Supplementary material

1. SUPPLEMENTARY METHODS

1.1. Data collection procedures

1.2. Independent validation cohorts

1.3. Prognostic parameters prospectively collected and assessed in the derivation cohort

1.4. Therapeutic protocol for interventions

1.5. Statistical analysis interpretation

1.6. Construction of the integrative score derived from the final multivariable Cox Model

1.7. TRIPOD Check list: Prediction Model Development and Validation

2. SUPPLEMENTARY TABLES

Supplementary Table A: Details of the Clinical trials depicting the population characteristics, clinical scenarios and interventions

Supplementary Table B: General transplant procedures and policies and allocation systems in the participating centres

Supplementary Table C: Baseline characteristics of the European validation centres

Supplementary Table D: Baseline characteristics of the North-American validation centres

Supplementary Table E: Independent determinants of kidney allograft loss in the derivation cohort stratified by centre: multivariable analysis

Supplementary Table F: Patients characteristics according to the therapeutic intervention.

Supplementary Table G: iBox risk score comparison of previously published risk scores

Supplementary Table H: Independent determinants of kidney allograft loss in the derivation cohort using histological diagnoses: multivariable analysis

3. SUPPLEMENTARY FIGURES

Supplementary Figure A: Density of risk evaluation time points after transplantation

Supplementary Figure B: iBox practical application for clinicians: Ready-to-use interface for clinicians.

Supplementary Figure C: Effect of treatment intervention on iBox risk score.

1. SUPPLEMENTARY METHODS

1.1 Data collection procedures

All data from Paris-Necker, Paris-Saint Louis, Foch and Toulouse hospitals were extracted from the prospective Paris Transplant Group Cohort data cohort. The database networks have been approved by the National French Commission for bioinformatics data and patient liberty and codes were used to ensure strict donor and recipient anonymity and blind access. Informed consent was obtained from the participants at the time of transplantation. The data are computerised in real time and at the time of transplantation, at the time of post-transplant allograft biopsies and at each transplant anniversary and are submitted for an annual audit.

1.2 Independent validation cohorts

In the European validation cohort, the French data from the Lyon, and Nantes Hospitals for donors and recipients were extracted from the DIVAT cohort (official website: www.divat.fr) and from the French national agency database CRISTAL (official website: https://www.sipg.sante.fr/portail/). The Belgian data and data from the North-American validation cohort were collected as part of routine clinical practice and entered in centres' databases in compliance with local and national regulatory requirements. They were sent anonymised to the Paris Transplant Group.

1.3 Prognostic parameters prospectively collected and assessed in the derivation cohort

Baseline recipient's characteristics:	
1.	Recipient's age
2.	Recipient's gender
3.	Recipient's height
4.	Recipient's weight
5.	Previous transplantation
6.	Delay between dialysis and transplantation
7.	Cause of end stage renal disease
8.	ABO blood group
9.	HLA genotype
10.	CMV serology
11.	HCV serology
12.	HBV serology
13.	HIV serology
Baseline donor's characteristics:	
14.	Donor's age
15.	Donor's gender
16.	Donor's height
17.	Donor's weight
18.	Type of donor: deceased vs living
19.	Cause of donor's death
20.	Double transplantation
21.	History of hypertension
22.	History of diabetes
23.	ECD status
24.	Serum creatinine
25.	ABO blood group
26.	HLA genotype
27.	CMV serology
28.	HCV serology
29.	HBV serology

30.	HIV serology
Immunological characteristics at the time of transplantation:	
31.	HLA mismatches A
32.	HLA mismatches B
33.	HLA mismatches Cw
34.	HLA mismatches DQ
35.	HLA mismatches DR
36.	HLA mismatches DP
37.	Anti-HLA DSA at the time of transplantation
38.	MFI of the anti-HLA DSA at the time of transplantation
39.	cPRA
Transplant characteristics:	
40.	Cold ischemia time
41.	Delayed graft function
42.	Induction treatment with anti-thymocyte globulin
43.	Induction treatment with basiliximab
44.	Steroid dose
Immunological data at the time of risk assessment (Luminex SA assessment A, B, C, DP, DQ, DR)	
45.	Anti-HLA DSA
46.	MFI of immunodominant anti-HLA DSA
Histological data according to the Banff classification:	
47.	g Banff score
48.	ptc Banff score
49.	t Banff score
50.	i Banff score
51.	cg Banff score
52.	v Banff score
53.	mm Banff score
54.	ci Banff score
55.	ct Banff score
56.	IFTA Banff score
57.	cv Banff score
58.	ah Banff score
59.	C4d ptc deposition
60.	Recurrence of ESRD
61.	Polyomavirus-associated nephropathy
62.	ABMR status
63.	TCMR status
64.	Borderline category
Follow-up variables:	
65.	Episodes of pyelonephritis
66.	Immunosuppression treatment
67.	Type of treatment: calcineurin inhibitors, mycophenolate mofetil, mTOR inhibitors or belatacept
68.	CNI blood through level at M12 and every year
69.	Steroid dose at M12 and every year
70.	Rejection therapy (e.g., steroid, plasma exchange, intravenous immunoglobulin)
71.	CMV prophylaxis
72.	BK viral load at M12 and every year

73.	CMV viral load at M12 and every year
74.	Allograft function at M12 and every year
75.	Proteinuria at M12 and every year
76.	Patient date and cause of allograft loss
77.	Patient date and cause of death

Detection and Characterisation of Donor-specific Anti-HLA Antibodies
All patients were tested for the presence of circulating anti-HLA donor-specific antibodies (DSAs) at the time of patient risk evaluation. The presence of circulating DSAs against HLA-A, HLA-B, HLA-Cw, HLA-DR, HLA-DQ and HLA-DP was retrospectively determined using single-antigen flow bead assays (One Lambda, Inc., Canoga Park, CA, USA) on a Luminex platform. Beads with a normalised mean fluorescence intensity (MFI), a measure of donor-specific antibody strength, of greater than 500 units were judged as positive as previously described. HLA typing of the transplant recipients and donors was performed using an Innolipa HLA Typing Kit (Innogenetics, Ghent, Belgium). In the validation cohorts, HLA genotyping and HLA antibody profiling were performed according to local centre practice.

