Supplementary Materials for

Identification of a Personalized eGFR Threshold Improves the Prediction of Kidney Failure Risk After Transplantation.

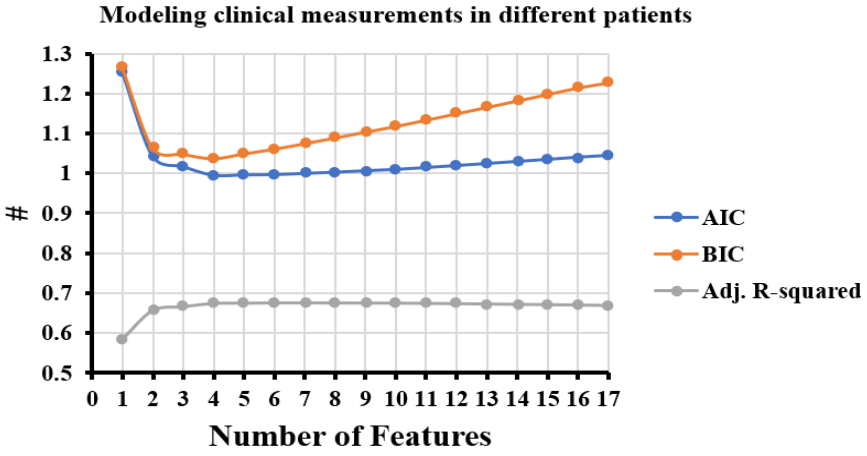
Symeon V. Savvopoulos[a,*], Irina Scheffner[b,*], Andreas Reppas[c], Wilfried Gwinner[b,+], Haralampos Hatzikirou[a,d,+].

a. Department of Mathematics, Khalifa University, Abu Dhabi, United Arab Emirates.
b. Department of Nephrology and Hypertension, Hannover Medical School, Hannover, Germany.
c. Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität, Berlin, Germany
d. Center for Interdisciplinary Digital Sciences (CIDS), Department Information Services and High Performance Computing (ZIH), TUD Dresden University of Technology, 01062 Dresden, Germany
* These authors contributed equally,
+ These authors jointly supervised this work
Corresponding author: Professor Haralampos Hatzikirou (haralampos.hatzikirou@ku.ac.ae)

• Supplementary Methods

1. Feature selection and multi-linear regression.

Patients without rejection from the training set were used for feature selection and prediction of GFR after 365 days postransplantation. The inputs were 1) patient age, 2) patient sex, 3) donor's age, 4) donor's sex, 5) deceased donor vs. living donor 6) cold ischemia time (hours), 7) initial function vs. delayed function of the graft, 8) HLA-mismatch at locus A, 9) HLA-mismatch at locus B, 10) HLA-mismatch at locus DR, 11) panel reactive antibodies (%), 12) blood transfusions before transplantation, 13) number of pregnancies, 14) number of previous transplantations, 15) highest GFR within six weeks after transplantation (ml/min/1.73 m^2), and 16) percentage change between the highest GFR within six weeks and the GFR at 100 d. All inputs were used in the AIC, BIC and adjusted R^2 for feature selection in Python. The results of the feature selection are shown in Supplementary Figure 1.



Supplementary Figure 1: Feature selection for 1-year GFR prediction. Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and adjusted R^2 as functions of the number of candidate features used to predict GFR at 365 days in non-rejection patients. Performance plateaus after three features, leading to the selection of highest GFR within six weeks, percentage change between six-week and 100-day GFR, and donor age as the most informative predictors.

Supplementary Table 1 contains the number of the features and the index of the features for the first five numbers.

Supplementary Table 1: Ranked importance of clinical predictors retained during feature selection. Summary of the most informative features selected in the first five steps of the forward feature-selection process for predicting eGFR at 365 days. The table lists the feature indices added at each step, highlighting that early post-transplant GFR (GFR within six weeks), its percentage change by day 100, donor age, pre-transplant blood transfusions, and HLA-B mismatch are the leading contributors. These results underpin the choice of three key features used in the final multilinear regression model.

Number of features	Index of the features
1	GFR_{in}[1]
2	GFR_{in}, % of GFR change
3	GFR_{in}, % of GFR change, donor's age
4	GFR_{in}, % of GFR change, donor's age, blood transfusions before transplantation
5	GFR_{in}, % of GFR change, donor's age, blood transfusions before transplantation, HLA-mismatch at locus B

According to Supplementary Figure1 and Supplementary Table 1, the selected number of features was three as it showed a good performance, and each criterion value varied insignificantly when the number of features increased. Thus, the most important features were the highest GFR within six weeks after transplantation (x_1), the percentage change between the highest GFR within six weeks and the GFR at 100 d (x_2), and donor's age (x_3).

The corresponding coefficients of the features with reference to multi-linear regression, and the intercept for predicting the GFR 365 days post transplantation, are shown in Supplementary Table 2.

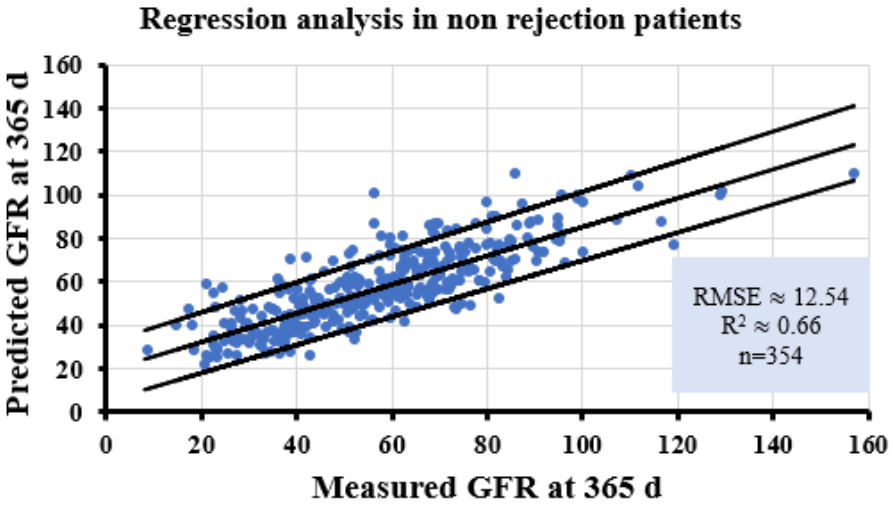
Supplementary Table 2: Multilinear regression coefficients for one-year eGFR prediction. Ordinary least-squares regression estimates, standard errors, test statistics, p-values, and confidence intervals for the intercept and three selected predictors: highest GFR within six weeks, percentage GFR change to day 100, and donor age. All coefficients are highly significant, with positive effects of early GFR level and GFR increase and a negative effect of donor age on one-year eGFR. These parameter estimates form the backbone of the endpoint regression used to initialize the dynamic model.

