to their risk: a high-risk group of 81 patients with 25% graft failure and a low-risk group of 236 patients with an 8% failure rate. Thus, although this clinical composite score predicts long-term graft survival, it needs validation in different patient groups throughout the world.

Kidney International (2010) **78,** 1288–1294; doi:10.1038/ki.2010.232; published online 22 September 2010

KEYWORDS: graft survival; prediction models; surrogate end points; transplantation

Correspondence: *Jean-Paul Soulillou, Institut de Transplantation et de Recherche en Transplantation-Urologie-Néphrologie, CHU Hôtel-Dieu 30 boulevard J. Monnet, 44093 Nantes cedex, France.*
E-mail: jean-paul.soulillou@univ-nantes.fr

Received 2 October 2009; revised 21 April 2010; accepted 4 May 2010; published online 22 September 2010

For decades, the conventional end point to estimate the efficacy of new drugs such as cyclosporine has been the 1-year graft survival.[1–3] However, the significant increase in graft survival during the cyclosporine era[4,5] led to the occurrence of acute rejection episodes during the first post-transplant year becoming a more popular end point of efficacy when appreciating the effect of novel therapeutics such as the anti-interleukin-2 receptor antibody,[6] mycophenolate mofetil,[7] and tacrolimus.[8] Nevertheless, the dramatic decrease in the incidence of acute rejection episodes over recent years has forged the challenge to define new pertinent surrogate end points to replace acute rejection or graft survival.[9] More recently, composite end points combining drug efficacy and safety have been used.[10] From a methodological standpoint, the main advantage of mixing different types of failures is the higher frequency of the end point. However, the definition of such a composite variable is often arbitrary and varies from one study to another. Protocol biopsies have also been proposed as a useful short-term end point of long-term outcome.[11,12] Despite a low risk/benefit ratio, biopsies remain a costly and invasive procedure.[13]

More recently, non-invasive (blood or urine) biomarkers have been proposed for the prediction of rejection episodes, such as perforin or granzyme B transcripts,[14] or for the prediction of long-term kidney graft outcome, such as blood levels of soluble CD30 (ref. 15) or donor-specific antibodies.[16] However, these biomarkers still need further validation in large patient cohorts, and often seem too late to be useful as surrogate markers. For example, the mean time to the appearance of donor-specific antibodies is 15–20 months.[17]

Currently, the 6- and 12-month post-transplant serum creatinine (Cr) level is considered to be the simplest marker that is significantly correlated with graft survival.[18–20]

However, although 6–12-month serum Cr level correlates with graft loss, this marker has been shown to be poorly predictive.[21] Several other pre- and post-transplant clinical covariates have also been associated with a reduced long-term graft loss, such as donor and recipient age,[22,23] HLA incompatibilities,[24,25] pre-transplant immunization,[26] and delayed graft function.[27] New early and predictive composite end points are thus needed for a more rapid evaluation of protocols and for decision-making in the clinical management of kidney graft recipients.[28]

In this study, we used a new strategy to characterize a clinical composite score at 1 year, called the Kidney Transplant Failure Score (KTFS). The KTFS takes into account a series of well-accepted pre- and post-transplant risk factors of graft loss. The score itself is calculated using the traditional multivariate Cox model[29] combined with a new statistical approach called the 'time-dependent receiver–operator characteristic (ROC) curves', making it possible to assess the predictive capacity of the surrogate marker that is being evaluated.[30]

RESULTS

Characteristics of the training and test sets

The characteristics of the population are shown in Tables 1 and 2. Among the 2169 training-set patients, the mean follow-up time was 5.1 years (± 2.7) and 380 patients still had a functional renal transplant at 8 years. The other recipients died ($n=72$), returned to dialysis ($n=169$), or were censored ($n=1548$). The censored patients were those who had a functioning graft at their last follow-up, but their follow-up was <8 years. In all, 82% of patients were recipients of a first kidney transplant. The mean donor age was 45.2 years (range 1–83), 63.3% were male and 31.7% died of vascular brain damage. Male recipients accounted for 61.9% of the population. The mean Cr value was 139.1 μmol/l at 6 months and 131.9 μmol/l at 1 year of follow-up.

Certain demographic characteristics of the test population differed from those of the training population. The transplant recipients and donors in the test set were significantly younger than those of the training set, and only 1.9% of patients were re-transplanted in the test set compared with 18.1% in the training set ($P<0.0001$). In addition, donor males were more frequent in the test set. This selection may explain the differences also observed for the parameters collected during the follow-up. For instance, the 1-year estimated graft filtration rate (eGFR) mean was 51.2 ml/min in the training set versus 62.4 ml/min in the test set ($P<0.0001$). We also observed a higher percentage of acute rejection episodes in the training set ($P=0.0072$). These differences endow our analysis with an advantage, because they provide an opportunity to test the robustness of the composite KTFS applied to different populations.