Kidney Allograft Phenotypes at time of risk assessment
In the derivation cohort, allograft biopsies were scored and graded from 0 to 3 according to the updated Banff criteria for allograft pathology for the following histological factors: glomerular inflammation (glomerulitis), tubular inflammation (tubulitis), interstitial inflammation, endarteritis, peritubular capillary inflammation (capillaritis), transplant glomerulopathy, interstitial fibrosis, tubular atrophy, arteriolar hyalinosis and arteriosclerosis. Additional diagnoses provided by the biopsy (e.g., the diagnoses of primary disease recurrence, BK virus nephropathy) were recorded. The biopsy sections (4 µm) were stained with periodic acid-Schiff, Masson's trichrome, and hematoxylin and eosin. C4d staining was performed via immunohistochemical analysis on paraffin sections using polyclonal human anti-C4d antibodies. Also, in the validation cohorts, the Banff criteria for the individual histological lesions were assessed in each biopsy included in the study.

1.4 Therapeutic protocol for interventions

We identified patients in the derivation cohort with therapeutic interventions: antibody-mediated rejection biopsy-proven according to the Banff Classification 2015 (n=425), T-cell mediated rejection biopsy-proven according to the Banff Classification 2015 (n=305) and calcineurin inhibitor weaning for calcineurin inhibitor toxicity with belatacept (n=114). iBox evaluations were performed at the time of treatment and after treatment.
All patients with antibody-mediated rejection received standard of care treatment including antibody targeting therapies consisting of 4 courses of high-dose intravenous immune globulin (2 g/kg of body weight over 96 hours), plasma exchange (5 rounds), and anti-CD-20 rituximab (Mabthera®, Roche, Meylan, France, 375 mg per square meter of body-surface area). Patients with a diagnosis of T cell-mediated rejection received 3 methylprednisolone pulses (500 mg/day) given intravenously together with oral steroid tapering (from 1 mg/kg BW to 10 mg per day over a month period). Last group consisted in adult recipients of a renal allograft from a living or deceased donor receiving calcineurin inhibitor-based maintenance immunosuppression at a stable dose during the month immediately before treatment and stable doses of background immunosuppression (mycophenolate mofetil, mycophenolic acid, sirolimus, or azathioprine) + corticosteroids, that were diagnosed with calcineurin inhibitor toxicity based on allograft biopsy assessment and converted to Belatacept (CTLA4-Ig). Belatacept 5 mg/kg was given by intravenous infusion on days 1, 15, 29, 43, and 57, and then every 28 days thereafter. Calcineurin inhibitor dose was tapered as follows: 100% on day 1, to 40 to 60% on day 15, 20 to 30% on day 23, and none on day 29 and beyond.

1.5 Statistical analysis interpretation

1.5.1 Continuous variables

When used as continuous variables in the Cox model, a potential non-linear relationship between predictors and allograft loss was first investigated using restricted cubic splines modelling. Secondly, a fractional polynomial method was applied to determine the best transformation for continuous variables. For donor age, recipient age, eGFR and HLA mismatches, a linear relationship with outcome was found to be a good approximation. A logarithmic transformation was necessary for proteinuria and time post-transplant.

1.5.2 Discrimination

The aim of discrimination is to distinguish between patients who experience an event and those who do not. The C-index estimates the proportion of all pairwise patient combinations from the sample data whose survival time can be ordered such that the patient with the highest predicted survival is the one who actually survived longer (discrimination). The C-index ($0 \leq C \leq 1$) is the probability of concordance between predicted and observed survival, with C-index=0.5 for random predictions and C-index=1 for a perfectly discriminating model.

1.5.3 Calibration

Calibration refers to the ability to provide unbiased survival predictions in groups of similar patients. It estimates how close the score-estimated risk is to the observed risk. A prediction model is considered "well calibrated" if the difference between predictions and observations in all groups of similar patients is close to 0 (perfect calibration). Any large deviation ($p<0.1$) indicates a lack of calibration.

1.5.4 Bootstrapping

Bootstrapping is the preferred simulation technique that was first described by Bradley Efron. The original dataset is a random sample of patients being representative of a general population. Bootstrapping means generating a large number of datasets, each of which with the same sample size as the original one, by resampling with replacement (i.e., a previously selected patient may be selected again).

1.5.5 Internal validation

Internal validation is useful to obtain an honest estimate of the model performance for patients who are similar to those in the development sample and to indicate an upper limit to the expected performance in other settings. The bootstrap approach is the preferred technique to assess internal validity. The internal validity of the final model was confirmed using a bootstrap procedure, which involved generating 1,000 datasets derived from resampling the original dataset and permitted the calculation of optimism-corrected performance estimates.

1.5.6 External validation

External validation may show different results from internal validation since many aspects may be different between settings, including selection of patients, definitions of variables, and diagnostic or therapeutic procedures. The strength of the evidence for the score validity is usually considered greater with a fully external validation (e.g., other investigators and centres).

1.6 Construction of the integrative score derived from the final multivariable Cox model

The construction of the score was performed with the sum of the beta derived from the final cox model for each patient. The survival probabilities, ranging from 0 to 100%, were performed using the baseline survival at 7 years and the sum of the beta.

1.7 TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic			Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	4
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5-6
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	6
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	7
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	7-8
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7-8
	5b	D;V	Describe eligibility criteria for participants.	7
	5c	D;V	Give details of treatments received, if relevant.	n/a
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	9
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	n/a
Predictors	7a	D;V	Clearly define all predictors used in developing the multivariable prediction model, including how and when they were measured.	8-9; Appendix material methods
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	n/a
Sample size	8	D;V	Explain how the study size was arrived at.	**n/a**
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	9
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	10
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	9-11
	10c	V	For validation, describe how the predictions were calculated.	10
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	10
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	**n/a**
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	11
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	7-8; Appendix table 2
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	12
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	12; Table 1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	Table 1
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	Table 2
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	Table 2
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	13; appendix material – methods p. 5
	15b	D	Explain how to use the prediction model.	13
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	13-14
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	n/a
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	19-20
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	17-18
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	17-20
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	19-20
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Supplementary appendix

Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	21

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

2 SUPPLEMENTARY TABLES

Supplementary Table A: Details of the Clinical trials depicting the population characteristics, clinical scenarios and interventions