Feature	Value	Std err	T	P>|t|	95% CI
Intercept	19.2214	3.543	5.425	0.000	[12.253 26.190]
x_1	0.7534	0.033	22.808	0.000	[0.688 0.818]
x_2	20.5617	2.437	8.436	0.000	[15.768 25.355]
x_3	-0.1543	0.046	-3.333	0.001	[-0.245 -0.063]

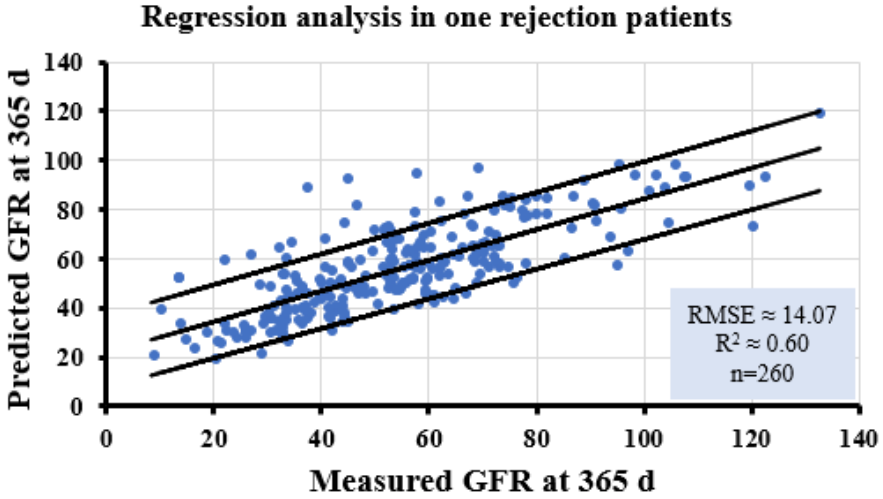
The results of the multilinear regression (Supplementary Table 2) are shown in patients without rejection from the training set in Supplementary Figure 2. By using the same coefficients in

[1] GFR_{in}: Highest GFR within the first 6 weeks posttransplantation

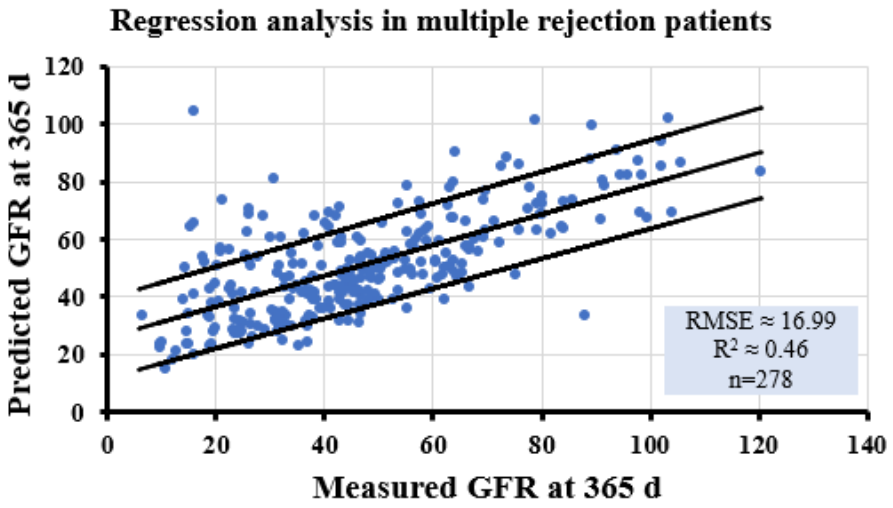
patients with one rejection and multiple rejections, the regression lines are shown in Supplementary Figures 3 and 4.



Supplementary Figure 2: Multilinear regression of 1-year GFR in non-rejection patients. Observed versus predicted GFR at 365 days in the training cohort of patients without rejection, as obtained from the multilinear regression model. The regression line and prediction interval illustrate good agreement between observed and predicted values, supporting the use of this model as the baseline for estimating downstream dynamic model parameters.



Supplementary Figure 3: Multilinear regression of 1-year GFR in patients with one rejection. Observed versus predicted GFR at 365 days in training-set patients who experienced a one rejection episode. The regression line and prediction interval are shown using the coefficients derived from the non-rejection cohort. Compared with non-rejection patients, the model exhibits reduced R^2 and increased error, indicating decreased predictive accuracy in the presence of rejection.



s

Supplementary Figure 4: Multilinear regression of 1-year GFR in patients with multiple rejections. Observed versus predicted GFR at 365 days in training-set patients with multiple rejection episodes using the same regression coefficients as in non-rejection patients. The further decrease in R^2 and increase in residual variability highlight the limitations of linear regression in capturing graft function under multiple rejection and motivate the use of dynamic modeling.

The results showed that linear regression can be used for GFR prediction at 365 days post transplantation in patients without rejection whereas R^2 significantly decreases and RMSE increases in patients with one and with multiple rejection.

Lastly, three different linear mixed models were used. Linear mixed models (LMMs) extend linear regression to data with grouped or repeated measurements (e.g., patients measured over time). They include fixed effects for population-level relationships and random effects for subject-specific deviations, which properly account for within-group correlation and allow partial pooling[1,2].

The first linear mixed model (LMM1) uses the GFR value at 42 days and captures the outcome for patient i at 365 days. Its equation is the following:

$$y_{\mathrm{it}} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1\mathrm{i}}) \cdot Time_{\mathrm{c,it}} + \varepsilon_{it} \qquad \text{(S. 1)}$$

Where the centered $Time_{\mathrm{c,it}}$ equals:

$$Time_{c,i365} = 1 - \frac{1}{2}\left(\frac{42}{365} + 1\right) = 0.4424 \qquad \text{(S. 2)}$$

u_{0i} and u_{1i} are equal to:

$$[u_{0i}\ u_{1i}] \sim N\left([0\ 0], \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}\right) \qquad \text{(S. 3)}$$

and ε_{it} is equal to:

$$\varepsilon_{it} \sim N(0, \sigma^2) \quad \text{(S. 4)}$$

τ_0 , τ_1, and τ_{01} are equal to :

$$\tau_0^2 = Var(u_{0i}) \quad \text{(S. 5)}$$

$$\tau_1^2 = Var(u_{1i}) \quad \text{(S. 6)}$$

$$\tau_{01} = Cov(u_{0i}, u_{1i}) \quad \text{(S. 7)}$$

The second linear mixed model (LMM2) uses GFR values at 42 and 100 days, along with baseline covariates, to capture the outcome for patient i at 365 days. Its equation is the following:

$$\mathrm{y_{it}} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1\iota}) \cdot Time_{\mathrm{c,it}} + \sum_k \delta_k Time_{\mathrm{c,it}}\, x_{ik} + a_1 \mathrm{y_{i42,\ t=365}} + a_2 \mathrm{y_{i100,\ t=365}} + \varepsilon_{it} \quad \text{(S. 8)}$$

Where x_{ik} are the baseline predictors, $Time_{c,it}x_{ik}$ are included when time interactions are enabled, and α_1 and α_2 are active at time t=365 days.

