

Appendix 1: Queries

1. For Scopus research

```
( TITLE-ABS ( transplant ) OR TITLE-ABS ( transplantation ) OR TITLE-ABS ( graft ) OR TITLE-ABS ( grafting ) )
AND
( TITLE-ABS ( kidney ) OR TITLE-ABS ( renal ) )
AND
( TITLE-ABS ( prediction ) OR TITLE-ABS ( predict ) OR TITLE-ABS ( predictive ) OR TITLE-ABS ( probability ) OR
TITLE-ABS ( prognosis ) OR TITLE-ABS ( prognostic ) OR TITLE-ABS ( prognostication ) OR TITLE-ABS ( score ) OR
TITLE-ABS ( scores ) )
AND
( TITLE-ABS ( model ) OR TITLE-ABS ( models ) OR TITLE-ABS ( regression ) OR TITLE-ABS ( equation ) OR TITLE-ABS
( equations ) OR TITLE-ABS ( modeling ) OR TITLE-ABS ( modelling ) OR TITLE-ABS ( score ) OR TITLE-ABS ( scores )
OR TITLE-ABS ( probability ) OR TITLE-ABS ( prognosis ) OR TITLE-ABS ( prognostic ) OR TITLE-ABS (
prognostication ) )
AND
( TITLE-ABS ( failure ) OR TITLE-ABS ( loss ) OR TITLE-ABS ( death ) OR TITLE-ABS ( mortality ) OR TITLE-ABS (
survival ) )
AND
( TITLE-ABS-KEY( prediction ) OR TITLE-ABS-KEY( predict ) OR TITLE-ABS-KEY( predictive ) OR TITLE-ABS-KEY (
predicting ) OR TITLE-ABS-KEY ( validation ) OR TITLE-ABS-KEY ( validity ) OR TITLE-ABS-KEY ( validated ) OR TITLE-
ABS-KEY ( cross-validation ) OR TITLE-ABS-KEY ( selection ) OR TITLE-ABS-KEY ( calibration ) OR TITLE-ABS-KEY (
discrimination ) OR TITLE-ABS-KEY( discriminates ) OR TITLE-ABS-KEY ( ROC ) )
```

2. For PubMed research

```
("transplant" [Title/Abstract] OR "transplantation" [Title/Abstract] OR "graft" [Title/Abstract] OR "grafting"
[Title/Abstract])
AND
("kidney" [Title/Abstract] OR "renal" [Title/Abstract])
AND
( " prediction" [Title/Abstract] OR " predict" [Title/Abstract] OR " predictive" [Title/Abstract] OR "probability"
[Title/Abstract] OR " prognosis" [Title/Abstract] OR "prognostic" [Title/Abstract] OR "prognostication"
[Title/Abstract] OR " score " [Title/Abstract] OR " scores" [Title/Abstract] )
AND
( "model" [Title/Abstract] OR "models" [Title/Abstract] OR "regression" [Title/Abstract] OR "equation"
[Title/Abstract] OR "equations" [Title/Abstract] OR "modeling" [Title/Abstract] OR "modelling" [Title/Abstract]
OR "score" [Title/Abstract] OR "scores" [Title/Abstract] OR "probability" [Title/Abstract] OR "prognosis"
[Title/Abstract] OR " prognostic" [Title/Abstract] OR "prognostication" [Title/Abstract] )
AND
( " failure" [Title/Abstract] OR " loss" [Title/Abstract] OR " death" [Title/Abstract] OR " mortality"
[Title/Abstract] OR " survival" [Title/Abstract] )
AND
( "prediction" [Title/Abstract] OR " predict " [Title/Abstract] OR "predictive" [Title/Abstract] OR " predicting "
[Title/Abstract] OR " validation" [Title/Abstract] OR " validity" [Title/Abstract] OR " validated " [Title/Abstract]
OR " cross-validation " [Title/Abstract] OR " selection " [Title/Abstract] OR " calibration" [Title/Abstract] OR "
discrimination" [Title/Abstract] OR " discriminates" [Title/Abstract] OR " ROC " [Title/Abstract] )
```

Appendix 2 : Data extraction table												
Author and years (journal)		Akl et al. 2008 (24) (Transplantation)		Bang et al. 2010 (60) (Yonei Medicine Journal)		Baskin-Bey et al. 2007 (25) (American Journal of Kidney Diseases)		Bodonyi-Kovacs et al. 2010 (26) (Human Immunoly)		Brown et al 2012 (57) (American Journal of Nephrology)		
Type of study		Development and validation of a new model		Validating with model updating		Development and validation of a new model		Development and validation of a new model		Development and validation of a new model		
Donor type		Living donors		Deceased donors		Deceased donors		Living donors		Deceased donors		
Recipient age range		Adult (Age range unspecified)		Adult (Age range unspecified)		Adult (Age range unspecified)		Unspecified		≥ 18 years		
OUTCOMES	Studied event		Graft failure (Dialysis/re-transplantation/death)		Death censored graft failure defined by eGFR less than 10 mL/min/1.73 m2 or conversion to dialysis		1.Death with or without functioning graft 2.Death censored graft failure (dialysis/re-transplantation)		1.Graft failure (dialysis/re-transplantation/death) 2.Death censored graft failure		Graft failure (dialysis/re-transplantation/death)	
	Competing event		No		Death and re-transplantation		1. Dialysis/re-transplantation 2. Death with functioning graft		1.No 2.Death with functioning graft)		No	
	Accounted for competing risk				Not clearly specified		Not clearly specified		Not clearly specified			
	Time horizon of prediction		5 years		6 months, 1year		8 years		1 – 4 years		1, 3, > 3	
Development cohort type		Single center cohort		Multi-centers cohort		Registry: United Network for Organ Sharing Standard Transplant Analysis and Research		Single center cohort		Registry: United States Renal Data System		
Years of Transplantation		1976 – 1992; 1995 - 2007		1994 - 2008		1995 - 2002		Unspecified		2000 - 2001		
Sample size		1581		337		36201		73 patients (but 75 allografts)		5144		
Statistical model used		1.Cox regression model 2. Artificial neural network (ANN)		Linear regression model		Cox regression model		Logistic regression model		Bayesian Belief Networks (BBNs) using prior information on train population (probability of graft failure at 1 year, 3 years and more than 3 years) Age at diabetes diagnosis, Recipient age at transplant , Body mass index, Cardiac arrest since neurological event that led to declaration of brain death, Cold ischemia pump ki, Combined Cold ischemia pump, Creatinine decline by 0.25 or more in the first 24 h, Donor blood type, Donor age, BMI of cadaveric donor, Cadaveric donor cause of death, Serum creatinine of cadaveric donor, Primary diagnosis, History of diabetes, History of hypertension, Cigarette use, Patient diabetes, Dialysis at listing (recipient), Length of time on dialysis pre-transplant in days, Was this a DCD donor?, Other drug use, Donor race, Donor gender, Estimated warm ischemic time, Graft thrombosis, Cocaine use – ever Induction medications combination, Procedure type, Pump, Are any medicines given currently for maintenance or anti-rejection , CNT HLA A, B, DR mismatch, Most recent serum creatinine prior to discharge, Most recent absolute creatinine at listing, Most recent USRDS PRA(%), Pre-transplant dialysis, Recipient blood type, Recurrent disease, Recipient race, Gender, Serum creatinine at time of transplant, Drug-treated systemic hypertension at listing, Warm ischemia anastomotic time		