STUDY	Trial #ID	Design	Clinical scenario	Target population	(n)	Time post-transplant of iBox risk score evaluation	Follow-up time post-transplant	iBox risk score C-Stat
CERTITEM*	**NCT 01079143**	Prospective, Randomised, open-label, multicentre trial	ISD minimisation	Recipients of renal transplants from a living or deceased donor	194	Median: 0.94 years IQR (0.92-0.98)	Median: 6.62 years IQR (2.82-7.34)	0.88
RITUX ERAH†	**Eudra CT 2007-003213-13**	Prospective, Randomised, multicentre, double-blind, placebo-controlled trial	Treatment of ABMR (preexisting DSA)	Recipients of renal transplants from a living or deceased donor with diagnosis of acute ABMR.	38	Median: 0.74 years IQR (0.53-1.10)	Median: 6.63 years IQR (4.03-7.69)	0.77
BORTEJECT‡	**NCT 01873157**	Prospective, Randomised, placebo-controlled, double-blind, single-centre trial	Treatment of ABMR (*de novo* DSA)	Recipients of renal transplants from a living or deceased donor with post-transplant *de novo* DSA detection	44	Median: 6.61 years IQR (4.04-15.41)	Median: 7.75 years IQR (5.32-16.41)	0.94

*: Rostaing, L., et al. "Fibrosis progression according to epithelial-mesenchymal transition profile: a randomised trial of everolimus versus CsA." American Journal of Transplantation 15.5 (2015): 1303-1312; †: Sautenet, B., et al. "One-year results of the effects of rituximab on acute antibody-mediated rejection in renal transplantation: RITUX ERAH, a multicentre double-blind randomised placebo-controlled trial." Transplantation 100.2 (2016): 391-399; ‡: Eskandary, Farsad, et al. "A Randomised Trial of Bortezomib in Late Antibody-Mediated Kidney Transplant Rejection." Journal of the American Society of Nephrology (2017): ASN-2017070818.

Supplementary Table B: General transplant procedures and policies and allocation systems in the participating centres

Transplant Referral Centres	Allocation system	Deceased / living donor rate	Expanded criteria donor rate	Dual kidney transplantation program	Paired donor exchange national program	ABO incompatible program	HLA incompatible program	Standard induction therapy Protocols ATG: Anti-thymocyte Globulin IL2R: interleukin 2 receptor
Paris Transplant Group Saint Louis, Necker, and Foch Hospitals, France	ABM: *Agence Française Biomédecine*[*]	84% / 16%	42%	**YES**	NO	**YES**	**YES**	**Induction rate 100%** (ATG or anti-IL2R)
Toulouse Hospital, France	ABM: *Agence Française Biomédecine*[*]	88% / 12%	41%	**NO**	**NO**	**YES**	**YES**	**Induction rate 85%** (ATG or anti-IL2R)
Nantes Hospital, France	ABM: *Agence Française Biomédecine*[*]	90% / 10%	50%	**NO**	**NO**	**NO**	**NO**	**Induction rate 80%** (ATG or anti-IL2R)
Lyon Hospital, France	ABM: *Agence Française Biomédecine*[*]	93% / 7%	24%	**YES**	**NO**	**YES**	**NO**	**Induction rate 100%** (ATG or anti-IL2R)
Leuven Hospital, Belgium	EuroTransplant: EU allocation system[†]	94% / 6%	30%	**NO**	**NO**	**YES**	**NO**	**Induction rate 40%** (anti-IL2R)
Johns Hopkins Medical Institute, Baltimore, USA	UNOS United Nations for Organ Sharing[‡]	49% / 51%	13%	**NO**	**YES**	**YES**	**YES**	**Induction rate 100%** (ATG or anti-IL2R)
Virginia, USA	UNOS United Nations for Organ Sharing[‡]	27% / 73%	10%	**NO**	**YES**	**YES**	**NO**	**Induction rate 100%** (ATG or anti-IL2R)
Mayo Clinic, Rochester, USA	UNOS United Nations for Organ Sharing[‡]	22% / 78%	4%	**NO**	**YES**	**YES**	**YES**	**Induction rate 100%** (ATG or anti-IL2R)

[*] http://sipg.sante.fr/portail/,
[†] http://www.eurotransplant.org/,
[‡] http://www.unos.org/

Supplementary Table C: Baseline characteristics of the European validation centres

	n	Nantes (France) (n=632)	n	Lyon (France) (n=608)	n	Leuven (Belgium) (n=889)
Recipient characteristics						
Age (years), mean (SD)	632	50.38 (13.57)	608	46.63 (13.28)	889	53.42 (13.30)
Gender male, No. (%)	632	404 (63.92)	608	386 (63.49)	889	543 (61.08)
ESRD causes	632		608		889	
Glomerulonephritis, No. (%)		179 (28.32)		151 (24.84)		254 (28.57)
Diabetes, No. (%)		55 (8.70)		188 (30.92)		73 (8.21)
Vascular, No. (%)		53 (8.39)		49 (8.06)		37 (4.16)
Other, No. (%)		345 (54.59)		220 (36.18)		525 (59.06)
Donor characteristics						
Age (years), mean (SD)	632	53.07 (14.99)	603	44.08 (16.55)	887	47.63 (14.89)
Male gender, No. (%)	631	354 (56.10)	605	395 (65.29)	888	476 (53.60)
Hypertension, No. (%)	620	185 (29.84)	607	101 (16.64)	649	164 (25.27)
Diabetes mellitus, No. (%)	481	36 (7.48)	343	11 (3.21)	889	0
Creatinine > 1.5 mg/dL, No. (%)	631	80(12.68)	605	95 (15.70)	700	18 (2.57)
Donor type						
Deceased donor, No. (%)	632	576 (91.14)	608	564 (92.76)	889	834 (93.81)
Death from cerebrovascular disease, No. (%)	576	323 (56.08)	564	257 (45.57)	834	413 (49.52)
Expanded criteria donor, No. (%)	574	248 (43.21)	608	142 (23.36)	828	238 (28.74)
Transplant baseline characteristics						
Prior kidney transplant, No. (%)	632	101 (15.98)	608	94 (15.46)	889	127 (14.29)
Cold ischemia time (hours), mean (SD)	632	18.75 (9.39)	599	13.68 (5.85)	862	14.37 (5.44)
Delayed graft function[*]**,** No. (%)	630	213 (33.81)	608	102 (16.78)	889	161 (18.11)
HLA-A/B/DR mismatch, mean (SD), number	632	3.28 (1.36)	608	3.58 (1.35)	843	2.75 (1.34)

Abbreviations: ESRD: end-stage renal disease; HLA: human leucocyte antigen.
[*] Delayed graft function was defined as the use of dialysis in the first postoperative week