Again, the centered $Time_{\mathrm{c,it}}$ equals:

$$Time_{c,i365} = 1 - \frac{1}{2}\left(\frac{42}{365} + 1\right) = 0.4424 \quad \text{(S. 9)}$$

u_{0i} and u_{1i} are equal to:

$$[u_{0i}\ u_{1i}] \sim N\left([0\ 0], \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}\right) \quad \text{(S. 10)}$$

and ε_{it} is equal to:

$$\varepsilon_{it} \sim N(0, \sigma^2) \quad \text{(S. 11)}$$

τ_0 , τ_1, and τ_{01} are equal to:

$$\tau_0^2 = Var(u_{0i}) \quad \text{(S. 12)}$$

$$\tau_1^2 = Var(u_{1i}) \quad \text{(S. 13)}$$

$$\tau_{01} = Cov(u_{0i}, u_{1i}) \quad \text{(S. 14)}$$

Lastly, the third linear mixed model (LMM3) uses the values of GFR at 42 and the percentage of GFR change between 42 and 100 days, and baseline covariates to capture the outcome for patient i at time 365 days. Its equation is the following:

$$y_{it} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i}) \cdot Time_{c,it} + \sum_k \gamma_k x_{ik} + \sum_k \delta_k Time_{c,it} x_{ik} + a_1 y_{i42,\ t=365} + a_2 perc_{i100,\ t=365} + \varepsilon_{it} \quad \text{(S. 15)}$$

Where x_{ik} are the baseline predictors, $Time_{c,it}x_{ik}$ are included when time interactions are enabled, and α_1 and α_2 are active at time t=365 days.

Again, the centered $Time_{c,it}$ equals:

$$Time_{c,i365} = 1 - \frac{1}{2}\left(\frac{42}{365} + 1\right) = 0.4424 \quad \text{(S. 16)}$$

u_{0i} and u_{1i} are equal to:

$$[u_{0i}\ u_{1i}] \sim N\left([0\ 0], \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}\right) \quad \text{(S. 17)}$$

and ε_{it} is equal to:

$$\varepsilon_{it} \sim N(0, \sigma^2) \quad \text{(S. 18)}$$

τ_0 , τ_1, and τ_{01} are equal to :

$$\tau_0^2 = Var(u_{0i}) \quad \text{(S. 19)}$$

$$\tau_1^2 = Var(u_{1i}) \quad \text{(S. 20)}$$

$$\tau_{01} = Cov(u_{0i}, u_{1i}) \quad \text{(S. 21)}$$

The non-rejection patients in the training cohort were used to evaluate the model's performance using 5-fold cross-validation. Supplementary Table 3 shows the performances of the three LMMs in non-rejection, one rejection, and multiple rejection patients, and also contains the performance of our end-point regression model with feature selection.

Supplementary Table 3: Comparison of linear mixed models and regression in non-rejection patients. Ranges of R^2 and RMSE values for three linear mixed-effects models (LMM1–LMM3) versus the simpler multilinear regression in non-rejection patients using 5-fold cross-validation. While more complex LMMs improve upon LMM1, they do not outperform the regression model, which achieves $R^2 = 0.66$ with competitive RMSE. This supports the choice of the parsimonious regression-based approach for one-year eGFR prediction.

Method	**R^2 range**	**RMSE range**
LMM1	0.353 – 0.672	11.310 – 15.173
LMM2	0.579 – 0.678	11.240 – 14.540
LMM3	0.560 – 0.704	11.748 – 13.591
Regression	0.660	12.540

As for the one rejection patients, the results are provided in Supplementary Table 4.

Supplementary Table 4: Performance of linear mixed models versus regression in one-rejection patients. R^2 and RMSE values for LMM1–LMM3 and the regression model in patients with a single rejection episode. The maximum R^2 and minimum RMSE values of the regression model show that mixed effect complexity in the added model does not produce better predictions for this subgroup. This confirms the original regression formulation is still valid despite variability due to rejection being included.

Method	**R^2**	**RMSE**
LMM1	0.432	17.082
LMM2	0.589	15.057
LMM3	0.541	15.901
Regression	0.600	14.070

As for the one rejection patients, the results are provided in Supplementary Table 5.

Supplementary Table 5: Performance of linear mixed models versus regression in multiple-rejection patients. Comparison of R^2 and RMSE for LMM1–LMM3 and the regression model in patients with multiple rejection episodes. Predictive performance is generally reduced compared with non-rejection patients, reflecting greater clinical complexity, but the regression model still performs comparably or better than mixed models. Overall, these analyses show no clear advantage of LMMs over the baseline regression.

Method	**R^2**	**RMSE**
LMM1	0.320	18.787
LMM2	0.483	17.938
LMM3	0.408	19.190
Regression	0.460	16.990

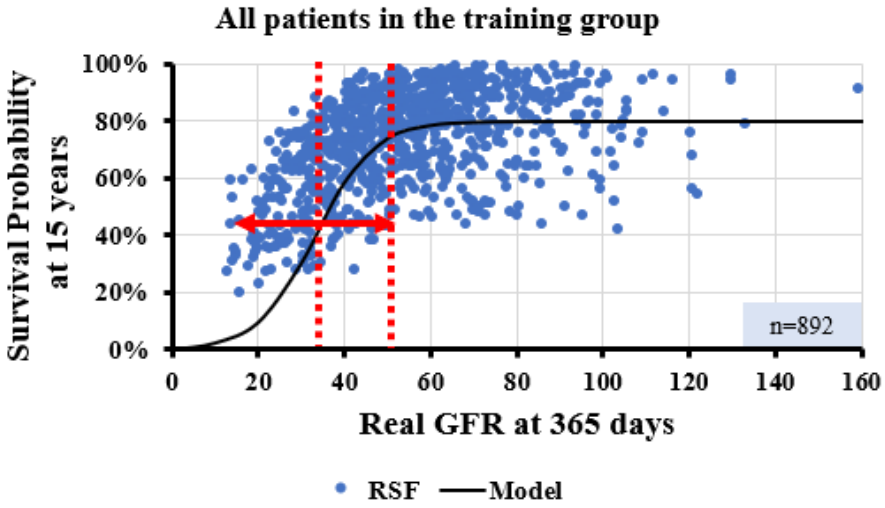
Across identical patient-grouped cross-validation splits, LMM1—the simplest mixed model—performed worst (lowest R^2, highest RMSE). In contrast, LMM2 and LMM3 achieved R^2 and RMSE values that were very close to those of our initial model, with only small differences. These results indicate that the linear mixed-model formulations did not improve 1-year eGFR prediction over the initial model, and therefore, we retain the initial model as a useful and parsimonious choice.