Description of the KTFS

Eight factors were retained after the multivariate analysis of kidney graft survival. The KTFS, weighted on the corrected log hazard ratios of the Cox model, is defined by the following formula:

$$\begin{aligned}\mathrm{KTFS} = &-0.75072 * \mathrm{Cr_D} - 1.02316 * \mathrm{Age_R} \\ &+1.17295 * \mathrm{Ntrans} + 0.22288 * \mathrm{AR} \\ &+0.01881 * \mathrm{Cr_{3m}} + 0.41551 * \sqrt{(\mathrm{Cr_{12m}})} \\ &-0.88001 * \mathrm{Gender} + 0.61121 * \mathrm{Pr_{12m}} \\ &+0.04077 * (\mathrm{Pr_{12m}})^2 + 0.48601 * \mathrm{Gender} * \mathrm{Pr_{12m}} \\ &-0.06115 * \mathrm{Gender} * (\mathrm{Pr_{12m}})^2.\end{aligned}$$

The definitions of the factors are provided in Table 3 and the complete Cox model is described in Table 4, with the corrected weights. When the corrected weight is positive, the KTFS increases with the value of this risk factor. For instance,

Table 1 | Demographic and clinical characteristics of quantitative factors (mean, interquartile interval)

	Learning sample (n=2169)	Testing sample (n=317)	P-value
Recipient age (years)	48.0 (38.0; 58.0)	45.3 (35.0; 55.7)	0.0005
Body mass index (kg/m^2)	23.6 (20.5; 26.0)	23.3 (20.5; 25.5)	0.3178
Delayed graft function (days)	6.4 (1.0; 10.0)	3.0 (1.0; 3.0)	<0.0001
Cold ischemia time (h)	23.5 (16.7; 30.8)	22.7 (17.0; 28.0)	0.3800
HLA-incompatibilities	3.0 (2.0; 4.0)	3.0 (2.0; 4.0)	0.9278
Last donor Cr (μmol/l)	98.6 (67.0; 112.0)	102.5 (70.0; 113.0)	0.4236
Panel reactive antibody on T cells	9.7 (0.0; 0.0)	7.7 (0.0; 0.0)	0.1403
Donor age (years)	45.2 (34.0; 56.0)	38.9 (26.0; 49.0)	<0.0001
3-month Cr (μmol/l)	142.4 (106.0; 165.0)	132.4 (105.0; 149.0)	0.0002
6-month Cr (μmol/l)	139.1 (106.0; 160.5)	131.9 (105.0; 144.0)	0.0183
1-year Cr (μmol/l)	139.8 (106.0; 160.0)	134.6 (106.0; 148.0)	0.1182
1-year eGFR (ml/min)	51.2 (38.9; 62.1)	62.4 (48.5; 73.8)	<0.0001
3-month Pr (g/day)	0.4 (0.1; 0.4)	0.3 (0.0; 0.3)	0.0044
6-month Pr (g/day)	0.4 (0.1; 0.4)	0.3 (0.0; 0.3)	0.0008
1-year Pr (g/day)	0.4 (0.1; 0.4)	0.3 (0.0; 0.3)	0.5469

Abbreviations: Cr, creatinine; eGFR, estimated graft filtration rate; HLA, human leukocyte antigen; Pr, proteinuria.
The demographic and biological characteristics of the training set (n=2169, from the hospitals of Nantes, Paris Necker, Nancy, Toulouse, and Montpellier) and the test set (n=317, from the transplantation centers of Caen, Grenoble, Tours, and Strasbourg). Only quantitative factors are included with the descriptions of means and interval interquartiles. The significance of the differences between means was statistically evaluated using a t-test, except for the body mass index and the delayed graft function, for which the P-values were obtained using a log-rank test.

Table 2 | Demographic and clinical characteristics (qualitative factors)

	Training set (%) (n=2169)	Test set (%) (n=317)	P-value
Male recipients	61.9	65.0	0.3101
Male donors	63.3	68.6	0.0820
First transplantation	81.9	98.1	<0.0001
Acute rejection episode	23.9	16.8	0.0072

The demographic and biological characteristics of the training set (n=2169, from the hospitals of Nantes, Paris Necker, Nancy, Toulouse, and Montpellier) and the test set (n=317, from the transplantation centers of Caen, Grenoble, Tours, and Strasbourg). Only qualitative factors are included with the listing of percentages. The significance of the differences between percentages was statistically evaluated using a χ^2-test.

Table 3 | Values for the risk factors included in the KTFS calculation

Abbreviation	Value
Gender	1 for male recipients and 0 for females
Cr_D	1 if the blood creatinine of the donor is >190 µmol/l, and 0 otherwise
Age_R	1 if the recipient age is >25 years, and 0 otherwise
Ntrans	1 if the number of previous transplantations is >2, and 0 otherwise
Cr_{3m}	Blood creatinine level measured 3-months post-transplantation in µmol/dl
Cr_{12m}	Blood creatinine level measured 1-year post-transplantation in µmol/l
Pr_{12m}	Proteinuria measured 1-year post-transplantation in g/day
AR	1 if an acute rejection occurs in the first year, and 0 otherwise

Abbreviation: KTFS, Kidney Transplant Failure Score; ROC, receiver-operator characteristic. These definitions are chosen in order to maximize the quality of adjustment to the Cox model and to maximize the area under the time-dependent ROC curve at 8 years. To calculate the KTFS for a subject, the abbreviations in the KTFS formula (page 2) are simply replaced by the corresponding value. If at least one variable is missing, the KTFS cannot be calculated.