Supplementary Table D: Baseline characteristics of the North-American validation centres

	Johns Hopkins (USA) (n=580)		Mayo Clinic (USA) (n=556)		Virginia (USA) (n=292)	
	n		n		n	
Recipient characteristics						
Age (years), mean (SD)	580	51.01 (14.70)	556	52.19 (13.74)	284	45.74 (12.88)
Gender male, No. (%)	580	321 (55.34)	556	340 (61.15)	292	169 (57.88)
ESRD causes	580		556		292	
Glomerulonephritis, No. (%)		147 (25.34)		162 (29.14)		56 (19.18)
Diabetes, No. (%)		116 (20.00)		106 (19.06)		49 (16.78)
Vascular, No. (%)		97 (16.72)		63 (11.33)		89 (30.48)
Other, No. (%)		220 (37.93)		225 (40.47)		98 (33.56)
Donor characteristics						
Age (years), mean (SD)	580	40.11 (14.78)	556	43.29 (13.00)	284	38.39 (17.13)
Male gender, No. (%)	580	279 (48.10)	556	258 (46.40)	284	157 (55.28)
Hypertension, No. (%)	578	73 (12.63)	429	50 (11.66)	280	66 (23.57)
Diabetes mellitus, No. (%)	577	30 (5.20)	419	3 (0.7)	280	14 (5.00)
Creatinine > 1.5 mg/dL, No. (%)	281	79 (28.11)	510	148 (29.02)	284	57 (20.07)
Donor type						
Deceased donor, No. (%)	580	283 (48.79)	556	123 (22.12)	292	214 (73.29)
Death from cerebrovascular disease, No. (%)	283	88 (31.10)	123	36 (29.27)	212	70 (33.02)
Expanded criteria donor, No. (%)	580	38 (6.55)	556	5 (0.90)	289	29 (10.03)
Transplant baseline characteristics						
Prior kidney transplant, No. (%)	580	99 (17.07)	544	78 (14.34)	284	58 (20.42)
Cold ischemia time (hours), mean (SD)	541	10.54 (13.35)	397	4.02 (6.97)	274	15.44 (10.70)
Delayed graft function[*]**,** No. (%)	576	35 (6.08)	556	3 (0.54)	292	120 (41.10)
HLA-A/B/DR mismatch, mean (SD), number	579	3.64 (1.73)	556	3.18 (1.86)	292	4.03 (1.61)

Abbreviations: ESRD: end-stage renal disease; HLA: human leucocyte antigen.
[*] Delayed graft function was defined as the use of dialysis in the first postoperative week

Supplementary Table E: Independent determinants of kidney allograft loss in the derivation cohort stratified by centre: multivariable analysis

		Number of patients	Number of events	HR	95% CI	p
Time from transplant to evaluation (year)		3,941	538	1.074	(1.017-1.134)	**0.0108**
eGFR (mL/min/1.73 m^2)		3,941	538	0.955	(0.949-0.961)	**<0.0001**
Proteinuria (log)		3,941	538	1.527	(1.414-1.648)	**<0.0001**
Interstitial fibrosis/	**0/1**	3,074	330	1	-	
Tubular atrophy (IFTA)	**2**	550	115	1.287	(1.029-1.610)	
	3	317	93	1.712	(1.321-2.220)	**0.0002**
Microcirculation	**0-2**	3,568	414	1	-	
Inflammation (g+ptc)	**3-4**	299	90	1.484	(1.142-1.930)	
	5-6	74	34	2.017	(1.358-2.997)	**0.0003**
Interstitial inflammation	**0-2**	3,559	447	1	-	
and tubulitis (i+t)	**≥3**	382	91	1.352	(1.071-1.706)	**0.0111**
Transplant	**0**	3,684	445	1		
Glomerulopathy (cg)	**≥1**	257	93	1.480	(1.140-1.921)	**0.0032**
Anti-HLA donor-specific	**<500**	3,265	387	1	-	
antibody mean	**≥500 – 3,000**	477	80	1.280	(0.986-1.661)	
fluorescence intensity	**≥3,000 – 6,000**	80	23	1.809	(1.167-2.803)	
	≥6,000	119	48	2.228	(1.591-3.120)	**<0.0001**

Supplementary Table F: Patients characteristics according to the therapeutic intervention.

	n	Overall population (n=844)	n	ABMR set (n=425)	n	TCMR set (n=305)	n	Minimization set (n=114)	P-Value
Recipient characteristics									
Age (years), mean (SD)	844	48.58 (14.14)	425	47.97 (14.24)	305	47.20 (13.19)	114	54.60 (14.83)	**<0.001**
Gender male, No. (%)	844	516 (61.14)	425	228 (53.65)	305	216 (70.82)	114	71 (63.16)	**<0.001**
ESRD causes	844		425		305		114		
Glomerulonephritis, No. (%)		240 (28.44)		122 (28.71)		91 (29.84)		27 (23.68)	
Diabetes, No. (%)		82 (9.72)		38 (8.94)		29 (9.51)		15 (13.16)	
Vascular, No. (%)		55 (6.52)		26 (6.12)		23 (7.54)		6 (5.26)	
Other, No. (%)		467 (55.33)		239 (56.24)		162 (53.11)		66 (57.89)	0.659
Donor characteristics									
Age (years), mean (SD)	844	52.17 (16.94)	425	50.86 (17.19)	305	50.10 (15.87)	114	62.61 (15.03)	**<0.001**
Gender male, No. (%)	844	445 (52.73)	425	228 (53.65)	305	160 (52.46)	114	57 (50.00)	0.781
Deceased donor, No. (%)	844	719 (85.19)	425	359 (84.47)	305	260 (85.25)	114	100 (87.72)	0.686
ECD, No. (%)	844	315 (37.32)	425	145 (34.12)	305	94 (30.82)	114	76 (66.67)	**<0.001**
Transplant baseline characteristics									
Cold ischemia time (hours), mean (SD)	838	17.05 (9.02)	425	17.39 (9.44)	304	15.95 (8.15)	109	18.80 (9.36)	**0.010**
HLA A/B/DR mismatch (number), mean (SD)	844	3.84 (1.32)	425	3.92 (1.28)	305	3.90 (1.31)	114	3.31 (1.39)	**<0.001**
Characteristics at the time of diagnosis									
Time of biopsy since Transplantation (years), mean (SD)	844	0.95 (1.51)	425	0.95 (1.40)	305	0.63 (0.85)	114	1.78 (2.58)	**<0.001**
eGFR (mL/min/1.73m^2), mean (SD)	844	44.36 (16.68)	425	47.07 (16.02)	305	45.58 (16.31)	114	31.01 (13.69)	**<0.001**
Proteinuria (g/g creatinine), mean (SD)	844	0.43 (0.73)	425	0.44 (0.63)	305	0.42 (0.88)	114	0.38 (0.61)	0.681
Anti-HLA DSA MFI at diagnosis, No. (%)	844		425		305		114		
<500		334 (39.57)		0		239 (78.36)		95 (83.33)	
≥500-3000		410 (48.58)		344 (80.94)		49 (16.07)		17 (14.91)	
≥3000-6000		39 (4.62)		32 (7.53)		6 (1.97)		1 (0.88)	
≥ 6000		61 (7.23)		49 (11.53)		11 (3.61)		1 (0.88)	**<0.001**
Delay of control biopsy after diagnosis (months)**,** mean (SD)	844	6.79 (8.29)	425	7.04 (8.35)	305	7.35 (9.32)	114	4.35 (3.14)	**0.003**
Graft losses, No. (%)	844	174 (20.62)	425	116 (27.29)	305	53 (17.38)	114	5 (4.39)	**<0.001**

