Continuous indices to assess the phenotypic spectrum of kidney transplant rejection

Vaulet, T.; Koshy, P.; Wellekens, K.; Aubert, O.; Bottomley, C.; Callemeyn, J.; ... ; Naesens, M.

nature communications

Article

https://doi.org/10.1038/s41467-025-65153-9

Continuous indices to assess the phenotypic spectrum of kidney transplant rejection

Received: 15 January 2025

Accepted: 3 October 2025

Published online: 26 November 2025

Check for updates

Thibaut Vaulet[1], Priyanka Koshy[1,2,3], Karolien Wellekens[1,3], Olivier Aubert[4,5], Charlotte Bottomley[6], Jasper Callemeyn[1,3], Evert Cleenders[1,7], Maarten Coemans[1,7], Lynn Cornell[8], Aiko P. J. de Vries[9,10], Gillian Divard[11], Marie-Paule Emonds[12], Sandrine Florquin[13,14], Mark Haas[15], Philip F. Halloran[16], Jesper Kers[9,10,17,18,19], Dirk Kuypers[1,3], Thangamani Muthukumar[20], Angelica Pagliazzi[1,3], Steven Salvatore[21], Olivier Thaunat[22,23], Surya V. Seshan[21], Elisabet Van Loon[1,3], Thomas Vanhoutte[1,3], Georg A. Böhmig[24], Friedrich A. von Samson-Himmelstjerna[25], Michelle Willicombe[6], Aravind Cherukuri[26], Alexandre Loupy[11], Candice Roufosse[6] & Maarten Naesens[1,3] ✉

The Banff classification for kidney transplant pathology dichotomizes the rejection continuum into distinct diagnostic categories, introducing artificial cutoff points and threshold effects. To better reflect the underlying disease spectrum, in this cohort study of 19,500 biopsies from 8873 patients across 10 centers worldwide, we developed two indices for quantifying antibody-mediated rejection/microvascular inflammation and T-cell-mediated rejection/tubulointerstitial inflammation from histological lesion scores and calculated indices for overall activity and chronicity. These indices demonstrate excellent discrimination for the main diagnostic categories of rejection (AUCs from 0.95 to 0.99), with consistent performance across derivation and validation datasets. These indices strictly confine intermediate phenotypes to low index values and are associated to graft failure even within the diagnostic categories, thus reflecting the underlying rejection continuum. In this work, we demonstrate that four continuous indices provide implementable and interpretable global evaluation of kidney transplant histology that align with the continuous nature of the rejection process regardless of the underlying disease cause.

Like many pathological processes, kidney transplant rejection is a gradual and continuous process that is more or less active depending on donor–recipient genetic disparity, highly heterogeneous immunological responses to this genetic mismatch, and the efficacy of the immunosuppression prescribed to the recipients[1,2]. The Banff classification for kidney transplant pathology is currently used for the diagnosis of kidney transplant rejection. This classification is an expert-based consensus framework for the diagnosis of allograft rejection and its subtypes, primarily antibody-mediated rejection (AMR) and T-cell-mediated rejection (TCMR), which are characterized histologically by microvascular and tubulointerstitial inflammation, respectively[3]. As such, this classification is essentially a dichotomization of the continuous rejection process into distinct diagnostic categories, which can be translated into complex decision trees such as the Banff Automation System[4].

A full list of affiliations appears at the end of the paper. ✉e-mail: maarten.naesens@kuleuven.be

This classification introduces artificial cutoff points that do not adequately reflect disease biology or the disease spectrum. The categorization of diseases into nonoverlapping classes aggregates highly diverse cases within singular diagnostic entities. Conversely, cases with very similar features but situated on opposite sides of a (consensus-based) diagnostic threshold are classified differently, while they are biologically equivalent[5]. To partly mitigate this issue, intermediate categories are adopted in many classifications and clinical guidelines to increase nuance.