Table 4 | The multivariate Cox model from the training set (n=2169) and the corresponding corrected weights (corrected log HR) obtained by maximization of the area under the ROC curve for a prognostic at 8 years

Variables	HR	log HR	$CI_{95\%}$ HR	P-value	Corrected weights
Cr_D	0.46	−0.76811	(0.19; 1.13)	0.0920	−0.75072
Age_R	0.37	−0.99039	(0.23; 0.61)	0.0001	−1.02316
Ntrans	2.94	1.07866	(1.68; 5.16)	0.0002	1.17295
AR	1.29	0.25468	(0.94; 1.77)	0.1100	0.22288
Cr_{3m}	0.96	−0.03844	(0.93; 0.99)	0.0098	0.01881
$\sqrt{(Cr_{12m})}$	1.55	0.44031	(1.43; 1.69)	<0.0001	0.41551
Gender	0.42	−0.86668	(0.28; 0.63)	0.0001	−0.88001
Pr_{12m}	1.73	0.55057	(1.19; 2.51)	0.0038	0.61121
Pr^2_{12m}	0.98	−0.02110	(0.93; 1.03)	0.0005	0.04077
Gender * Pr_{12m}	1.66	0.50685	(1.00; 2.75)	0.0490	0.48605
Gender * Pr^2_{12m}	0.93	−0.07623	(0.85; 1.00)	0.0520	−0.06115

Abbreviations: CI, confidence interval; HR, hazard ratio; ROC, receiver-operator characteristic.

the KTFS increases with the number of previous transplantations (Ntrans) and with the blood Cr values at 3 and 12 months (Cr_{3m} and Cr_{12m}). On the contrary, if the corrected weight is negative, the KTFS decreases with the value of each risk factor. For example, the KTFS is lower for male recipients compared with female recipients. More accurately, when the KTFS increases by 0.7, the risk of graft failure increases by twofold (P<0.0001). This increase is represented in Figure 1, for which the score was categorized into 10 classes. The mean KTFS was 4.00 (± 1.34), and 50% of the patients had a KTFS >3.73 (range 1.23–15.33). This distribution is shown in Figure 2.

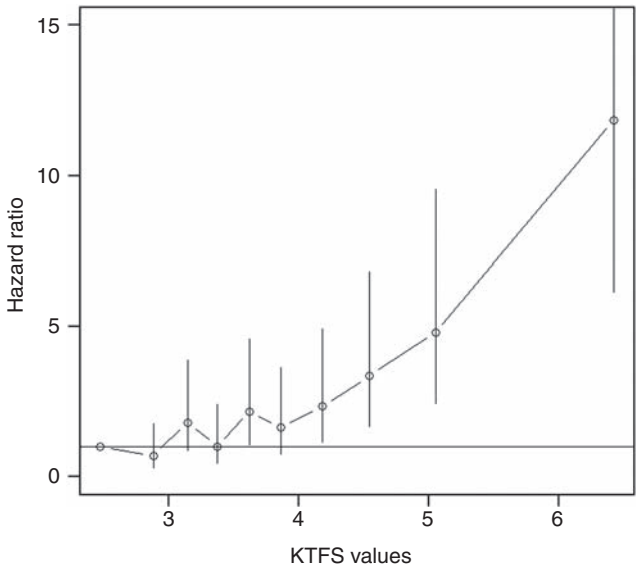

Figure 1 | Evolution of the risk to return in dialysis according to the 10 Kidney Transplant Failure Score (KTFS) categories. The hazard ratios according to the 10 KTFS categories are defined by quantiles. The reference group is the lower KTFS interval with hazard ratio = 1 (horizontal line). The vertical lines represent the 95% confidence intervals of each hazard ratio (n = 2169; 871 patients excluded from the computation of the KTFS owing to the missing values for certain parameters).

Predictive properties of the KTFS in the training set

As explained in the methods section of the time-dependent ROC methodology, the correlation of the KTFS with the risk of graft failure (P<0.0001) is not sufficient to demonstrate its capacity to predict graft survival. Lachenbruch *et al.*[16] also insisted on this distinction between 'correlation' and 'prediction'. Thus, we next evaluated whether it could also predict graft survival. The area under the ROC curve (AUC) of the KTFS was 0.78 (confidence interval $(CI)_{95\%} = (0.73, 0.80)$), indicating that the KTFS was a powerful predictor of graft failure before the eighth anniversary. In comparison, the AUC of the 1-year Cr was 0.73 ($CI_{95\%} = (0.67, 0.76)$), but this predictive capacity was significantly lower than that of the KTFS (P<0.0001, one-tailed test by bootstrap resampling). The prognostic accuracy of the 1-year eGFR was equivalent to that of Cr (AUC = 0.70, $CI_{95\%} = (0.66, 0.75)$). Finally, the ΔCr (from 6 to 12 months) was even less powerful (AUC = 0.60, $CI_{95\%} = (0.58, 0.69)$) than the KTFS (P<0.0001). The corresponding ROC curves are presented in Figure 3a.