2. Graft Survival probabilities for all patients of the training group.

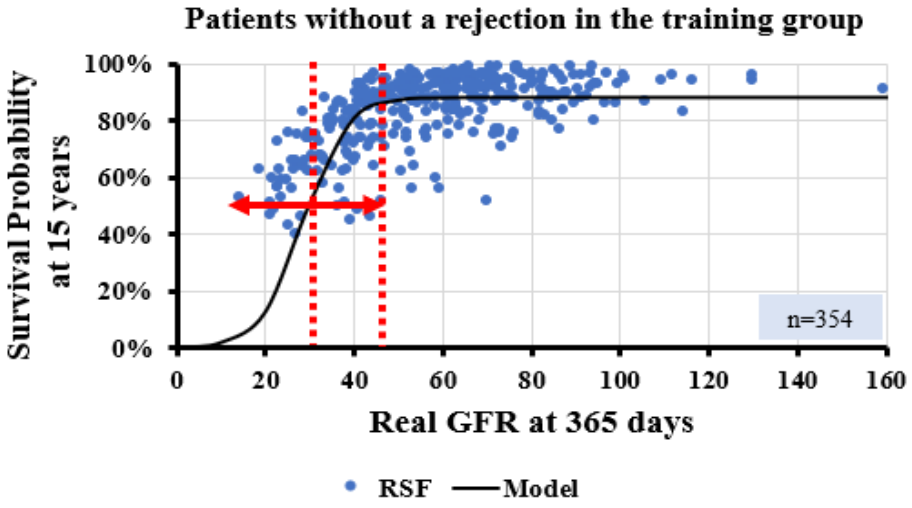
Here, we employ a survival random forest analysis developed by Gwinner et al. to assess 15-year graft survival outcomes[3]. By leveraging an extensive dataset from a single center[4], observed over up to 15 years and including several patients, the study developed a machine-learning model to predict clinically meaningful graft outcomes. The model integrated diverse clinical and laboratory variables from the first transplant year and identified key risk factors for premature graft failure. In Gwinner et al.'s analysis[3], the model demonstrated robust predictive

performance, with a 71% death-censored graft survival at 15 years in the training cohort, significant concordance indices in both training (0.79) and validation (0.74) cohorts, and strong calibration. The findings underscored the model's ability to stratify patients into different risk levels for targeted monitoring and intervention strategies, emphasizing the prognostic importance of eGFR at one year, patient demographics, types of rejections, infections, and glomerulonephritis.

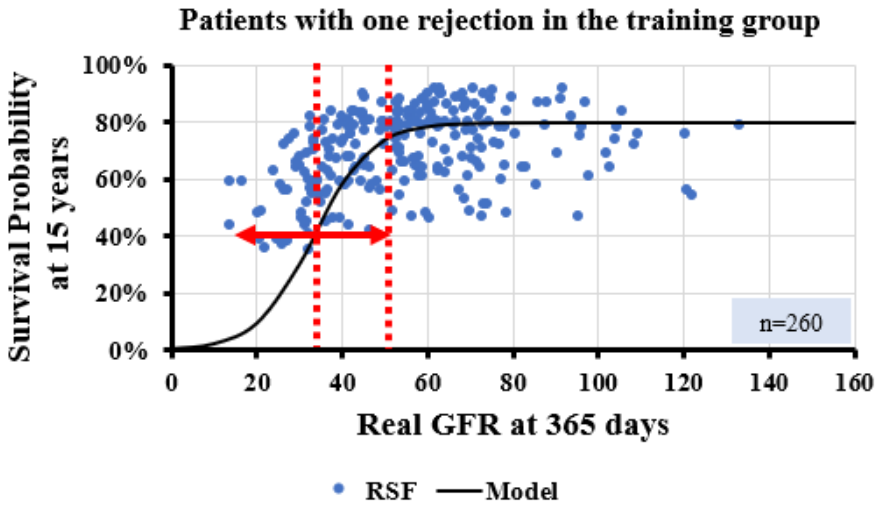
Therefore, the following data illustrates graft survival after 15 years across all patients (Supplementary Figure 5), with our analysis further detailing survival outcomes in patients categorized by rejection history within the first year (no rejection, one rejection, and multiple rejection in Supplementary Figures 6, 7, and 8, respectively) .



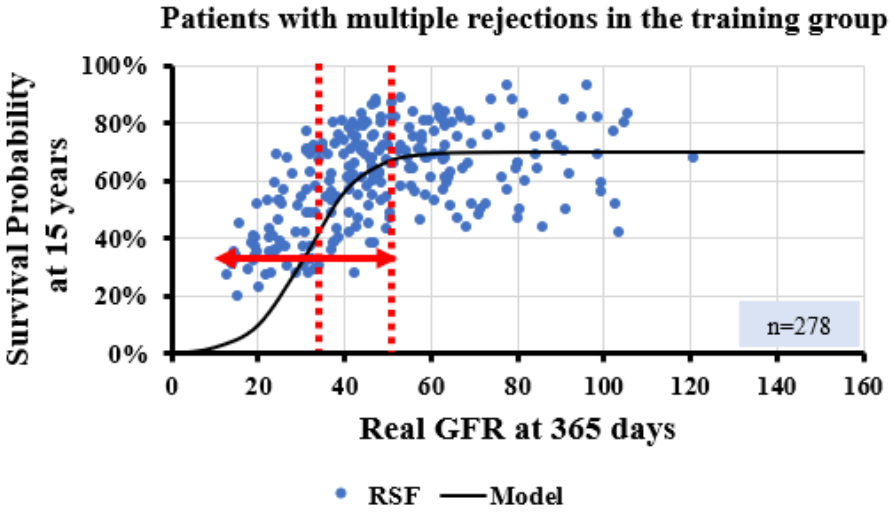
Supplementary Figure 5: Survival model fit for all training-cohort patients. Random survival forest (RSF)–based graft survival probability as a function of one-year GFR for all patients in the training cohort. The sigmoidal fit illustrates a gradual decline in long-term graft survival with decreasing GFR, consistent with the existence of a critical GFR range for failure risk.



Supplementary Figure 6: Survival model fit in non-rejection patients. RSF-based graft survival probability versus one-year GFR for patients without rejection. The fitted sigmoid curve is less noisy in this subgroup and supports a relatively well-defined GFR range within which graft failure probability increases steeply, forming the basis for the critical threshold interval used in the main analysis.



Supplementary Figure 7: Survival model fit in patients with one rejection. RSF-based graft survival probability as a function of one-year GFR in patients with a single rejection episode. The sigmoidal dependence is maintained, but the curve is noisier, reflecting added variability from rejection-related injury while still indicating a transition range in GFR associated with increased failure risk.



Supplementary Figure 8: Survival model fit in patients with multiple rejections. RSF-based graft survival probability versus one-year GFR for patients with multiple rejection episodes. Although the fit is noisy, the overall sigmoidal pattern persists. Across Figures S5–S8, a practical GFR failure-threshold range of 30–50 mL/min/1.73 m² is identified and later used in the uncertainty analysis of the dynamic model.

Based on real data spanning 15 years of graft performance and annual GFR values, sigmoidal behavior is observed. However, the fitting is rather noisy, making it difficult to pinpoint an exact inflection point (GFR threshold). In light of this uncertainty, a range of critical GFR thresholds (30–50 mL/min/1.73 m²) was selected for use in the uncertainty analysis later in the study. This approach accounts for variability and provides a more robust framework for predicting graft performance.