Specifically, for kidney transplantation, intermediate diagnostic categories such as "probable AMR" and "borderline for TCMR", as well as acute/active, chronic-active, and chronic subcategories of AMR and TCMR, were introduced in the Banff classification to reflect the temporal stage of disease. The clinical significance of such intermediate and temporal subcategories remains ambiguous and leads to recurring discussions on the choice of the boundaries between the presence/absence of disease and different disease stages[3,6].

With a more continuous quantification of the underlying disease process, which is closer to actual pathological observations and to biological reality, the problems created by the artificial boundaries in typical disease classification models could be overcome[7–9]. In that context, we recently developed activity[10] and chronicity indices[11] from Banff lesion scores of kidney transplant biopsies, representing the overall amount of inflammation and of chronicity observed in a kidney transplant biopsy, respectively. Others have developed mathematical models for quantifying and distinguishing AMR and TCMR[12]. As acknowledged in recent Banff consensus meetings, these scores hold promise for stratifying biopsies by disease severity, independent of the histological diagnosis[3,6]. However, experts in the field highlighted that, to date, no study has demonstrated the added value of continuous indices for the evaluation of kidney transplant pathology in comparison with the current rule-based classification[6].

Therefore, we hypothesized that continuous indices derived from routinely assessed histological lesion scores would enable the quantification of the global spectrum of kidney transplant rejection, primarily distinguishing microvascular inflammation (MVI) from tubulointerstitial inflammation (TI) patterns. Second, we hypothesized that the Banff diagnostic subcategories of rejection, such as active/acute, chronic active, and chronic, could be replaced by more continuous activity and chronicity indices calculated from the histological lesion scores, adding more detailed evaluation of disease stage and severity.

To test these hypotheses, we developed two easily calculable continuous indices representing the rejection spectra using a large derivation cohort. We validated these indices with two additional large multicentric validation cohorts by analyzing (1) the relationship of these indices with the Banff diagnostic categories and (2) the ability of these indices to discriminate different clinical outcomes *within* these categories. Finally, we assessed and validated (3) the associations of the continuous indices with graft failure rates, independent of the Banff (sub)categories. These four continuous indices, which are derived from standard lesion scores, elegantly capture the full histological spectrum and severity of kidney transplant rejection more comprehensively than the Banff (sub)categories do.

Results

Patient and biopsy characteristics

The patient characteristics are reported in Table 1. After patients without posttransplantation biopsies ($N = 77$) or biopsies of inadequate quality ($N = 119$) were excluded, a total of 6272 post-transplantation biopsies from 1814 patients were included in the derivation cohort. Validation was performed on two external cohorts: the European validation cohort consisted of 11,043 biopsies from 5898 transplants, and the US validation cohort consisted of 2185 biopsies from 1161 transplants.

Individual Banff lesion scores were assessed by local pathologists at each participating center. The final diagnosis, based on the integration of these scores and the human leukocyte antigen (HLA)-donor-specific antibody (DSA) status, was assigned by two expert reviewers (K.W. and M.N.) using the Banff 2022 classification criteria[3] into the following, potentially coexisting, diagnostic categories: AMR (containing active, chronic active, and chronic AMR); probable AMR (hereafter referred to as "Probable AMR"); DSA-negative, C4d-negative microvascular inflammation (hereafter referred to as "$MVI_{DSA-/C4d-}$"); borderline/suspicious for TCMR (hereafter referred to as "Borderline TCMR"); acute TCMR; and polyomavirus-associated nephropathy (hereafter referred to as PVAN); and normal or other changes as defined by Banff category 6 (hereafter referred to as "No rejection"). Following the reappraisal in recent studies of the role of the v-lesion in TCMR, biopsies with $v > 0$, t0-3i0 or t0i0-3, not meeting Banff AMR-MVI criteria, were considered as isolated v, and not as TCMR[13,14]. In this study, "mixed rejection" strictly refers to the concomitant presence of AMR and TCMR; "(Borderline) TCMR" refers to Borderline TCMR or TCMR.