As the KTFS is a quantitative variable, it is important to identify subgroups according to their risk of graft failure.

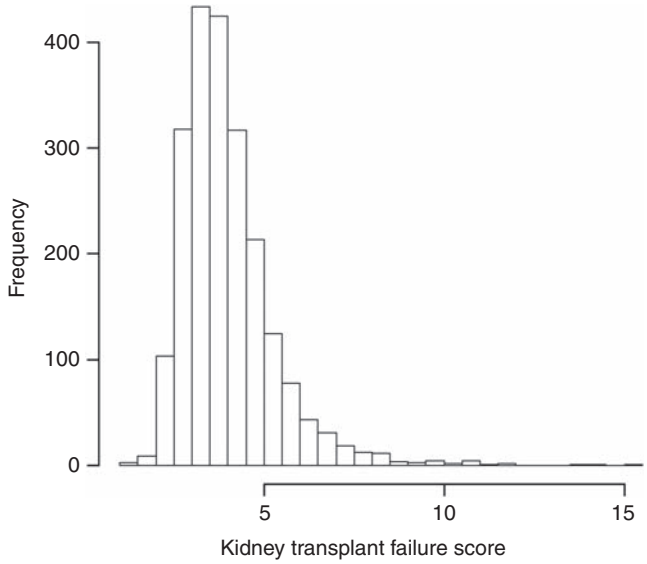

Figure 2 | Distribution of the Kidney Transplant Failure Score (KTFS) calculated in the training population ($n = 2169$). Histogram of the KTFS calculated in the training population, representing the distribution of the composite end point ($n = 2169$).

The KTFS decision threshold was calculated at 4.17, which corresponds to the maximum value of sensitivity (0.72) and specificity (0.71). Two grades were thus defined. Grade I was the group with the lower risk of graft failure (KTFS $\leqslant 4.17$) and grade II was the group with the higher risk (KTFS > 4.17). The graft survival curves are presented in Figure 4a. One can see that at 8 years, the graft failure rate was 7.0% for grade I versus 29.8% for grade II ($P < 0.0001$). In all, 65% of patients were classified as grade I and 35% as grade II. A total of 31% of patients in grade I had kidneys from deceased donors who met the expanded criteria donors, whereas the percentage was 51.8% for patients in grade II ($P < 0.0001$). This difference in distribution of the expanded criteria donor transplants did not explain the predictive capacities of the KTFS, because graft survival between the expanded criteria donor and non-expanded criteria donor transplants was not significant ($P = 0.2520$).

Validation in the independent test set

The capacity of the KTFS to predict 8-year graft failure was also analyzed for the test set (317 independent patients from four different centers). Figure 3b shows the ROC curves and also that the KTFS was still the predictor with the best accuracy (AUC = 0.78, $CI_{95\%} = (0.71, 0.86)$). It outperformed the Cr (AUC = 0.67, $CI_{95\%} = (0.58, 0.78)$), the eGFR (AUC = 0.67, $CI_{95\%} = (0.56, 0.78)$), and the ΔCr between 6 and 12 months post-transplant (AUC = 0.61, $CI_{95\%} = (0.53, 0.79)$). The prognostic capacity of the KTFS was thus still higher than the 1-year Cr ($P = 0.0050$), the 1-year eGFR ($P = 0.0083$), and the ΔCr ($P = 0.0300$).

Figure 4b also shows that the threshold of 4.17 discriminated the two groups according to graft survival.