Abbreviations: ABMR: Antibody-mediated rejection; TCMR: T-cell mediated rejection; CI: confidence interval; DSA: donor-specific anti-HLA antibodies; ECD: Expanded Criteria Donor; eGFR: estimated Glomerular Filtration Rate; ESRD: end-stage renal disease; HLA: human leucocyte antigen; MFI: mean fluorescence intensity

Supplementary Table G: iBox risk score comparison of previously published risk scores. A comprehensive search strategy was conducted through several databases (PubMed, Medline, Embase, Cochrane, and Scopus) without date restrictions for publications up to July 25, 2018 for allograft survival scoring systems among kidney transplant recipients. We used the search terms “kidney transplantation”, “allograft survival” and “prognostic score”. Out of 460 articles identified, 11 were related to long-term allograft survival, 5 were externally validated and only 2 comprised immunological parameters. They are presented in the table and compared with the iBox risk prediction score. The two studies identified: i) were not derived from patient cohorts with systematic monitoring and specific design towards risk stratification; ii) did not integrate a large spectrum of potential prognostic factors, iii) were not validated in multiple large cohorts worldwide with different transplant allocation systems and management practices, iv) were not validated in randomised controlled therapeutic clinical trials (RCTs).

STUDY	Trial Registration /protocol	Study Design	Population	Number of Centres involved	External Validation cohort	Time of risk evaluation	Follow-up time post-transplant	Validation in therapeutic randomised controlled clinical trials (RCT)	Candidate Predictors Evaluated	Data set qualification (data quality)	Allograft phenotypes at the time of risk assessment	CSTAT validation in the iBox cohort	Individual risk prediction tool and interfacing for patients and clinicians
Gonzales et al*	None	Retrospective	n=556 (1999 – 2008)	1	No	Fixed at 1 year after transplant	Median: not applicable	No	17	Not audited	Yes (Banff international classification)	0.69§	No
Premaud et al†	None	Retrospective	n=664 (1984 – 2011)	3	Yes n=896 France only	Fixed at 1 year after transplant +2 adjustable variables	Median: 6.4 years	No	12	Not audited	No	0.67‖	No
iBox Risk score trial	Clinical trial.gov #NCT03474003	Prospective observational	n=4,000 (2000 – 2014)	10	Yes n=3,557 Europe and US	Time adjusted‡	Median: 7.65 years (IQR: 5.20-10.30)	Yes 3 RCT (NCT01079143, EudrCT 2007-003213-13 and NCT01873157)	33	Annual audit	Yes (Banff international classification)	-	iBox risk prediction score individual calculation tools for clinicians and patients

* Gonzales MM, Bentall A, Kremers WK, Stegall MD, Borrows R. Predicting Individual Renal Allograft Outcomes Using Risk Models with 1-Year Surveillance Biopsy and Alloantibody Data. J Am Soc Nephrol. 2016;27(10):3165-74, † Premaud A, Filloux M, Gatault P, Thierry A, Buchler M, Munteanu E, et al. An adjustable predictive score of graft survival in kidney transplant patients and the levels of risk linked to de novo donor-specific anti-HLA antibodies. PloS one. 2017;12(7):e0180236, ‡ see **Appendix Figure 1** for the distribution of iBox time post-transplant risk evaluation, **Table 2B** for the inclusion of “time of risk evaluation post-transplant” in the final model, and **Appendix Figure 3** showing examples of time updated iBox risk evaluation in patients (Patient #**3** and #**4**)

§Performance in the iBox cohort of the Mayo Histology-based model (NBMM) published by Gonzales et al. This risk score is applied in a retrospective cohort including immunological and histological parameters (7 prognostic variables assessed at 1-year post transplant: eGFR (mL/min/1.73m²), proteinuria (estimated with urinary albumin-to-creatinine ratio UACR, g/L), acute rejection, race, recipient age, g Banff score and ci Banff score). The C-stat of the final model= 0.90. This score was assessed in patients who received a solitary kidney transplant between 1999 and 2008. Out of the iBox derivation cohort, 3,569 patients fitted the inclusion & exclusion criteria (231 patients excluded because of no information on albuminuria for NBMM calculation).

- Calculation of the NBMM score in the IBox cohort.
Authors based NBMM score on survival Cox models (univariate and multivariable analyses). They provided in the article the weighted coefficients for the 7 variables associated with death-censored transplant failure and the transformation needed for each variable.