DEVELOPMENT COHORT	Predictors used in the final model		Predictors measured at transplantation		1. For Cox model: recipient age, donor age, haplotype, time to diuresis, total steroid dose, immunosuppression, acute tubular necrosis , number of acute rejection 2. For artificial neural networks: recipient age, donor age, haplotype, time to diuresis, total steroid dose, immunosuppression, acute tubular necrosis, number of acute rejection, number of blood transfusions, HLA-A, B, HLA-DR		Age, creatinine clearance, cause of donor death, history of hypertension, HLA mismatch, body weight		Recipient age, history of diabetes mellitus, history of angina, time on dialysis therapy		At two years: 1. Graft failure: gene expression (TNF- α , Bcl-2, IFN- γ), clinical variables (Delayed Graft Function, acute Rejection, systolic blood pressure) 2. Death censored graft failure: no significant covariate	
											At 3 and 4 years: No significant covariate	
											Systolic and diastolic blood pressure values around the 1-year after transplantation	
											Not clearly specified	
Method used to account for post-transplant predictors										Not clearly specified		
Prediction tools presentation form		1.Nomogram for Cox model 2.Original model for ANN		Score (Nyberg scoring system modified)		Recipient risk score (RRS) Threshold defined to separate into 4 risk groups based on expected survival after transplantation. Four recipients groups (RG): - RG1 = RRS<2.555 - RG2 = 2.555 < RRS < 3.308 - RG3 = 3.308 < RRS < 3.802 - RG4 = RRS >3.802 Decreasing median survival from RG1 to RG4		Original model		Original model		
Proposed thresholds for clinical use / Proposed software or online-calculator												
VALIDATION COHORT	Internal validation		Yes		No		Updated version 1-year RRS (Lorent et al Plos One 2016) in an online calculator at www.divat.fr		Yes		Yes	
	Internal validation type		Unspecified				Unspecified		10 fold cross validation		1.10 fold cross validation 2.Split sample 70%-30%	
	External validation		Yes		Yes		No		No		Yes	
	External validation type		Temporal		Spatial						Temporal	
	External validation cohort (sample size)		Transplanted in 1992 -1995 (319)		Transplanted in Asan medical center in 1995 – 2003 (102)						Transplanted in 1997, 2002 and 2003	
PER	Overall											
	Discrimination		validation models AUC Sensitivity /specificity (%)				AUC=0.692 for death model		AUC at two years prediction for graft failure Gene expression only: 0.577		Validation Year AUC Sensitivity/speci ficity	

Author and years (journal)		Carl van Walraven et al. 2010 (27) (Canadian Medical Association Journal)	Dahle et al 2015(29) (Transplantation)	De Vusser et al 2013 (30) (Journal of American Society of Nephrology)	Einecke et al. 2010 (31) (Journal of Clinical Investigation)	Foucher et al. 2010 (33) (Kidney International)
Type of study		Development and validation of a new model (in both transplanted and dialysis patients)	Development and validation of a new model	Development and validation of a new model	Development and validation of a new model	Development and validation of a new model
Donor type		Living and deceased donors	Unspecified	Unspecified	Unspecified	Deceased donors
Recipient age range		Age ≥ 18 years and < 100 years	≥16 years	Adult (Age range unspecified)	Unspecified	≥ 18 years
OUTCOMES	Studied event	Death with or without functioning graft	Death (Any cause)	1. Graft failure (dialysis/re-transplantation/death) 2. Death-censored graft failure	Death censored graft failure (dialysis-re-transplantation)	Death censored graft failure (Dialysis/re-transplantation)
	Competing event	Dialysis/re-transplantation	Dialysis/re-transplantation	1. No 2. Death with functioning graft	Death with functioning graft	Death with functioning graft
	Accounted for competing risk	Not clearly specified	Not clearly specified	Not clearly specified	Not clearly specified	Not clearly specified
Time horizon of prediction			Not reported (5 years an example in supplement material)	3, 5 and 10	Not reported	8 years (after first anniversary of transplantation)
Development cohort type		Multi-center cohort	Single center cohort	Single center	Multi – center	Multi – center Cohort
Years of transplantation		1995 - 2006	2007 - 2012	1991 – 2004	2004 – 2007 (Alberta) 2006- 2007 (Illinois)	1996 - 2007
Sample size		84724	1497	181 transplanted before 2004	105 biopsies	2169
Statistical model used		Cox regression model	Cox regression model	Logistic regression model (Used logistic regression to build a score after Cox model analysis)	Cox regression model	Cox regression model
DEVELOPMENT COHORT	Predictors measured at transplantation	Age, race, cause of kidney failure, body mass index, comorbid disease, smoking, employment status, serum albumin level, year of first renal replacement therapy, kidney transplantation, time to transplant waiting-list, time on the waiting list	Recipient age and sex, coronary heart disease, smoking status, time in renal replacement therapy, creatinine, diabetes, pulse pressure	For graft failure: interstitial fibrosis, tubular atrophy, donor age, glomerulosclerosis For death censored graft failure: interstitial fibrosis, tubular atrophy, glomerulosclerosis	Molecular risk score, peritubular capillary basement membrane multi-layering, arteriolar hyalinosis, proteinuria. Note: date at biopsy is a baseline date	Donor creatinine, recipient age, recipient sex, number of previous transplantations,
	Predictors measured after transplantation		Pulse wave velocity at 8 weeks after transplantation			Acute rejection during the first year post-transplant, Creatinine at 3 month, creatinine at 12 month, proteinuria at 12 month