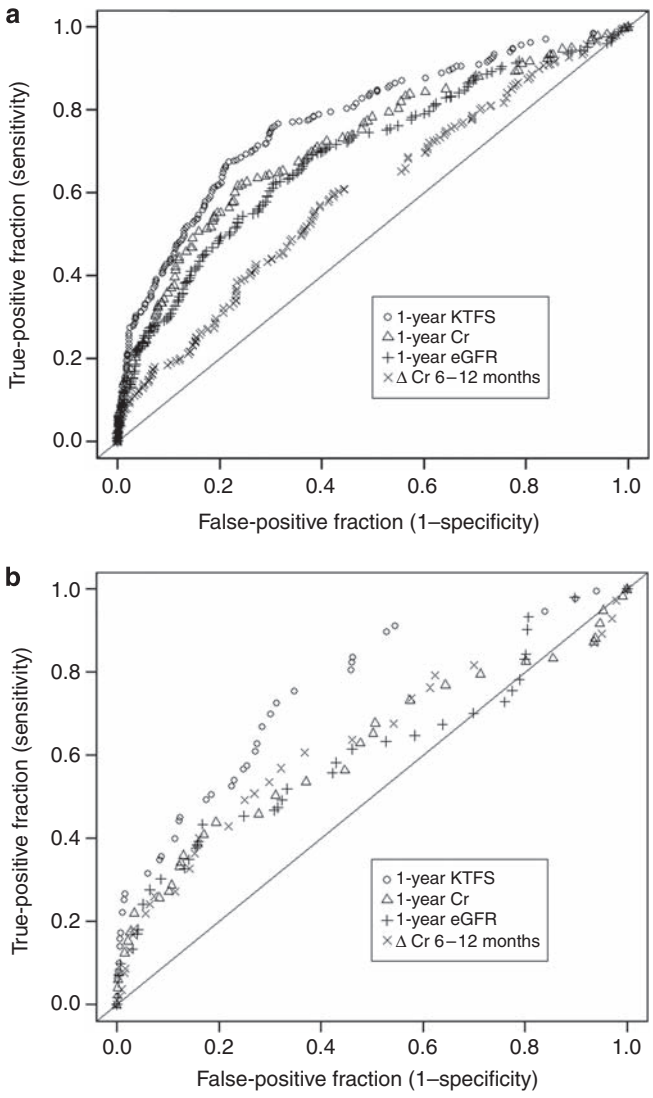

Figure 3 | 8-Year predictive properties of the Kidney Transplant Failure Score (KTFS). (**a**) Receiver–operator characteristic (ROC) curves for 8-year predictions to evaluate the prognostic capacity of the different markers. The KTFS (black line, AUC = 0.78) is the most predictive marker. The usual surrogate markers, 1-year Cr (red line, AUC = 0.73) and 1-year estimated graft filtration rate (eGFR; green line, AUC = 0.70), show lower predictive capacities. The evolution of creatinine (Cr) between 6 and 12 months is the worst marker (blue lines, AUC = 0.60). The area under the ROC curve of the KTFS is higher than those of the others ($P < 0.0001$, one-tailed test by bootstrap resampling). (**b**) ROC curves for 8-year prediction applied to the independent testing set. The results are similar to those obtained in the training set. The KTFS (black line, AUC = 0.78) is the most predictive marker. The usual surrogate markers, 1-year Cr (red line, AUC = 0.67) and 1-year eGFR (green line, AUC = 0.67), show lower predictive capacities. The evolution of Cr between 6 and 12 months is the worst marker (blue line, AUC = 0.61). The area under the ROC curve of the KTFS is higher than those of the others ($P < 0.05$, one-tailed test by bootstrap resampling).

The graft failure rate of patients in grade I was 8.2% at 8 years versus 25.3% in grade II ($P < 0.0001$). In this test set, 74.5% of patients were classified as grade I and 25.5% as grade II.

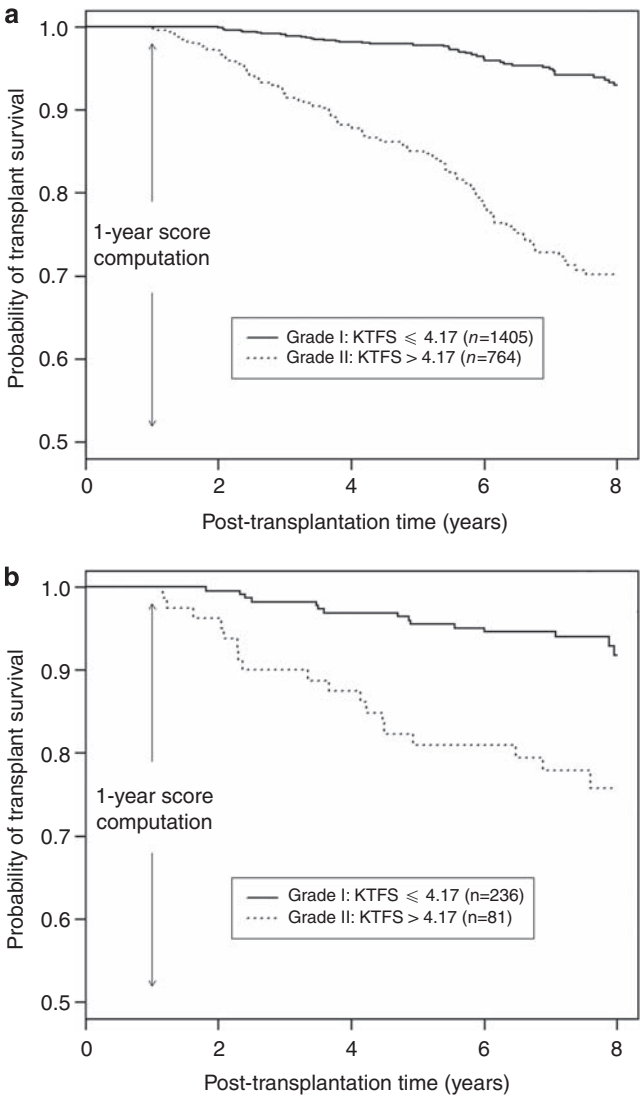

Figure 4 | Graft survival functions according to the Kidney Transplant Failure Score (KTFS) grades (threshold calculated at 4.17 by maximizing both the sensitivity and the specificity). (**a**) Estimation of transplant survival during the first 8 years post-transplantation and according to the two grades defined at 1-year post-transplantation. The dark line indicates the low-risk subgroup (grade I), with a 93% probability of having a functional kidney at 8 years post-transplantation. The survival of the grade II subgroup (red line) is lower, with an 8-year probability of survival estimated at 70.2%. The difference between the two curves is highly significant (log-rank test, $P<0.00001$). (**b**) These are the same survival curves, but estimated from the test set. The dark line represents the low-risk subgroup (grade I), with a 91.8% probability of having a functional kidney at 8 years post-transplantation. The survival of the grade II subgroup (red line) is lower with an 8-year probability of survival estimated at 75.8%. The difference between the two curves is still highly significant (log-rank test, $P<0.00001$).