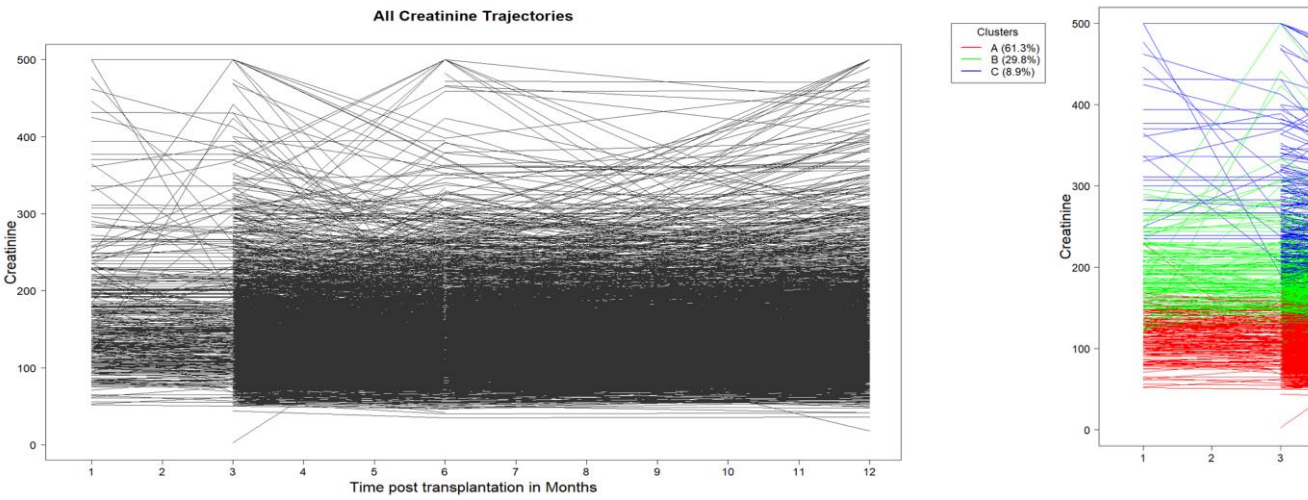
The formula applied in the iBox cohort was therefore the following one:

$$\mathbf{NBMM_{IBox\ cohort}} = \beta_{UACR} * (\log_{10}(UACR-46)) + \beta_{eGFR} * ((eGFR-47)/10) + \beta_{eGFR^2} * ((eGFR-47)/10)^2 + \beta_{Rejection} * Rejection + \beta_{Black\ ethnicity} * Black\ ethnicity + \beta_{Recipient\ Age} * ((Recipient\ age-46)/10) + \beta_{UACR\ with\ rejection\ interaction} * (\log_{10}(UACR-46) - 0.46\ for\ rejection) + \beta_{g\ score} * g + \beta_{ci\ score} * ci$$

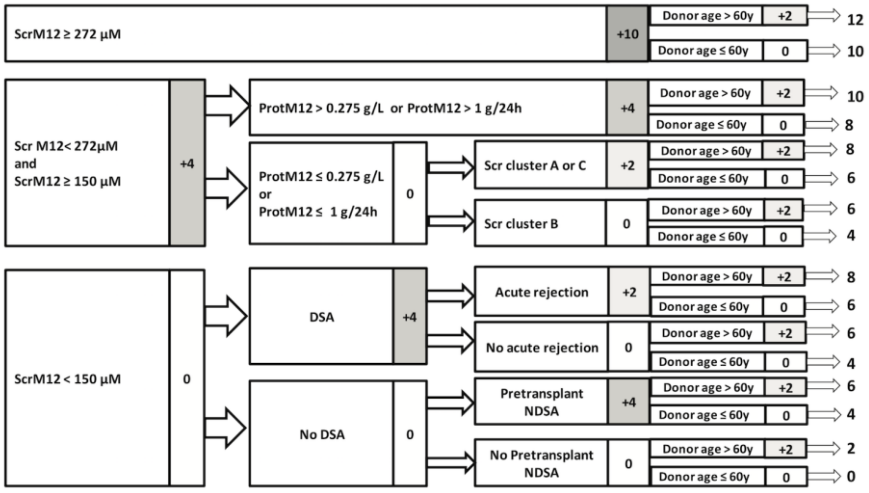
With βs representing the weighted coefficients.

We finally calculated the NBMM score in the iBox cohort and determined the C stat of model= 0.69.

|| **Performance in the iBox cohort of the AdGFS score published by Premaud et al.** The AdGFS includes 7 prognostic variables: i) 5 assessed at 1-year post transplant: serum creatinine (μM), proteinuria (g/L), creatinine cluster, donor age and pretransplant sensitization. ii) 2 adjustable variables: de novo anti-HLA DSA and acute rejection episode (C-stat of the final model=0.83). This score was assessed only in patients without DSA at the time of transplant and who exclusively received kidneys from deceased donors. The patients were transplanted between 1984 and 2011.
Population of IBox reference set fitting the inclusion exclusion criteria of AdGFS score: n= 2,855 (All with available data for AdGFS calculation).
- Building of the creatinine cluster in the iBox cohort
We applied the creatinine clusters defined in the AdGFS creation process by calculating their centres in the specified timepoints (M1, M3, M6, Y1) and assessing in the IBox cohort the clusters of assignment based on k-means analysis.



- Calculation of the AdGFS score in the IBox cohort.
We used the scoring system defined by the authors to calculate the AdGFS score in the IBox cohort.