	Method used to account for post-transplant predictors		Used the predictor as a baseline variable			Only patients alive with transplant function at 12 months were selected to develop the model
Prediction tools presentation form		Index score	Original model	Index score (Leuven donor risk score)	Score	Kidney Transplant Failure Score (KTFS) Threshold defined at 4.77 after sensitivity/specificity analysis. Two risk grade groups: - Grade1 : KTFS less than 4.77 (Lower risk group) - Grade2 : KTFS more than 4.77 (Higher risk group)
Proposed thresholds for clinical use / Proposed software or online-calculator				Threshold defined at 47 after sensibility/specificity analysis	Threshold defined at median score. Two risk groups: - Lower risk group less than score median - Higher risk group more than score median	Online calculator at www.divat.fr
VALIDATION COHORT	Internal validation	Yes	Yes	No	Yes	Yes
	Internal validation type	Split-sample 50%, 50%	Bootstrapping		10 fold Cross-validation	Cross-validation
	External validation	No	No	Yes	Yes	Yes
External validation type				Temporal	Spatial	Spatial
External validation cohort (sample size)				Transplanted after 2004 (n=367)	Recipients transplanted between 2006 and 2007 from Minnesota (48 biopsies)	Transplanted from Caen, Grenoble, Tours, and Strasbourg (317)
Overall			Nagelkerke's R2 : 0.12 0.12 for prediction at 5 years			
PERFORMANCES	Discrimination	AUC 0.746	AUC 0.78 for overall model 0.78 for prediction at 5 years	AUC for graft failure: For development cohort 0.70 at 3yrs, 0.67 for 5 years and 0.60 for 10 yrs For validation cohort 0.70 at 3 yrs, 0.81 at 5 yrs, Not calculated at 10 years because 10-year survival data were not available in the validation cohort Sensitivity/specificity at 5 years for validation cohort 85%/81% at 47 threshold: 90%/60% at 60 threshold	AUC 0.83 for internal validation sensitivity of 0.83 and specificity 0.63	AUC at 8 years: - 0.78 for internal validation - 0.78 for external validation For internal validation (KTFS cut off at 4.17): Sensibility 0.72, specificity 0.71
Calibration		Agreement between index score and risk of death at five years	Calibration slope 0.89 for overall model; 0.87 for prediction at 5 years		Agreement of predicted risk and score of graft failure	
Other performance indicator						
Other studies used for comparison				AUC evaluated in same population At 3 year At 5 years At 10 years Remuzzi et al 2006 0.62 0.62 0.59 Lopes et al 2005 0.55 0.59 0.59 Snoeijns et al 2008 0.6 0.6 0.59		
Number of citation in google scholar at 31/04/16		30	5	18	91	25

Author and years (journal)		Fritsche et al. 2005 (34) (Transplantation Proceeding)	Grams et al. 2012 (35) (Clinical Investigation)	Greco et al. 2010 (36) (Transplantation Proceeding)	Gourishankar et al 2013 (59) (Clinical Transplantation)	Gusukuma LW, 2014 (54) (Journal Brasileiro de Nefrologia)
Type of study		Development and validation of a new model	Development and validation of a new model	Development and validation of a new model	Validation without any model updating	Development and validation of a new model
Donor type		Living and deceased donor	Donor type unspecified	Donor type unspecified	Deceased donor	Deceased donor
Recipient age range		Adult (Age range unspecified)	aged > 65	≥ 18 years	Adult (Age range unspecified)	>18 years
OUTCOMES	Studied event	Death censored graft failure (Dialysis/re-transplantation)	Death with or without functioning graft	Graft failure (undefined)	Death censored graft failure (dialysis/re-transplantation)	Graft failure (death or graft with creatinine > 1.5 mg/dl)
	Competing event	Death with functioning graft	Dialysis/re-transplantation	Depends on graft failure definition	Death with functioning graft	Return to dialysis /re-transplantation
	Accounted for competing risk	Not clearly specified	Not clearly specified	Not clearly specified	Not clearly specified	Not clearly specified
Horizon of prediction		4 years	3 years (1, 2, and 5 years for sensitivity)	5 years	Long-term (unspecified)	6 months
DEVELOPMENT COHORT	Development cohort type	Multi-center Cohort	Registry: United States Renal Data System	Single center cohort	Single center cohort	Single center
	Years of transplantation	1981 - 2004	1999 - 2006	Unspecified	1990 - 2004	February - November 2011
	Sample size	497	6988	194		311
	Statistical model used	Decision-tree Hill-Climbing algorithm	Logistic regression model	Decision tree model	Calculation of individual score value	Logistic regression
	Predictors measured at transplantation		Age, dialysis time, sex, congestive heart failure, cardiac arrhythmia, complicated diabetes mellitus, chronic pulmonary disease, coronary artery disease, hypertension, drug dependence, peripheral vascular disease, polycystic kidney disease, deficiency anemia, other neurological disorder, liver disease, depression, peptic ulcer disease, current smoking, year of transplantation	Delayed graft function, acute rejection episode, and chronic allograft nephropathy, body mass index	Deceased donor score (DDS) (Nyeberg et al, 2003) Donor risk score (DRS) (Schold et al, 2005)(51) kidney donor risk index (KDRI){ (Rao et al, 2009)(45)	Sex, recipient weight, DR mismatch, public aid/welfare, patient monthly income, to have a children, family support, ECG alteration, donor death etiology, donor age
	Predictors used in the final model	Cold ischemia time, single kidney transplantation, creatinine mean and slope				
	Predictors measured after transplantation	Creatinine at 60 to 180 day post-transplant (Used median and slope creatinine between day 60 and 180 after transplantation)				
	Method used to account for post-transplant predictors	Not clearly specified				
	Prediction tools presentation form	Original model	Original model	Original model	Existing score	Score
	Proposed thresholds for clinical use / Proposed software or online-calculator					No threshold, but proposed a simple score derivation (range 0-56), and a probability of success of a transplantation for a given score