DISCUSSION

The early prediction of long-term outcome in kidney transplantation has become a major issue, not only for guiding clinical management but also for shortening the survey period. In this paper, we present a non-invasive composite clinical tool, called the KTFS, as a surrogate for long-term graft outcome. Although our paper describes the KTFS in a given clinical context, the method has the potential to be substantially improved and adapted to other types of transplantation or with other cohorts (different variables or patients).

The KTFS was calculated from 2169 kidney recipients. It combines risk factors of graft loss, including Cr values (at 3 and 12 months), proteinuria value (at 12 months), recipient age and gender, number of previous transplantations, donor Cr value, and rejection episodes. All these parameters were collected within the first post-transplant year. The KTFS has been shown to have properties that are useful for the prediction of graft failure up to the eighth post-transplant year. A patient with a KTFS <4.17 is predicted to have a functioning graft on the eighth anniversary of transplantation, with a 93% chance of accuracy. However, a patient with a KTFS >4 has a 29.8% risk of renal failure. We also compared the prognostic capacities of the KTFS with various other markers of kidney graft function, namely 1-year Cr, 1-year eGFR, as well as the change in Cr level between 6 and 12 months. The capacity of the KTFS to predict graft failure was significantly better. It comes as no surprise that the KTFS works better than the 1-year Cr, eGFR, or Cr change between 6 and 12 months, because these variables are included in the KTFS. However, this is the first time that a composite approach has been proposed and evaluated properly.

In contrast to the study by Kaplan *et al.*,[21] we showed that Cr and its eGFR are acceptable predictors of graft survival. This difference is probably because of the statistical modelling, as Kaplan *et al.*[21] used conventional ROC curves. In contrast, we have used in the present paper an ROC method adapted to survival data.[30]

The robustness of the results obtained from the training set was consolidated by including a test set of 317 patients, which was completely independent. This validation analysis yielded similar results even though the patients differed for several parameters included in the KTFS, such as younger age of recipient and donor, higher percentage of male donors, and less number of re-transplanted patients.

The analysis has been also developed in accordance with the intended clinical application, and the KTFS can be used with a threshold value for medical decision-making. The KTFS could thus be adapted with the aim of providing a simple and useful tool for physicians to predict, at 1 year, the future outcome of the patient.

One limitation of the KTFS is that it is restricted to patients who have maintained their graft function for at least 1 year post-transplantation. For this reason, patients who had lost their kidney transplants during the first year were excluded from this study (22 deaths with a functioning graft and 216 transplant failures). However, when we analyzed the causes of graft failure during the first year, we identified specific events including 84 vascular thromboses, 45 irreversible rejection episodes, 15 early relapses of initial disease, 16 never-functioning kidneys, 10 severe surgical

complications, 23 severe dysfunctions, and 43 failures of unknown cause. These early graft failures are clearly different from those associated with long-term survival. Although the aim of the KTFS described here was to predict long-term graft failure as early as possible, it would also be feasible to test a score based on parameters collected at 3 or 6 months post-transplant or even only on pre-transplant parameters.

In summary, we have built a new, simple, and non-invasive clinical score to more accurately predict the long-term graft outcome. We believe that this score could be used as a new end point for clinical trials or as a decision tool in the clinical management of kidney transplant recipients. But, even if the KTFS was already validated on an independent cohort, it would also be useful to see some further validations and improvements in different cohorts throughout the world.

MATERIALS AND METHODS

Study sample and Données Informatisées et VAlidées en Transplantation (DIVAT) data bank

We based our analysis on data regarding kidney transplant recipients prospectively collected between January 1996 and November 2007 and computerized in the DIVAT data bank. The biological and clinical data within DIVAT are prospectively recorded according to a common thesaurus at the participating centers (since 1996 for Nantes and Paris Necker, since 1998 for Nancy, and since 2003 for Toulouse and Montpellier). Yearly audits between all participating centers showed $<1\%$ error in the collected data. Recorded parameters included donor age, gender, and last donor Cr level before kidney retrieval, as well as recipient age, weight, size, gender, previous transplantations, cold ischemia time, delayed graft function (defined by the time taken to attain a calculated clearance $\geqslant 10$ ml/min), highest level of panel reactive antibodies on T cells, and HLA A-B-DR incompatibilities.[27] Post-transplant acute rejection episodes are additionally collected in real time. The French law does not authorize the storage of patient ethnicity. Finally, Cr and daily proteinuria were also recorded at 3 and 6 months as well as annually. The eGFR was calculated using the four-variable Modification of Diet in Renal Disease formula.[31] Only adult recipients of organs from heart-beating deceased donors with a functional transplant on the first anniversary of their transplantation were included in the study. With regard to all these selection criteria, 2169 patients with all parameters collected were included.