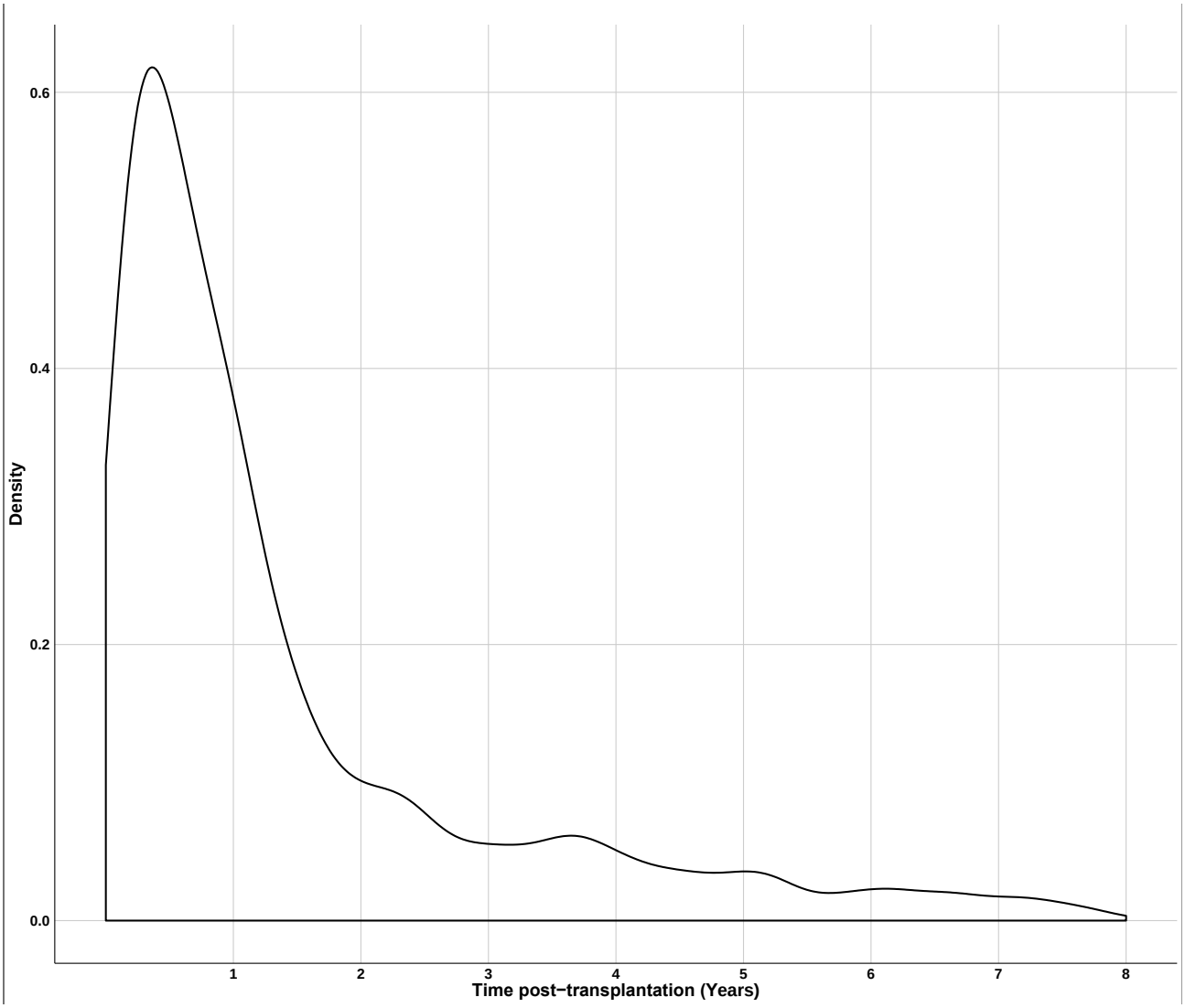
- We finally calculated the AdGFS score in the IBox cohort: C stat of the model= 0.71.

Supplementary Table H: Independent determinants of kidney allograft loss in the derivation cohort using histological diagnoses instead of Banff international classification lesions grading system: multivariable analysis

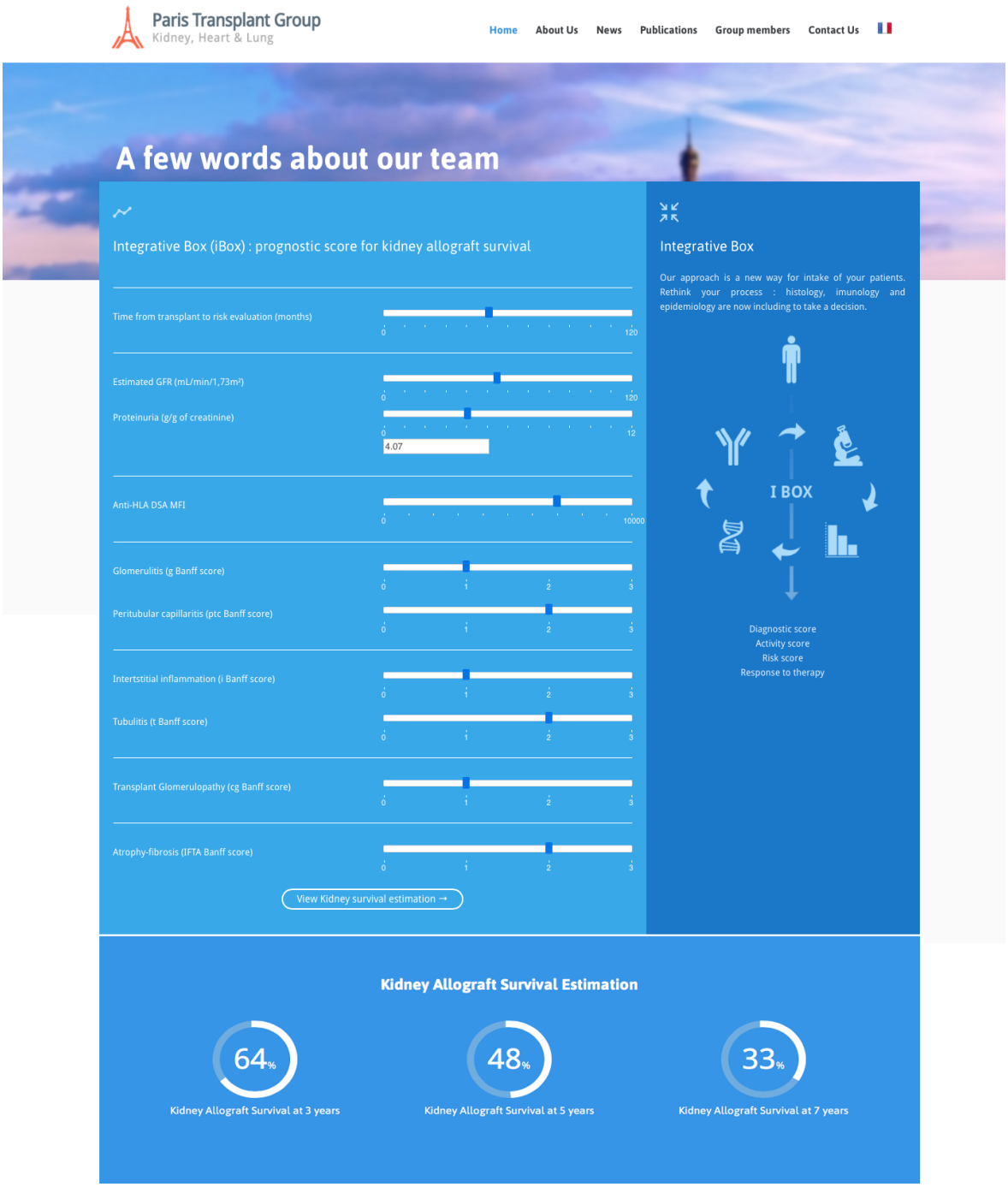
		Number of patients	Number of events	HR	95% CI	p
Time from transplant to evaluation (year)		3,997	548	1.097	(1.043-1.153)	**0.0003**
eGFR (mL/min/1.73 m^2)		3,997	548	0.955	(0.949-0.961)	**<0.0001**
Proteinuria (log)		3,997	548	1.552	(1.443-1.670)	**<0.0001**
Antibody-mediated rejection	**No**	3,398	368	1	-	
	Yes	599	180	1.811	(1.475-2.223)	**<0.0001**
T-cell mediated rejection	**No**	3,810	502	1	-	
	Yes	187	46	1.369	(1.007-1.861)	**0.0453**
Nephropathy Recurrence	**No**	3,867	510	1	-	
	Yes	130	38	1.680	(1.199-2.355)	**0.0026**
BK virus associated nephropathy	**No**	3,900	517	1		
	Yes	97	31	1.450	(1.000 -2.107)	**0.0500**
Anti-HLA donor-specific antibody mean fluorescence intensity	**<500**	3,309	393	1	-	
	≥500 – 3,000	483	82	1.220	(0.946-1.572)	
	≥3,000 – 6,000	82	24	1.527	(0.993-2.348)	
	≥6,000	123	49	1.985	(1.432-2.753)	**0.0003**