VALIDATION COHORT	Internal validation	Yes	Yes	Yes	No	Yes
	Internal validation type	Split-sample 50%, 50%	1. Five-fold cross-validation 2. Random forest	Leave-one out cross validation		unspecified
	External validation	No	Yes	No	Yes	No
	External validation type		Spatial		Fully	
	External validation cohort (sample size)		Transplanted from scientific transplant registry database (2728)		730	
PERFORMANCES	Overall					
	Discrimination	When decision tree algorithm applied: Sensitivity 42.3% for creatinine cut-off at 3.1mg/dl When Hill-Climbing algorithm applied for creatinine cut-off at 1.8 mg/dl: Sensitivity 69.5%, specificity 79.0%	AUC: - 0.66 in development cohort - 0.68 by cross-validation	88.2% sensitivity 73.8% specificity		AUC=0.817
	Calibration		Hosmer-Lemeshow test (p-value=0.44) for 3 years		Agreement between each score level and observed graft survival	Hosmer-lemeshow test (p=0.672) Agreement between estimated probability and observed graft survival (r2=0.982)
	Reclassification					
Other performance indicator						
Other studies used for comparison						
Number of citation in google scholar at 31/04/16		8	29	6	7	0

Author and years (journal)		Hemke et al. 2013 (37) (BMC Nephrology)	Hernández et al. 2005 (38) (Transplantation)	Hernández et al. 2009 (39) (Transplantation)	Ho et al. 2013 (40) (Transplantation)	Jassal et al. 2005 (62) (American Journal of Kidney Diseases)
Type of study		Development and validation of a new model (in both transplanted and dialysis patients)	Development and validation of a new model		Development and validation of a new model	Validation with updating the model
Donor type		Donor type unspecified	Deceased donor	Donor type unspecified	Donor type unspecified	Living and deceased donor
Recipient age range		≥ 16 years	Adult (Age range unspecified)	≥ 18 years	Adult (Age range unspecified)	Adult (Age range unspecified)
OUTCOMES	Studied event	Death with or without functioning graft	Death(any cause)	Death(any cause)	Death censored graft failure (dialysis/re-transplantation)	Death with or without functioning graft
	Competing event	Dialysis/re-transplantation	Dialysis/re-transplantation	Dialysis/re-transplantation	Death with functioning graft	Dialysis/re-transplantation
	Accounted for competing risk	Not clearly specified	Not clearly specified	Not clearly specified	Not clearly specified	Not clearly specified
Time horizon of prediction		3, 5 and 10 years	2, 5 and 8 years	3 years	Long-term (unspecified)	Not reported
DEVELOPMENT COHORT	Development cohort type	Registry: Dutch renal replacement registry	Single center Cohort	Single center Cohort	Single center Cohort	Registry: Canadian Organ Replacement Registry (CORR)
	Years of transplantation	1995 – 2005 (date of start RRT)	1981 - 2001	1990 - 2002	1997 - 2008	1988 - 1998
	Sample size	6934	646	2452	231	6324
	Statistical model used	Cox regression model	Cox regression model	Cox regression model	Cox regression model	Cox regression model Comorbidity index:
	Predictors measured at transplantation	Age, sex, primary renal disease and therapy at 90 days (Therapy at 90 days was used as baseline predictor)	Age, pretransplant cardiovascular disease, cardiac hypertrophy, vascular calcification, diabetes, time on dialysis	Recipient age, hepatitis C infection and pre-transplant diabetes	1.for overall population: Delayed Graft Function and Donor-specific antibody 2.for recipient who have available clinical data at six months.; Delayed Graft Function and Recipient age	Model 1. Charlson Comorbidity Index (CCI) (Charlson et al, 1987) Model 2. Khan index (Khan et al, 1993) Model 3. Davies index (Davies et al 2002) Model 4. Modified CCI (Hemmelgarn et al, 2003) Model 5. Categorized CCI Model 6: Log-transformed CCI
	Predictors used in the final model					All models were adjusted for covariates (age, sex, cause of renal disease)
	Predictors measured after transplantation		Acute tubular necrosis and renal function at discharge expressed as serum creatinine in the first week after transplantation	Proteinuria and serum creatinine at one-year post-transplant, use of tacrolimus or mycophenolate mofetil during the first year post-transplant, new onset of diabetes after transplantation	Urinary Monocyte chemoattractant protein 1 (CCL2), at 6 months	
	Method used to account for post-transplant predictors		Not clearly specified	Not clearly specified	Only patients alive with transplant function at 6 months were selected to develop the model	
	Prediction tools presentation form	Index score	Index score	Index score	Original model	Updating existing index
	Proposed thresholds for clinical use / Proposed software or online-calculator	Threshold defined at tertiles points Calculate individual prognostic score	Threshold defined at tertiles points Three risk group (score total = 8 points): - Lower risk :Index score < 2nd tertile - Median risk : Index score between 1th and 2nd tertile - High risk: Index score > 1th tertile	Threshold defined at quintiles points. Four risk groups (score rang: 40- 200): - Group1 : index score <40 - Group2 : index score from 40 to 120 - Group3 : index score from 120 to 200 - Group4 : index score >200		
VALIDATION COHORT	Internal validation	Yes	Yes	Yes	Yes	Yes
	Internal validation type	Split-sample 50%, 50%	Split-sample 50%, 50%	Split-sample 50%, 50%	unspecified	Cross-validation
	External validation	No	No	No	No	No
	External validation type					
	External validation cohort (sample size)					
PERFORMANCES	Overall					
	Discrimination	AUC 0.70 at 3 years 0.72 at 5 years 0.72 at 10 years	AUC 0.60 for development group 0.63 for validation group	AUC 0.75 for development group 0.74 for validation group	AUC 0.89 for overall population 0.87 for recipient who have available clinical data at six months	AUC 0.74 for model 6
	Calibration	Calibration slope: 0.948 for 3 years, 0.99 for 5 years and 1.025 for 10 years. Agreement between observed outcomes and predicted survival probabilities	Agreement between observed outcomes and predicted survival probabilities			