Independent test sample for external validation

An additional sample set of 317 independent kidney transplantations was used to validate the results obtained from the training sample (DIVAT). For this test sample, the same data (double-checked) from four other French transplantation centers (Caen, Grenoble, Tours, and Strasbourg) were collected between April 1995 and January 2006.

Calculation of the KTFS

The KTFS is a score calculated to optimize the prediction of the time between transplantation and graft failure (that is, return to dialysis). Death is considered as a censored event. A multivariate Cox model[29] was performed to initially integrate all significant risk factors ($P<0.15$). The proportional hazard assumption was tested using weighted residuals.[32] As the objective was not to interpret the different factors but to use them to construct a score, they were included in the model using more flexible transformations (polynomial and logarithmic functions, interactions between variables) than those in classical survival analyses. However, if no transformation was adequate to respect the log linearity, the quantitative variables were categorized. These transformations or cutoff estimations were performed to maximize the partial likelihood of the model. It must be noted that we did not use the usual first error risk threshold of 0.05, a value adapted to prove a correlation between the covariate and the survival with a very low risk of error. In fact, our objective was to obtain the most complete 1-year composite end point. The resulting KTFS is equal to the sum of the risk factor values, multiplied by the corresponding log hazard ratios. Thus, as the graft failure risk increases, so does the KTFS. To maximize the predictive capacity of this score, the weight (i.e., log hazard ratios) of each parameter was corrected. The corrected weights were obtained by maximizing the area under the ROC curve (see the following paragraph).

Testing the predictive capacity of the KTFS

Diagnostic ROC curves. Usually, the ROC analysis assumes that the disease status does not change over time, which is the case when the diagnostic marker is measured at the same time as the disease. However, this was not the case in our study, in which all patients had a functioning kidney transplant, but this could fail during the course of their follow-up, requiring adaptation of the ROC curve for prognosis.

The adaptation of ROC. The time-dependent ROC method was used here to analyze the capacity of a marker to predict a kidney transplant failure. This method is based on the classical Kaplan and Meier estimator[33] converted using the Akritas approach.[34] The procedure takes into account the complexity of survival analysis (censoring of the follow-up). Within this context, we addressed the question of how well the KTFS, assessed at 1-year post-transplantation, could discriminate subjects who returned to dialysis before their eighth transplantation anniversary from those who did not. We chose the maximum prognostic time at 8 years, as only a few patients were still being followed up after this time. We estimated the decision threshold values of the KTFS by maximizing the sensitivity and the specificity. For prognosis on the eighth anniversary of transplantation, the sensitivity represents the proportion of at-risk patients among those who have returned to dialysis before this date. The specificity is the proportion of risk-free patients among those who have not returned to dialysis before this eighth anniversary.

Comparison with traditional markers

Finally, as the 1-year serum Cr and eGFR levels, and evolution of the Cr between 6 months and 1-year post-transplantation (ΔCr) have previously been proposed as markers of late graft loss,[21,35] we performed ROC analysis for these three markers so as to compare their prediction capacities with that of the KTFS. These comparisons were made using 1000 non-parametric bootstrap re-samples, in which the differences between areas under the curve were calculated at each iteration (one-tailed test).

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

This work was partly supported by the RTRS, the Fondation de Coopération Scientifique 'CENTAURE' and Roche Laboratory. YF is a

recipient of a post-doctorate Grant in biostatistics from RTRS 'CENTAURE'. We also thank Dr Joanna Ashton-Chess for editing the paper.