3. Parameters of the model.

The duffing oscillator model is given by:

$$\frac{d^2GFR}{dt^2} + \lambda\frac{dGFR}{dt} + f(GFR) = 0 \quad \text{(S. 22)}$$

Where:

$$f(GFR) = aGFR^3 + bGFT^2 + \gamma GFR + \delta \quad \text{(S. 23)}$$

Which derives:

$$\frac{dGFR}{dt} = u \quad \text{(S. 24)}$$

$$\frac{du}{dt} = -\lambda u - f(GFR) \quad \text{(S. 25)}$$

If c=*GFR* at 365 d, and θ is the threshold, then at steady-state, the model should satisfy equations S. 26, S. 27, and S. 28.

$$f(0) = 0 \;\; => \;\; \delta = 0 \quad \text{(S. 26)}$$

$$\begin{gathered} f(c)=0 => ac^3 + bc^2 + \gamma c = 0 \;\Leftrightarrow \\ \Leftrightarrow \; ac^2 + bc + \gamma = 0 \;\Leftrightarrow \\ \Leftrightarrow \; a = -\frac{bc+\gamma}{c^2} \end{gathered} \quad \text{(S. 27)}$$

$$f(\theta) = 0 \;\; \Leftrightarrow \;\; a\theta^2 + b\theta + \gamma = 0 \quad \text{(S. 28)}$$

At GFR at 365 days:

$$\frac{df}{dx} < 0 \;\; \Leftrightarrow \; 3\alpha c^2 + 2bc + \gamma < 0 \quad \text{(S. 29)}$$

At the threshold θ:

$$\frac{df}{dx} > 0 \;\; \Leftrightarrow \; 3\alpha\theta^2 + 2b\theta + \gamma > 0 \quad \text{(S. 30)}$$

Also, the curvature changes at the threshold

$$\begin{gathered} \frac{d^2f}{dx^2} = 0 \;\; \Leftrightarrow \; 6\alpha\theta + 2b = 0 \Leftrightarrow \\ \Leftrightarrow \theta = -\frac{b}{3\alpha} \end{gathered} \quad \text{(S. 31)}$$

From equations S. 28 and S. 31:

$$\frac{ab^2}{9\alpha^2} - \frac{b^2}{3\alpha} + \gamma = 0 \Leftrightarrow$$
$$\Leftrightarrow b^2 - 3b^2 + 9\alpha\gamma = 0 \Leftrightarrow \quad \text{(S. 32)}$$
$$\Leftrightarrow b = \pm 3\sqrt{\frac{\alpha\gamma}{2}}$$

In equations S. 23 and S. 25, γ should be positive as it represents the decay of *GFR* in the patients. Thus, from equation S. 32, α is positive. Based on equations S. 31 and S. 32, $b = -3\sqrt{\frac{\alpha\gamma}{2}}$ because $\theta > 0$.

θ then equals:

$$\theta = -\frac{b}{3\alpha} = -\frac{-3\sqrt{\frac{\alpha\gamma}{2}}}{3a} \Leftrightarrow \quad \text{(S. 33)}$$
$$\Leftrightarrow \theta = \sqrt{\frac{\gamma}{2\alpha}}$$

From equation S. 27, by replacing b, γ as a function of c and a can be approximated by:

$$ac^2 - 3\sqrt{\frac{\alpha\gamma}{2}}c + \gamma = 0 \Leftrightarrow$$
$$\Leftrightarrow ac^2 - \frac{3}{\sqrt{2}}c\sqrt{a}\sqrt{\gamma} + \gamma = 0 \Leftrightarrow$$
$$\Leftrightarrow ac^2 - 2c\sqrt{a}\sqrt{\gamma} + \gamma + 2c\sqrt{a}\sqrt{\gamma} - \frac{3}{\sqrt{2}}c\sqrt{a}\sqrt{\gamma} = 0 \Leftrightarrow \quad \text{(S. 34)}$$
$$\Leftrightarrow ac^2 + \gamma + \left(\frac{2\sqrt{2} - 3}{\sqrt{2}} - 2\right)c\sqrt{a}\sqrt{\gamma} = 0 \Leftrightarrow$$
$$\Leftrightarrow ac^2 + \gamma - \frac{3}{\sqrt{2}}c\sqrt{a}\sqrt{\gamma} = 0$$

The last part is a second order polynomial of γ, which equals:

$$\sqrt{\gamma} = \frac{\frac{3}{\sqrt{2}}c\sqrt{a} \pm \sqrt{\frac{9}{2}c^2a - 4ac^2}}{2} \Leftrightarrow$$

The two solutions of γ are: (S. 35)

$$\gamma = 2c^2a \text{ and } \gamma = \frac{c^2a}{2}$$

From equations S. 27 and S. 31:

$$c^2 - 3\theta c + \frac{\gamma}{\alpha} = 0 \Leftrightarrow$$
$$\Leftrightarrow \theta = \frac{c^2 + \gamma/\alpha}{3c} \quad \text{(S. 36)}$$

Using the different solutions of γ in the equation S. 35 to S. 34, θ can also be approximated from measurable c by:

$$\theta = \frac{c}{2} \; when \; \gamma = \frac{c^2 a}{2} \tag{S. 37}$$

It is noticeable that α and b parameters depend on γ. Based on Supplementary Table 6 if patients with relevant complications which are known to affect the graft (rejection, BK virus nephropathy, glomerulonephritis, severe infections and other severe extra-renal diseases and urinary tract infections) are excluded the average annual loss is 1.58 ml/min 1.73 m^2/yr.

Supplementary Table 6: Average loss per year in patients without rejection, BK virus nephropathy, glomerulonephritis, severe infections and other severe extra-renal diseases and urinary tract infections. Percentiles of annual eGFR change in patients without rejection, BK virus nephropathy, glomerulonephritis, severe infections, extra-renal diseases, or urinary tract infections, estimated by two methods (grounded and Tukey). Median loss is about 1.6 mL/min/1.73 m² per year.

		Percentile						
		5	**10**	**25**	**50**	**75**	**90**	**95**
Grounded method	Average loss per year	-11.2680	-7.4477	-3.7441	-1.5803	0.9553	3.4796	5.2999
Tukey test	Average loss per year			-3.6809	-1.5707	1.0690		

Based on Supplementary Table 7, the whole cohort of patients has an average annual GFR loss of 2.07 ml/min/1.73 m^2/yr).

Supplementary Table 7: Average loss per year in the whole training cohort of patients (Table 1). Distribution of yearly eGFR change for all patients in the training cohort, including those with complications, again reported as percentiles for two estimation methods. The median decline is steeper (≈2.1 mL/min/1.73 m² per year) than in the no rejection group, reflecting the cumulative impact of graft injuries.

		Percentile						
		5	**10**	**25**	**50**	**75**	**90**	**95**
Grounded method	Average loss per year	-17.8305	-10.6935	-4.7898	-2.0732	0.0798	1.9119	3.5672
Tukey test	Average loss per year			-4.7898	-2.0587	.0798		

The model under consideration is characterized by two key scales: T, which represents time measured in days, and G, which corresponds to the GFR units in mL/(min 1.73 m²). These scales serve as the foundation for understanding the system's dynamics. The model involves several parameters, each with specific units. The parameter α has units of $1/(G^2\,T^2)$, while β is measured in $1/(G\,T^2)$. The parameter γ is expressed in $1/T^2$, and λ has units of $1/T$.