	Reclassification					
	Other performance indicator					
Other studies used for comparison						
Number of citation in google scholar at 31/04/16		2	32	26	12	27

Author and years (journal)		Kasiske, et al. 2010 (41) (American Journal of Kidney Diseases)	Kikić et al. 2014 (42) (European Journal for Clinical Investigation)	Krikov et al.2007 (43) (American Society for Artificial Internal Organ Journal)	Lin, et al.2008 (32) (Journal of Biomedical Informatics)	Loupy et al. 2014 (46) (Journal of American Society of Nephrology)
Type of study		Development and validation of a new model	Development and validation of a new model	Development and validation of a new model	Development and validation of a new model	Development and validation of a new model
Donor type		Deceased donor	Donor type unspecified	Donor type unspecified	Living and deceased donors	Donor type unspecified
Recipient age range		≥ 18 years	Adult (Age range unspecified)	Pediatric and Adult (Age range unspecified)	≥ 18 years	Adult (Age range unspecified)
OUTCOMES	Studied event	1. Graft failure (dialysis/re-transplantation/death) 2. Death censored graft failure (dialysis/re-transplantation) 3. Death with functioning graft	1.Death (with and without function graft 2.Death censored graft failure	Death censored graft failure (dialysis/re-transplantation)	1.Death-censored graft failure (dialysis/re-transplantation) 2.Death (any cause)	Death censored graft failure (defined by return to dialysis)
	Competing event	1. No 2. Death with functioning graft 3. Dialysis/re-transplantation	1.Dialysis/re-transplantation 2.Death with functioning graft	Death with functioning graft	1.Death with functioning graft 2 Dialysis/re-transplantation	Death with functioning graft and re-transplantation
	Accounted for competing risk	Not clearly specified	Not clearly specified	Not clearly specified	Not clearly specified	Not clearly specified
Horizon of prediction		5 years	2 years	1, 2, 3, 5, 7, and 10 years	1, 3, 5, 7 years	8 years
DEVELOPMENT COHORT	Development cohort type	Registry: United States Renal Data System	Single center Cohort	Registry: USRDS	Registry: United States Renal Data System	Single center Cohort (Necker Hospital)
	Years of transplantation	2000 - 2006	1999 - 2002	1990 - 1999	1995 - 2002	2004 - 2010
	Sample size	59091 for model at transplantation 57603 for model at 7 days 43743 for model at 1 year	392	92844	5389	939
	Statistical model used	Cox regression model	Logistic regression model	Tree based model analysis	Logistic regression model Cox regression model Artificial Neural Networks (ANNs)	Cox regression model
	Predictors measured at transplantation	Model at transplantation: donor age, recipient race, first versus subsequent transplant, prior years on renal replacement therapy, recipient age, primary cause of CKD, hepatitis C virus antibody status, donor history of hypertension, recipient primary insurance coverage, donor cause of death, and total HLA antigen mismatches.	Recipient age, living donor , peripheral vascular dis, oral anticoagulation, smoking, haemodialysis, serum calcium, serum albumin	Recipient variable: recipient race, gender, age, height, weight, recipient having a transplant before the current one, total number of transplants, the time waiting list, predominant RRT, percent time on Peritoneal Dialysis, number of RRT used, specific combination of RRT, recipient comorbidity score, cardiovascular disease, unstable angina, diabetes, hypertension, presence of hepatitis B core antibodies, hepatitis C antibodies, peak and most recent level of panel reactive antibodies, and pay for medical services.	Recipient variables: age; gender; race; height; weight; cause of ESRD; history of hypertension, diabetes, or cardiovascular disease; duration between date of current transplantation and failure date of the previous transplantation, dialysis modality, primary source of pay for treatment.	1.Model with Antibody-mediated rejection (ABMR) Molecular Score : donor age, the humoral histologic score, and the ABMR molecular score
	Predictors used in the final model	Model at 7 days: eGFR at hospital discharge, donor age, primary cause of CKD, recipient race, recipient age, and years on renal replacement therapy. Model at 1 year: eGFR at 1 year post-transplant, recipient race, hospitalization during year 1 post-transplant, primary cause of CKD, recipient age, and recipient primary insurance coverage.		Donor variables: donor race, gender, age, height, weight, donor type. Transplant procedure variables: cold ischemia time and number of matched HLA antigens, using MMF in the immunosuppressive regimen	Donor variables: donor type, age, gender, race, height, weight, cause of death. Transplantation parameters: number of matched HLA antigens, cold storage time, procedure type.	2.Model with endothelial DSA-selective transcripts: donor age, the humoral histologic score, and the endothelial DSA-selective transcripts
	Predictors measured after transplantation	Characteristics available at 7 days then at 1 years				eGFR at time of rejection , Antibody-mediated rejection (ABMR) within first years after transplantation
	Method used to account for post-transplant predictors	Only patients alive with transplant function at 7 days, then at 12 months were selected to develop the model				Not clearly specified
	Prediction tools presentation form	Original model	Score	Original model	Original model	Original model
	Proposed thresholds for clinical use / Proposed software or online-calculator		Threshold defined at quartiles points. Three risk groups: - Lower risk group: score between 0 – 2.5 - Intermediate risk group: score between 3 – 8.5 - High risk group: score ≥9			
PER FOR VALIDATION COHORT	Internal validation	Yes	No	Yes	Yes	Yes
	Internal validation type	Split-sample 70%, 30%		Split-sample 67%, 33%	10 fold cross-validation	Bootstrapping
	External validation	Yes	Yes	No	No	Yes
	External validation type	Spatial	Temporal			Spatial