REFERENCES

1. Canadian Multicentre Transplant Study Group. A randomized clinical trial of cyclosporine in cadaveric renal transplantation. *N Engl J Med* 1983; **309**: 809–815.
2. Merion RM, White DJ, Thiru S *et al.* Cyclosporine: five years' experience in cadaveric renal transplantation. *N Engl J Med* 1984; **310**: 148–154.
3. Hall BM, Tiller DJ, Hardie I *et al.* Comparison of three immunosuppressive regimens in cadaver renal transplantation: long-term cyclosporine, short-term cyclosporine followed by azathioprine and prednisolone, and azathioprine and prednisolone without cyclosporine. *N Engl J Med* 1988; **318**: 1499–1507.
4. Suthanthiran M, Strom TB. Renal transplantation. *N Engl JMed* 1994; **331**: 365–376.
5. Hariharan S, Johnson CP, Bresnahan BA *et al.* Improved graft survival after renal transplantation in the United States, 1988–1996. *N Engl J Med* 2000; **342**: 605–612.
6. Soulillou JP, Cantarovich D, Le Mauff B *et al.* Randomized controlled trial of a monoclonal antibody against the interleukin-2 receptor (33B3.1) as compared with rabbit antithymocyte globulin for prophylaxis against rejection of renal allografts. *N Engl J Med* 1990; **322**: 1175–1182.
7. Sollinger HW. Mycophenolate mofetil for the prevention of acute rejection in primary cadaveric renal allograft recipients. US Renal Transplant Mycophenolate Mofetil Study Group. *Transplantation* 1995; **60**: 225–232.
8. Pirsch JD, Miller J, Deierhoi MH *et al.* A comparison of tacrolimus (FK506) and cyclosporine for immunosuppression after cadaveric renal transplantation. FK506 Kidney Transplant Study Group. *Transplantation* 1997; **63**: 977–983.
9. Meier-Kriesche HU, Schold JD, Kaplan B. Long-term renal allograft survival: have we made significant progress or is it time to rethink our analytic and therapeutic strategies? *Am J Transplant* 2004; **4**: 1289–1295.
10. Tedesco-Silva H, Szakaly P, Shoker A *et al.* FTY720 versus mycophenolate mofetil in *de novo* renal transplantation: six-month results of a double-blind study. *Transplantation* 2007; **84**: 885–892.
11. Nankivell BJ, Borrows RJ, Fung CL *et al.* Natural history, risk factors, and impact of subclinical rejection in kidney transplantation. *Transplantation* 2004; **78**: 242–249.
12. Hertig A, Anglicheau D, Verine J *et al.* Early epithelial phenotypic changes predict graft fibrosis. *J Am Soc Nephrol* 2008; **19**: 1584–1591.
13. Wilkinson A. Protocol transplant biopsies: are they really needed? *Clin J Am Soc Nephrol* 2006; **1**: 130–137.
14. Li B, Hartono C, Ding R *et al.* Noninvasive diagnosis of renal-allograft rejection by measurement of messenger RNA for perforin and granzyme B in urine. *N Engl J Med* 2001; **344**: 947–954.
15. Pelzl S, Opelz G, Wiesel M *et al.* Soluble CD30 as a predictor of kidney graft outcome. *Transplantation* 2002; **73**: 3–6.
16. Lachenbruch PA, Rosenberg AS, Bonvini E *et al.* Biomarkers and surrogate endpoints in renal transplantation: present status and considerations for clinical trial design. *Am J Transplant* 2004; **4**: 451–457.
17. Hourmant M, Cesbron-Gautier A, Terasaki PI *et al.* Frequency and clinical implications of development of donor-specific and non-donor-specific HLA antibodies after kidney transplantation. *J Am Soc Nephrol* 2005; **16**: 2804–2812.
18. Giral M, Taddei C, Nguyen JM *et al.* Single-center analysis of 468 first cadaveric kidney allografts with a uniform ATG-CsA sequential therapy. *Clin Transplant* 1996; 257–264.
19. Nicol D, MacDonald AS, Lawen J *et al.* Early prediction of renal allograft loss beyond one year. *Transplant Int* 1993; **6**: 153–157.
20. Hariharan S, McBride MA, Cherikh WS *et al.* Post-transplant renal function in the first year predicts long-term kidney transplant survival. *Kidney Int* 2002; **62**: 311–318.
21. Kaplan B, Schold J, Meier-Kriesche HU. Poor predictive value of serum creatinine for renal allograft loss. *Am J Transplant* 2003; **3**: 1560–1565.
22. de Fijter JW, Mallat MJ, Doxiadis II *et al.* Increased immunogenicity and cause of graft loss of old donor kidneys. *J Am Soc Nephrol* 2001; **12**: 1538–1546.
23. Meier-Kriesche HU, Ojo AO, Cibrik DM *et al.* Relationship of recipient age and development of chronic allograft failure. *Transplantation* 2000; **70**: 306–310.
24. Cecka JM, Terasaki PI. The UNOS Scientific Renal Transplant Registry–1991. *Clin Transplant* 1991; 1–11.
25. Opelz G, Döhler B. Effect of human leukocyte antigen compatibility on kidney graft survival: comparative analysis of two decades. *Transplantation* 2007; **84**: 137–143.
26. Terasaki PI. Humoral theory of transplantation. *Am J Transplant* 2003; **3**: 665–673.
27. Giral-Classe M, Hourmant M, Cantarovich D *et al.* Delayed graft function of more than six days strongly decreases long-term survival of transplanted kidneys. *Kidney Int* 1998; **54**: 972–978.
28. Rabb H. Surrogate endpoints for long-term allograft survival: an AST/ASTS sponsored consensus conference. *Am J Transplant* 2004; **4**: 1170.
29. Cox D. Regression models and life-tables. *J R Stat Soc Ser B* 1972; **34**: 187–229.
30. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56**: 337–344.
31. Levey A, Coresh J, Greene T *et al.* Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann Intern Med* 2006; **145**: 247–254.
32. Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81**: 515–526.
33. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. *J Am Stat Assoc* 1958; **53**: 457–481.
34. Akritas M. Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann Stat* 1994; **22**: 1299–1327.
35. Hariharan S, McBride MA, Cohen EP. Evolution of endpoints for renal transplant outcome. *Am J Transplant* 2003; **3**: 933–941.