To explore the model in dimensionless form, we introduce two dimensionless variables. The dimensionless GFR is denoted as $\widetilde{GFR}$, and the dimensionless time is represented as τ. These are defined by scaling the GFR and time according to the relations:

$$GFR = G\,\widetilde{GFR} \tag{S. 38}$$

$$t = \frac{\tau}{\sqrt{\gamma}} \tag{S. 39}$$

By substituting these dimensionless variables into the original model, we obtain a dimensionless equation that retains the structure of the system but is now independent of specific units. The resulting equation is:

$$\begin{aligned} &\frac{1}{\gamma}\frac{d^2GFR}{dt^2} + \frac{\lambda}{\gamma}\frac{dGFR}{dt} + \frac{a}{\gamma}GFR^3 + \frac{b}{\gamma}GFR^2 + GFR = 0 \;\Leftrightarrow \\ \Leftrightarrow\; &G\frac{d^2\widetilde{GFR}}{d\tau^2} + \frac{\lambda\,G}{\sqrt{\gamma}}\frac{d\widetilde{GFR}}{d\tau} + \frac{a}{\gamma}G^3\,\widetilde{GFR}^3 + \frac{b}{\gamma}G^2\widetilde{GFR}^2 + G\,\widetilde{GFR} = 0 \end{aligned} \tag{S. 40}$$

Next, we introduce the substitution:

$$G = \sqrt{\frac{\gamma}{\alpha}} \tag{S. 41}$$

to fully nondimensionalize the model. This leads to the dimensionless form of the equation:

$$\frac{d^2\widetilde{GFR}}{d\tau^2} + \tilde{\lambda}\frac{d\widetilde{GFR}}{d\tau} + \widetilde{GFR}^3 + \tilde{B}\,\widetilde{GFR}^2 + \widetilde{GFR} = 0 \tag{S. 42}$$

In this equation, the dimensionless parameters:

$$\tilde{\lambda} = \frac{\lambda}{\sqrt{\gamma}} \tag{S.43}$$

$$\tilde{B} = \frac{\beta}{\gamma}G = \frac{\beta}{\gamma}\sqrt{\frac{\gamma}{\alpha}} = \frac{\beta}{\sqrt{\gamma a}} \tag{S.44}$$

are derived from the original model parameters.

Since GFR typically decreases over time, as evidenced by Supplementary Table 7, an average value for γ can be estimated based on empirical data. By using the time scale expression of Equation S.18 and considering the median GFR decline rate, we calculate:

$$\sqrt{\gamma} = \frac{2.07}{365} \;\Leftrightarrow\; \gamma = 0.000032\ d^{-2} \tag{S.45}$$

This value provides an important reference for analyzing the temporal dynamics of GFR within the dimensionless model.

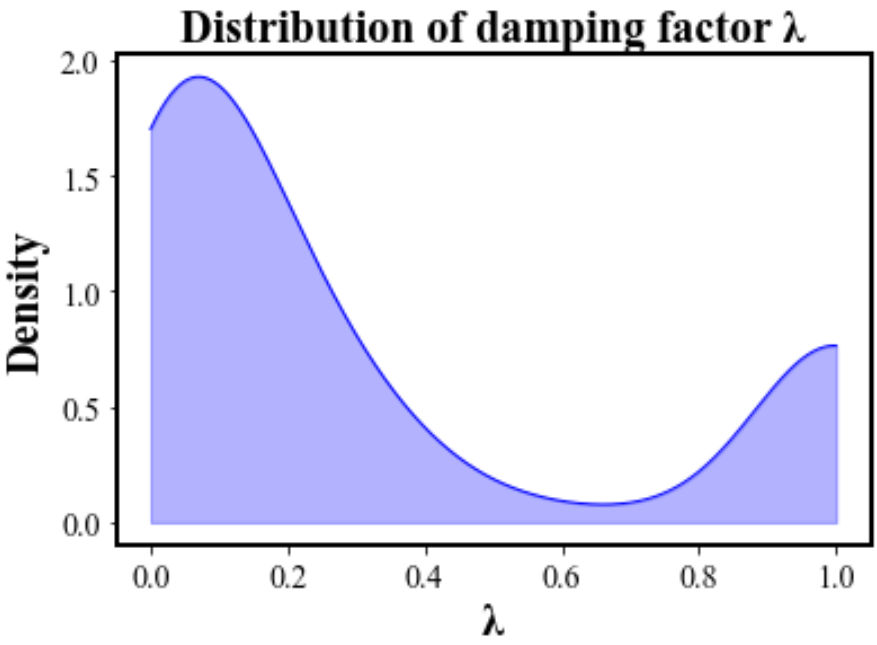
Equation S. 37, which defines the threshold, is crucial for the uncertainty analysis in our study. According to Figures S. 5–S. 8, which depict the graft survival probability in relation to the annual GFR, the inflection point (50%) of the sigmoidal curve varies by ±20 ml/min/1.73 m^2 due to patients' heterogeneity. Therefore, based on equation S. 16, parameters α, and b should be estimated for GFR_{365} ± 40 ml/min/1.73 m^2, with a minimum GFR_{365} value of 5 ml/min/1.73 m^2 when GFR_{365} is less than 50 ml/min/1.73 m^2 to avoid negative GFR_{365} values.

4. Distribution of λ.

Before proceeding with the estimation of the distribution of λ, a sensitivity analysis was conducted to evaluate the influence of four model parameters on GFR values. The Random

Sampling–High Dimensional Model Representation (RS-HDMR) method was selected due to its effectiveness in handling nonlinear models. Assuming an annual GFR of 74.45 mL/min/1.73 m^2 and a γ value of 0.000032 d^{-2} for a patient with an initial GFR of 91.23 mL/min/1.73 m^2 and an initial rate of change of -0.33 (mL/min/1.73 m^2)/d, the nominal values for α and b were set to $1.15 \cdot 10^{-8}$ (1.73 $m^2 \cdot$ min)^2 /(d·mL)^2 and -1.29×10^{-6} (1.73 m^2·min)/(d·mL), respectively. In the RS-HDMR analysis, the parameters α, b, and γ were varied by ±50% from their nominal values. To account for a wide range of possible variations in GFR dynamics in the studied case, the damping factor λ was systematically varied between 10^{-4} and 0.2 d^{-1}. The average values and standard deviations of the first- and second-order sensitivity indices over a two-year period were as follows: 0.173 ± 0.072 for α, 0.303 ± 0.119 for b, 0.021 ± 0.007 for γ, and 0.439 ± 0.222 for λ. These results indicate that γ is the least influential parameter and was thus held constant across all patient-specific estimations. Consequently, from predicting an annual GFR values, α and b were estimated using the nonlinear expressions in Equations S.32 and S.37, while λ was estimated using a Particle Swarm Optimization (PSO) approach.

The distribution of λ obtained from the parameter estimation of 354 patients without rejection is depicted in Supplementary Figure 9.