	External validation cohort (sample size)		Transplanted in January 2007 – December 2007 (n=157)			321 transplanted at Saint-Louis
Overall						

Discrimination	AUC for graft failure by internal validation: - 0.649 for model at transplantation - 0.674 for model at 7 days - 0.716 for model at 1 years		AUC - 0.63 at 1 year - 0.64 at 2 years 0.71 at 3 years 0.82 at 5 years 0.90 at 10 years	AUC (for 1, 2, 3, 5, 7 years): 1. Death-censored graft failure Logistic: 0.71 - 0.72 - 0.75 - 0.81 Cox: 0.72 - 0.73 - 0.74 - 0.80 ANNs: 0.61 - 0.68 - 0.73 - 0.82 2.Death (any cause) Logistic: 0.71 - 0.73 - 0.77 - 0.81 Cox: 0.72 - 0.73 - 0.76 - 0.80 ANNs: 0.59 - 0.66 - 0.75 - 0.82		AUC for model with ABMR Molecular Score: 0.81 for internal validation
	AUC for graft failure by external validation: - 0.61-0.70 for model at transplantation - 0.60-0.72 for model at 7 days - 0.64-0.78 for model at 1 years					
Calibration	Slope of prognostic index for graft failure - 1.04 for model at transplantation - 0.99 for model at 7 days - 0.96 for model at 1 year		Agreement between predicted and observed graft survival (correlation r2=0.94, r2=0.98, r2=0.99, r2=0.93, and r2=0.98) for 1,2,3,5 and 10 years respectively		Hosmer-Lemeshow for all analysis was Chisq > 10 (P>0.05)	
Reclassification					For the model with ABMR Molecular Score: NRI= 1.01, IDI= 0.16	
Other performance indicator						
Other studies used for comparison						
Number of citation in google scholar at 31/04/16	21	0	17	24	29	

Author and years (journal)		Lowsky et al. 2012 (47) (Statistics in Medicine)	MacHnicki et al. 2009 (48) (American Journal of Transplantation)	Moore et al. 2011 (44) (American Journal of Kidney Diseases)	Moore et al. 2011(61) (Experimental and Clinical Transplantation)	Munivenkatappa et al. 2008 (49) (American Journal of Transplantation)
Type of study		Development and validation of a new model	Development and validation of a new model	Development and validation of a new model	Validation with model updating	Development and validation of a new model
Donor type		Living and deceased donor	Deceased donor	Living and deceased donor	Living and deceased donor	Deceased donor
Recipient age range		Unspecified	≥18 years	≥ 18 years	≥ 16 years	Adult (Age range unspecified)
OUTCOMES	Studied event	Graft failure (dialysis/re-transplantation/death)	1.Graft failure (dialysis/re-transplantation/death) 2.Death with functional graft 3.Death (including death after return to dialysis)	1.Graft failure (dialysis/re-transplantation/death) 2.Death censored graft failure	Death with functional graft	Death censored graft failure (return to dialysis, or a decline in renal function to a glomerular filtration rate (GFR), of <20 mL/min)
	Competing event	No	1.No 2. Dialysis/re-transplantation 3. Dialysis/re-transplantation	1.No 2.Death with functioning graft	Dialysis/re-transplantation	Death with functioning graft and re-transplantation
	Accounted for competing risk		Not clearly specified	Not clearly specified	Not clearly specified	Not clearly specified
	Horizon of prediction	5 years	9 years	7 years	3 and 5 years	5 years
DEVELOPMENT COHORT	Development cohort type	Registry: United States Renal Data System	Registry: United States Renal Data System	Multi-centers LOTESS (novates multicentre study)	LOTESS database (novates multicentre study)	Single center (Maryland Medical Center)
	Years of transplantation	1996-1999	1995 - 2002	1995 - 1998	1995 - 1998	1999 - 2005
	Sample size	30051	25270	2763	2033	371
	Statistical model used	Mahalanobis K-nearest neighbor (MKNN) Random survival forests (RSF) Cox regression model	Cox regression model	Cox regression model	Cox regression model	Cox regression model
	Predictors measured at transplantation	Age, pre-event dialysis time , Blood type, Peak panel-reactive antibody, Body mass index, Cause of end-stage renal disease, Pre-transplant blood transfusion, Previous transplant, Donor type, Expanded criteria donor, and Year of transplant	Model 1— Organ Procurement Transplant Network (OPTN) variables (OPTN only) Model 2— Clinical Classifications Software (CCS) disease categories + OPTN variables (OPTN+CCS) Model 3—Charlson comorbidities plus OPTN variables Model 4—Elixhauser comorbidities plus OPTN	For death censored graft failure: age, Race, eGFR For graft failure: recipient age, Sex, eGFR	Model 1: Recipient Risk Score (RRS, Baskin-Bey et al. 2007 (25) Model 2: Age-adjusted Charlson Comorbidity Index Model 3: Foley Score Model 4: Wright-Khan Index Model 5: Charlson Comorbidity Index Model 6: Modified End-Stage Renal Diseases Charlson Comorbidity Index Model 7: Davies Index Models 2-7 included: - Donor: age, sex, race, and source. - Recipient: age, sex, race, duration of dialysis, dialysis modality, prior transplant, induction, use of, and body mass index	Arteriolar hyalinosis, periglomerular fibrosis, presence of scar , glomerulosclerosis, wall-to-lumen ratio
	Predictors used in the final model		All models were adjusted for donor, recipient and transplant characteristics			
	Predictors measured after transplantation			Data were collected prospectively in a cohort: eGFR at data collection, serum urea nitrogen, serum albumin at 6 month before data collection, percentages decrease in eGFR during preceding 6 months before data collection and acute rejection Only patients alive with transplant function at 12 months was selected to develop the model		
	Method used to account for post-transplant predictors					
	Prediction tools presentation form	Original model	Original model	Score	Existing model (Model 1 - RRS) Updated model (Models 2-7)	Score: Maryland Aggregate Pathology Index (MAPI)
	Proposed thresholds for clinical use / Proposed software or online-calculator			No threshold, but proposed a simple score derivation (range 0 to 181 for death censored and 0 to 179 for graft failure)		Threshold defined after sensibility/specificity analysis. Three groups (score range: 0-15): - Lower risk group: MAPI from 0 to 7 - Intermediate risk group: MAPI from 8 to 11 - Higher risk group: MAPI from 12 to 15