3 SUPPLEMENTARY FIGURES

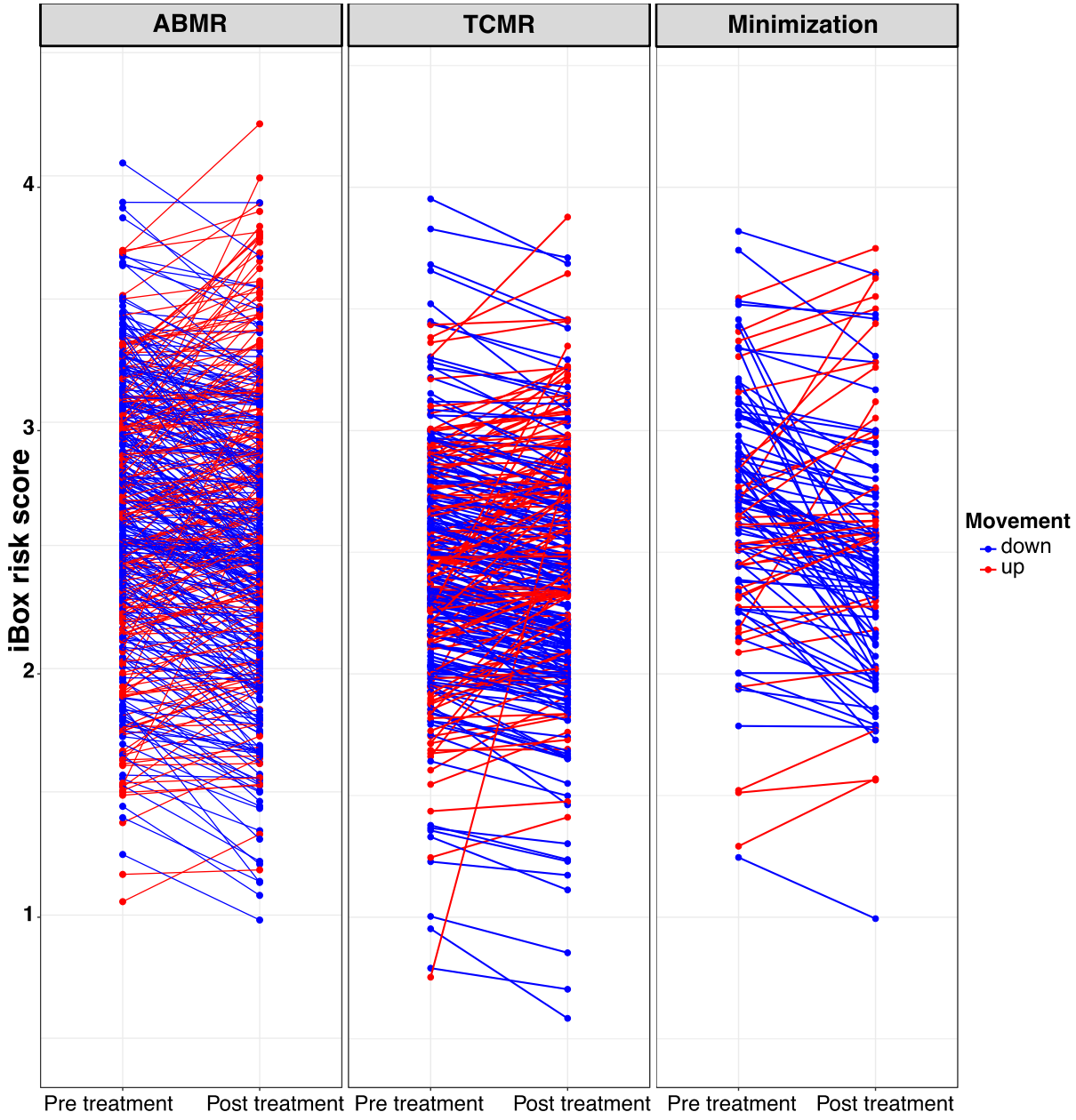
Supplementary Figure A: Density of risk evaluation time points after transplantation



Supplementary Figure B: iBox practical application for clinicians: Ready-to-use interface for clinicians.



Supplementary Figure C: Effect of treatment intervention on iBox risk score. This analysis shows the iBox risk score assessed at the time of therapeutic intervention and after therapeutic intervention in the 3 clinical scenarios including antibody–mediated rejection, T-cell-mediated rejection and calcineurin inhibitor minimization. Blue lines correspond to a decrease in iBox risk score after treatment. Red curves correspond to an increase in iBox risk score after treatment. The iBox prediction capability post-treatment was accurate in these 3 therapeutic scenarios (C-index 0.81; 95% bootstrap percentile CIs=0.77 to 0.85). The calibration plot showed a good agreement between the iBox prediction model after therapeutic intervention and the actual observation of kidney allograft loss (calibration intercept at 7-year: 0.0121 ; calibration slope at 7-year: 0.971, 95%CI [0.719 to 1.224]).



Abbreviations: ABMR, antibody-mediated rejection; TCMR, T Cell mediated rejection.