Supplementary Figure 9: Distribution of patient-specific damping parameters (λ). Distribution of the damping factor λ obtained from dynamic model calibration in 354 non-rejection patients. Small λ values correspond to underdamped dynamics with oscillatory GFR courses over time, whereas large λ values capture overdamped dynamics with rapid relaxation to a steady GFR level. This heterogeneity in λ underpins the diversity of observed GFR trajectories and is explicitly incorporated into the Monte Carlo simulations used for individualized risk estimation.

Small values of λ capture GFR courses that oscillate over time (underdamped oscillator)., whereas for high λ the model could fit patients whose GFR course relaxes fast to a certain value (overdamped oscillator).

- **Supplementary Notes**

1. Histological rejection profiles and transplantation types across cohorts

Supplementary Table 8 summarizes the histological assessment of acute rejection in both the training and validation cohorts. It reports the number and proportion of biopsies showing acute T cell-mediated rejection (borderline, IA/B, IIA/B) and acute antibody-mediated rejection (ABMR) within the first post-transplant year, thereby illustrating how the distribution of rejection phenotypes relates to subsequent graft outcomes. Supplementary Tables 9 and 10 describe the distribution of isolated kidney versus combined organ transplantations (including pancreas, liver, heart, lung and multiorgan procedures) in the training and validation datasets, respectively, stratified by rejection status (none, one, multiple episodes). In the training cohort (Supplementary Table 9), combined procedures are relatively rare and similarly distributed across rejection groups, suggesting that multiorgan transplantation is not a major determinant of rejection category. In contrast, in the validation cohort (Supplementary Table10), some differences in the frequency of pancreas and multiorgan transplantations across rejection strata reach statistical significance, indicating that certain transplant constellations may be associated with a higher risk of rejection in this independent dataset.

Supplementary Table 8: Details to the patient groups with one or more rejection within the first-year post-transplantation, according to the histological type and severity. Biopsies were graded according to the Banff classification, denoting acute antibody-mediated rejection (ABMR) and acute T cell-mediated rejection (TCMR) with severity grades from borderline (BL), TCMR to IA/B (tubulointerstitial) and II A/B (vascular) TCMR.

	Training dataset N=3147 biopsies	**Validation dataset N=2495 biopsies**
Acute TCMR, n (%)		
BL-TCMR	466 (14.8)	431 (17.3)
TCMR IA/B	294 (9.3)	101 (4.0)
TCMR IIA/B	114 (3.6)	87 (3.5)
Acute ABMR, n (%)	398 (9.0)	294 (11.7)

Parentheses include the percentages of each category.

Supplementary Table 9: Kidney transplant patients with combined organ transplantation and their distribution across the subgroups with no rejection, one rejection and multiple rejections in the training cohort. One-sided χ^2 test p-values are included. The table shows that combined procedures are relatively infrequent and similarly distributed across the no rejection and rejection groups, suggesting that combined organ transplantation is not a primary driver of the rejection category in this cohort. Shown are numbers and percentages in parentheses.

	No rejection	**One rejection**	**Multiple rejections**	**p-values**
N	354	260	278	
Pancreas transplantation				
Yes	28 (7.9)	19 (7.3)	18 (6.5)	0.79
No	326 (92.1)	241 (92.7)	260 (41)	
Liver transplantation				
Yes	2 (0.6)	2 (0.8)	0 (0)	0.38
No	352 (99.4)	258 (99.2)	278 (100)	
Heart transplantation				
Yes	5 (1.4)	1 (0.4)	1 (0.4)	0.23
No	349 (98.6)	259 (99.6)	277 (99.6)	
Lung transplantation				
Yes	2 (0.6)	2 (0.8)	1 (0.4)	0.81
No	352 (99.4)	258 (99.2)	277 (99.6)	
Multiorgan transplantation				
Yes	36 (10.2)	24 (9.2)	20 (7.2)	0.42
No	318 (89.8)	236 (90.8)	258 (92.8)	

Parentheses include the percentages of each category. No statistically significant one-sided χ^2 test p-value (> 0.05) is found.

Supplementary Table 10: Kidney transplant patients with combined organ transplantation and their distribution across the subgroups with no rejection, one rejection and multiple rejections in the validation cohort. One-sided χ^2 test p-values are included. The table shows that combined procedures are relatively infrequent and similarly distributed across the no rejection and rejection groups, suggesting that combined organ transplantation is not a primary driver of the rejection category in this cohort. Shown are numbers and percentages in parentheses.

	No rejection	**One rejection**	**Multiple rejections**	**p-values**
N	410	256	181	
Pancreas transplantation				
Yes	36 (8.8)	10 (3.9)	3 (1.7)	**0.00**[a]
No	374 (91.2)	246 (96.1)	178 (98.3)	
Liver transplantation				
Yes	2 (0.5)	3 (1.2)	0 (0)	0.27
No	408 (99.5)	253 (98.8)	181 (100)	
Heart transplantation				
Yes	2 (0.5)	0 (0)	0 (0)	0.34
No	408 (99.5)	256 (100)	181 (100)	
Lung transplantation				
Yes	0 (0)	0 (0)	0 (0)	1.00
No	410 (100)	256 (100)	181 (100)	
Multiorgan transplantation				
Yes	40 (9.8)	13 (5.1)	3 (1.7)	**0.00**[a]
No	370 (90.2)	243 (94.9)	178 (98.3)	

Parentheses include the percentages of each category.
[a] Statistically significant one-sided χ^2 test p-value ≤ 0.05 (bold).

- **Supplementary Results**

1. Classification with regular XGBoost, and Random Forest.

In this section, the prediction of high risk patients was treated as a very simple classification problem. High-risk patients were those whose annual GFR value was below a specified threshold, with thresholds of 30, 40, and 50 mL/min/1.73 m² obtained from Section 2 of SI.

$$failure = \begin{cases} 1, \ when\ measured\ GFR_{365\,d} \leq Clin_{thr} \\ 0, \ when\ measured\ GFR_{365\,d} > Clin_{thr} \end{cases} \quad \text{(S.46)}$$

For classification XGBoost and Random Forest were performed in the patient with one, with multiple and without rejection, and the NearMiss algorithm was used due to imbalanced data. The hyperparameter optimization for the XGBoost model was conducted using the following parameter grid: the number of estimators was set to [100, 200, 300], the maximum tree depth ranged from [3, 4, 5], the learning rate was varied across [0.01, 0.1, 0.2], the subsample values were chosen from [0.8, 0.9, 1.0], and the column sampling by tree values were selected from [0.8, 0.9, 1.0]. For the Random Forest model, the hyperparameter optimization utilized a parameter grid consisting of: the number of estimators set to [100, 200, 300], maximum depth options of [3, 4, 5], the minimum number of samples required to split a node ranging from [2, 5, 10], the minimum number of samples per leaf set to [1, 2, 4], and both bootstrap settings [True, False] were evaluated.

The results are shown in Supplementary Tables 11-13.