VALIDATION COHORT	Internal validation	Yes	Yes	Yes	Yes	Yes
	Internal validation type	Split sample 80% -20% (7512)	Cross-validation Bootstrapping	Random split 70% - 30%	Unspecified	Split sample 67%-33%
	External validation	Yes	No	Yes	Yes	No
	External validation type	Temporal		Spatial	Fully	
	External validation cohort (sample size)	13525 transplanted at 1999		Transplanted in Elisabeth hospital, Birmingham in 1996 – 2006 (n=731)		
PERFORMANCES	Overall		AUC	AUC	AUC	
	Discrimination		Models Graft loss DWF Death	Internal 0.73 for graft failure 0.87 for death censored graft failure	Models For model original Include age for updated model	AUC 0.70 for development cohort 0.74 for validation cohort

Author and years (journal)		Rao et al 2009 (45) (Transplantation)	Schnitzler et al. 2012(50) (Transplantation)	Schold et al. 2005 (51) (American Journal of Transplantation)	Shabir et al. 2014 (53) (American Journal of Kidney Diseases)	Szabo et al. 2011(58) (Journal of Nephrology)
Type of study		Development and validation of a new model	Development and validation of a new model	Development and validation of a new model	Development and validation of a new model	Validation without any model updating
Donor type		Deceased donor	Living and deceased donor	Deceased donors	Living and deceased donor	Deceased donor
Recipient age range		≥ 18 years	Unspecified	Adult (Age range unspecified)	≥ 18 years	≥ 18 years
OUTCOMES	Studied event	Graft failure (dialysis/re-transplantation/death)	Graft failure (dialysis/re-transplantation/death)	1.Graft failure (dialysis/re-transplantation/death) 2.Death censored graft failure	1.Death censored graft failure 2.Graft failure (dialysis/re-transplantation/death)	Death (unspecified)
	Competing event	No	No	1.No 2.Death with functioning graft	1.Death with functioning graft 2.No	Dialysis/re-transplantation
	Accounted for competing risk			Not clearly specified	Not clearly specified	Not clearly specified
Time horizon of prediction		Long-term (unspecified)	to 9 years	1 and 5 years	5 years	5 years
DEVELOPMENT COHORT	Development cohort type	Registry: SRTR/OPTN	Registry: United States Renal Data System	Registry: Scientific Registry of Transplant Recipients	Multi-center cohort	Single center
	Years of Transplantation	1995 - 2005	1995 - 2004	1996-2002	1999 - 2006	1991 - 2004
	Sample size	69440	87575	45850	651	
	Statistical model used	Cox regression model	Cox regression model	Cox regression model	Cox regression model	Cox regression model
	Predictors measured at transplantation	Donor age, race, history of hypertension, history of diabetes, serum creatinine, cerebrovascular cause of death, height, weight, donation after cardiac death, hepatitis C virus status, human leukocyte antigen-B and DR mismatch, cold ischemia time, and double or in bloc transplant.	1. Standard criteria donor (SCD): Recipient variables (age, sex, pre-transplant dialysis, body mass index, Cause of ESRD glomerulonephritis, diabetes, previous transplant, peak panel-reactive antibody, HLA mismatches, race, year of transplant). Donor variable (hypertension, cytomegalovirus serology, race, cause of death), 2. Expanded criteria donor (ECD): Recipient variables (age, sex, Body mass index, cause of ESRD glomerulonephritis, diabetes, previous transplant, peak panel-reactive Antibody, HLA mismatches, race, year of transplant), donor race,	Donor/Recipient CMV match, donor race, donor age, cause of Death, HLA, cold ischemia time, donor history of hypertension, donor history of diabetes	Sex, Race	Age, pre-transplant cardiovascular disease, serum creatinine at discharge, left ventricular hypertrophy, vascular calcification, diabetes before renal transplantation, time on dialysis > 48 months, acute tubular necrosis
	Predictors used in the final model					
	Predictors measured after transplantation		eGFR at 1 years, acute rejection within 1 year		Recipient age, acute rejection, transplant function (eGFR), serum albumin level and urine albumin-creatinine ratio at 6 month and 12 month	
	Method used to account for post-transplant predictors		Only patients alive with transplant function at 12 months was selected to develop the model		Only patients alive with transplant function at 12 months was selected to develop the model	
	Prediction tools presentation form	Score (KDRI)	Original model	Score	Original model	Existing score developed on Spanish population (Hernandez et al, 2005)
	Proposed thresholds for clinical use / Proposed software or online-calculator	Threshold defined at quintiles points. Five risk groups: - Group1: KDRI from 0.45 to 0.79 - Group2: KDRI from 0.79 to 0.96 - Group3: KDRI from 0.96 to 1.15 - Group4: KDRI from 1.15 to 1.45 - Group5: KDRI > 1.45		Threshold defined by cluster analysis. Five risk classes: - Class I : score from 0 to 0.234 - Class II : score from 0.234 to 0.524 - Class III : score from 0.524 to 0.853 - Class IV : score from 0.853 to1.17 - Class V : score >1.17		
VALIDATION COHORT	Remapping of the KDRI (KDPI): https://optn.transplant.hrsa.gov/resources/allocation-calculators/kdpi-calculator/					
	Internal validation	Yes	Yes	Yes	No	No
	Internal validation type	5 fold cross-validation (one half)	Unspecified	Split random 67% - 33%		
	External validation	No	Yes	Yes	Yes	Yes
	External validation type		Spatial	Spatial	Spatial	Fully
PERFORMANCES	External validation cohort (sample size)		From clinical trial: BENEFIT (n=345) and BENEFIT-EXT (n=244)	Naïve cohort (unspecified) (n=2230)	Tours in France (n=736), Leeds in UK (n=787) and Halifax in Canada (n=475)	339 transplanted in Hungaria
	Overall			Coefficient of determination (R2)=96.0%		
	Discrimination	AUC 0.62			AUC 1.For death censored graft failure: Tours: 0.89; Leeds: 0.78; Halifax: 0.90 2.for graft failure Tours: 0.80; Leeds: 0.76; Halifax: 0.81	AUC 0.65 Sensitivity / specificity (%) Cut-off 84 27 0.12 56 66 0.25 40 82 0.46 24 90 0.70 9 98 0.93