Supplementary Table 11: XGBoost, and Random Forest accuracies for different critical thresholds ($Crit_{thr}$) in non-rejection patients. The F1 score is in parentheses. AUC and F1 scores for XGBoost and Random Forest classifiers predicting high-risk patients (annual eGFR below 30, 40, or 50 mL/min/1.73 m²) among non-rejection patients. Although both methods achieve reasonable discrimination, their AUC and F1 values remain below those of the dynamic-model–based pipeline reported in the main text, especially at higher thresholds, underscoring the benefit of mechanistic modeling plus uncertainty analysis.

Classifier	**AUC with F1 score for $Crit_{thr}$ equal to 30 mL/min/1.73m²**	**AUC with F1 score for $Crit_{thr}$ equal to 40 mL/min/1.73m²**	**AUC with F1 score for $Crit_{thr}$ equal to 50 mL/min/1.73m²**
XGBoost	0.65 (0.59)	0.73 (0.75)	0.76 (0.76)
Random Forest	0.71 (0.66)	0.74 (0.73)	0.79 (0.79)

Parentheses include the F1 score.

Supplementary Table 12: XGBoost, and Random Forest accuracies for different critical thresholds ($Crit_{thr}$) in one rejection patients. Classification performance of XGBoost and Random Forest in patients with one rejection episode across three candidate eGFR thresholds. AUC and F1 scores vary by threshold and method, reaching up to ~0.88–0.89, but still do not consistently surpass the individualized-threshold approach.

Classifier	**AUC with F1 score for $Crit_{thr}$ equal to 30 mL/min/1.73m²**	**AUC with F1 score for $Crit_{thr}$ equal to 40 mL/min/1.73m²**	**AUC with F1 score for $Crit_{thr}$ equal to 50 mL/min/1.73m²**
XGBoost	0.77 (0.76)	0.74 (0.70)	0.89 (0.88)
Random Forest	0.88 (0.86)	0.68 (0.66)	0.87 (0.87)

Parentheses include the F1 score.

Supplementary Table 13: XGBoost, and Random Forest accuracies for different critical thresholds (Critthr) in multiple rejection patients. AUC and F1 scores for XGBoost and Random Forest classifiers in the multiple-rejection subgroup. Overall performance is slightly lower and more variable than in non- or single-rejection groups, reflecting greater heterogeneity.

Classifier	AUC with F1 score for $Crit_{thr}$ equal to 30 mL/min/1.73m^2	AUC with F1 score for $Crit_{thr}$ equal to 40 mL/min/1.73m^2	AUC with F1 score for $Crit_{thr}$ equal to 50 mL/min/1.73m^2
XGBoost	0.77 (0.74)	0.74 (0.72)	0.74 (0.75)
Random Forest	0.77 (0.74)	0.68 (0.69)	0.71 (0.73)

Parentheses include the F1 score.

Based on the results on Supplementary Tables 11-13, regardless of considering the issue of data imbalance, on average, both the ROC AUC and F1 scores were lower than those achieved by our procedure.

Also, we did a sensitivity analysis to evaluate the model's robustness in the training and validation cohorts by stratifying by patient sex, donor sex, and patient age (<50 and ≥50 years). The model was tested for performance after 1 to 2 years of transplant for each stratum in both the training (Supplementary Table 14) and validation (Supplementary Table 15) sets with regards to the values for both AUC and F1 scores. Across subgroup comparisons and both timeframes, AUC and F1 scores remained at a high level and had minimal variability across the different demographic characteristics, thereby confirming that demographics have little impact on model performance. This level of consistency indicates that the findings can be generalised across the range of patient and donor characteristics and also across the short- and mid-term follow-up timeframes.

Supplementary Table 14: Sensitivity analysis of predictive performance in the training cohort. AUC and F1 for predicting graft failure following one (1) year and two (2) years in a training cohort (N_1) of patients, grouped by Patient Sex, Donor Sex and Patient Age. While there are differences across subgroups, there are substantially increased AUC (≈0.79 – 0.88) and good F1 scores for the discriminatory abilities of this model across all of the demographic categories. Therefore, the individualized threshold method is robust and generalizable in the training population.

Factor	Categories	Time (day)	AUC	F1 score
Patient sex	Male	365	0.88±0.02	0.84±0.04
	Female	365	0.85±0.03	0.73±0.06
Donor sex	Male	365	0.87±0.02	0.82±0.04
	Female	365	0.87±0.03	0.78±0.05
Patient age	< 50 years	365	0.85±0.03	0.84±0.05
	≥ 50 years	365	0.87±0.01	0.76±0.05
Patient sex	Male	730	0.84±0.02	0.82±0.04
	Female	730	0.80±0.02	0.72±0.05
Donor sex	Male	730	0.83±0.02	0.81±0.04
	Female	730	0.81±0.02	0.76±0.05
Patient age	< 50 years	730	0.79±0.03	0.82±0.05
	≥ 50 years	730	0.83±0.02	0.74±0.04

AUC and F1 score values are provided as mean with standard deviation in (±).

Supplementary Table 15: Sensitivity analysis of predictive performance in the validation cohort. AUC and F1 scores for graft failure prediction at 365 and 730 days in the validation cohort (N_2), stratified by patient sex, donor sex, and age. Performance remains consistently high across all subgroups, with AUCs around 0.84–0.92 and strong F1 scores, confirming that the model generalizes well to a temporally distinct cohort. This analysis supports the applicability of the personalized eGFR threshold approach across different patient demographics.

Factor	Categories	Time (day)	AUC	F1 score
Patient sex	Male	365	0.90±0.04	0.86±0.03
	Female	365	0.92±0.06	0.80±0.04
Donor sex	Male	365	0.91±0.04	0.86±0.03
	Female	365	0.90±0.06	0.82±0.03
Patient age	< 50 years	365	0.90±0.05	0.88±0.03
	≥ 50 years	365	0.90±0.05	0.80±0.04
Patient sex	Male	730	0.86±0.04	0.85±0.03
	Female	730	0.89±0.03	0.79±0.04
Donor sex	Male	730	0.87±0.04	0.85±0.03
	Female	730	0.87±0.03	0.81±0.03
Patient age	< 50 years	730	0.84±0.01	0.86±0.03
	≥ 50 years	730	0.89±0.05	0.79±0.04

AUC and F1 score values are provided as mean with standard deviation in (±).

Supplementary References

1. Lin, Y.-C. *et al.* The impact of baseline glomerular filtration rate on subsequent changes of glomerular filtration rate in patients with chronic kidney disease. *Sci Rep* **11**, 7894 (2021).

2. Fahrmeir, L., Kneib, T., Lang, S. & Marx, B. D. *Regression*. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2021). doi:10.1007/978-3-662-63882-8.

3. Gwinner, W. *et al.* Modeling long-term kidney graft survival with first transplant-year parameters. *Under review*.

4. Scheffner, I., Gietzelt, M., Abeling, T., Marschollek, M. & Gwinner, W. Patient Survival after Kidney Transplantation: Important Role of Graft-sustaining Factors as Determined by Predictive Modeling Using Random Survival Forest Analysis. *Transplantation* 1095–1107 (2020) doi:10.1097/TP.0000000000002922.