	Calibration		Internal cohort: Agreement between predicted and observed graft survival at 5 years. (r2 =99.9%) by plot External validation: Agreement between predicted and observed graft survival at each years	Agreement between each score level and observed graft survival (survival plot by score class)	Hosmer-Lemeshow test (p-value) 1. for death-censored graft failure Tours: 0.4; Leeds: 0.1; Halifax: 0.3 2. for graft failure Tour: 0.09; Leeds: 0.03; Halifax: 0.6	Hosmer-Lemeshow test (p-value<0.001)

Author and years (journal)		Tang et al. 2011 (63) (American Society for Artificial Internal Organ Journal)						Tang et al. 2011 (55) (American Society for Artificial Internal Organ Journal)				Tiong et al. 2009 (28) (Journal of Urology)				Watson et al. 2012 (52) (Transplantation)			
Type of study		Development and validation of a new model						Development and validation of a new model				Development and validation of a new model				Development and validation of a new model			
Donor type		Living donor						Donor type unspecified				Living donor				Deceased donor			
Recipient age range		Adult (Age range unspecified)						≥ 18 years				Adult (Age range unspecified)				≥ 18 years			
OUTCOMES	Studied event	Death censored graft failure						Graft failure (dialysis/re-transplantation/death)				Graft failure (Dialysis/death/re-transplantation)				Graft failure (dialysis/re-transplantation/death)			
	Competing event	Death with functioning graft						No				No				No			
	Accounted for competing risk	Not clearly specified																	
Time horizon of prediction		1,3, 5,7,10 years						3				5				0- 9 years (overall) and 90 days, 3 month to 3 years, over 3 years			
DEVELOPMENT COHORT	Development cohort type	Registry: United States Renal Data System						Registry: United States Renal Data System				Registry: United Network for Organ Sharing				Registry: UK Transplant Registry			
	Years of transplantation	1990 - 2004						1985 - 2002				2000 - 2003				2000 - 2007			
	Sample size	92844 for 1 year 73672 for 3 year 58005 for 5 year						46791 for 7 years 35279 for 10 years 4754				20085				4570			
	Statistical model used	Tree based model						Classification trees, Logistic regression, and Artificial neural networks				Cox regression model				Cox regression model			
	Predictors used in the final model	Predictors measured at transplantation Recipient variables: age, gender, race, height, weight, history of hypertension, diabetes, unstable angina, cardiovascular or peripheral vascular diseases, peak panel reactive antibody, Recipient serum albumin , Dialysis modality predominant dialysis modality, total time on waiting list, immediately before, Transplantation, Predominate ESRD modality, Recipient dialysis modality used before transplantation for at least 60 days, primary source of payment for treatment, if recipient has US citizenship, a comorbidity score, peak PRA level, and most recent PRA level. Donor variables: donor type, age, gender, race, height, and weight. Transplantation parameters: the degree of human leukocyte antigen (HLA) match, donor cold storage time, history of previous transplantation, total number of transplantations, and if the recipient used mycophenolate in immunosuppressive therapy						Model with 38 variables: Recipient age, Recipient race, Recipient ethnicity, Recipient height, Recipient weight, Duration of ESRD, Charlson index, Recipient history of cardiovascular diseases, Recipient history of diabetes, Recipient history of hypertension, Number of pre-transplant blood transfusion, Number of previous pregnancies, Dialysis was required during first post-transplant week, Number of matched human leukocyte antigens, Number of previous transplantations, Maintenance regimen includes, Induction regimen includes, Donor cold ischemia time, Donor age, Donor race, Donor gender, Donor height, Donor weight, Donor type, Donor cause of death, Donor history of hypertension				Recipient : age, gendre, race, BMI, HLA, Donor age, donor gender, donor race, donor BMI, donor serum creatinine, nephrectomy type, cause of graft failure, depleting antibodies, IL2 receptor, azathioprine, mycophenolatemofetil, rapamycin, calcineurin inhibitor				Recipient age, donor age, history of hypertension, donor weight, days in hospital, adrenaline			
	Predictors measured after transplantation											Delayed graft function, any treated rejection episode in 6 months and eGFR at 6-month							
	Method used to account for post-transplant predictors											Not clearly specified							
	Prediction tools presentation form	Original model						Original model				Nomogram (two proposed)				Score: United Kingdom Kidney Donor Risk Index (UKKDRI) for overall Threshold defined at quartile points. Four groups: - Group 1: UKKDRI <0.87 - Group 2: UKKDRI from 0.88 to 1.02 - Group 3 : UKKDRI from 1.03 to 1.34 - Group 4: UKKDRI ≥ 1.35			
	Proposed thresholds for clinical use / Proposed software or online-calculator											No threshold but proposed a simple score derivation (range 0 to 100)							
	VALIDATION COHORT	Internal validation	Yes						Yes				Yes				Yes		
Internal validation type		Split random 66% - 33%						10 fold cross validation				10 fold cross-validation				Split-sample 60% - 40%			
External validation		Yes						No				No				No			
External validation type		Spatial																	
External validation cohort (sample size)		UUHSC (University of Utah health science center) (854)																	
PERFORMANCES	Overall	AUC						AUC											
	Discrimination	years	1	3	5	7	10	Model with 38 variables	Model with 6 variables	AUC	0.71 for pre-transplant nomogram 0.78 for 6 months post-transplantation			AUC	0.62 for overall				
		External validation	0.54	0.58	0.58	0.61	0.70	Logistic regression	0.74	0.73									
							Artificial neural networks	0.71	0.73										
	Calibration											Agreement of predicted and observed graft survival (calibration plot)							
	Reclassification																		
Other performance																			
Other studies used for comparison																			
Number of citation in google scholar at 31/04/16		4						5				25				30			