Most transplants in the derivation cohort were from deceased donors (92.1%). In total, 273 (15.0%) recipients underwent retransplantations, and 151 (8.3%) recipients had pretransplantation HLA-DSAs. Among all the biopsies, 1,666 (26.6%) were performed for clinical indications, whereas the other 4,561 (72.7%) biopsies were performed per protocol. TCMR (7.2%) and Borderline TCMR (6.9%) were the most common rejection phenotypes. AMR occurred in 5.5%, $MVI_{DSA-/C4d-}$ in 4.4%, Probable AMR in 1.6%, isolated v in 2.4% and mixed rejection in 1.5% of the cases. A total of 4610 (73.5%) biopsies did not meet the criteria for the diagnostic categories and were classified as No rejection biopsies. The details of the diagnostic categories in the validation cohorts are reported in Table 1.

During the period of observation, 255 (14.1%) kidney allografts failed in the derivation cohort. In the European and US validation cohorts, 1009 (17.1%) and 171 (14.7%) kidney allografts failed, respectively.

Distribution of the continuous indices in relation to the Banff diagnostic (sub)categories

Given that AMR and TCMR constitute continuous disease spectra, AMR and TCMR can be represented by latent continuous variables instead of being regarded as strictly binary phenomena (present vs. absent). As such, we developed two indices based on a latent variable approach, where the unobserved latent component is inferred from a set of relevant, observed Banff lesions, with the binary diagnostic categories (AMR vs. no AMR, and TCMR vs. no TCMR) serving as the dependent variable to guide the inference of the latent variable (Online Methods). The activity and chronicity indices were adapted from previously published models[10,11]. The continuous indices included in the analyses, shown in Table 2, were calculated for all biopsies. The distribution of these indices in the derivation cohort is shown in Fig. 1.

The index trained on AMR vs. no AMR gradually increased from No rejection to Probable AMR to AMR. As $MVI_{DSA-/C4d-}$ cases also presented high indices, we referred to this as the AMR/MVI index. The index trained on TCMR vs. no TCMR demonstrated a gradual increase from No rejection to Borderline TCMR to TCMR. In PVAN cases, this index was more broadly distributed and was significantly higher than in No rejection cases (Student t-test: $p < 0.001$), because the tubulointerstitial inflammation (TI) in a subset of PVAN biopsies was similar to those in TCMR biopsies. This second index was designated as the TCMR/TI index.

The Banff diagnostic categories were logically distributed in a two-dimensional space on the basis of a combination of the AMR/MVI index and the TCMR/TI index (Fig. 2). The AMR and $MVI_{DSA-/C4d-}$ cases demonstrated significant overlap of AMR/MVI indices. The highest AMR/MVI indices were observed for AMR cases, but this is inherent to

https://doi.org/10.1038/s41467-025-65153-9

Table 1 | Demographic, clinical, and histological characteristics of the patients and biopsies included in the derivation and validation cohorts

Cohort characteristics	Derivation cohort, *n* (%) *n* = 1814	European validation cohort, *n* (%) *n* = 5898	*P* value[a]	US validation cohort, *n* (%) *n* = 1161	*P* value[a]
Donor demographics					
Donor type			<0.001		<0.001
Donation after brain death	1315 (72.5)	2900 (50)		526 (45.6)	
Donation after cardiac death	355 (19.6)	1045 (18)		169 (14.6)	
Living donation	144 (7.9)	1862 (32)		460 (39.8)	
Missing	0 (0)	91 (1.5)		6 (0.5)	
Age (years), mean ± SD	49.0 ± 14.4	52.2 ± 15.3	<0.001	42.0 ± 13.9	<0.001
Missing	0 (0)	61 (1)		16 (1.1)	
Male	985 (54.3)	2931 (50.5)	0.008	599 (51.6)	0.19
Missing	0 (0)	97 (1.6)		0 (0)	
Recipient demographics					
Age (years), mean ± SD	54.5 ± 12.8	50.7 ± 14.3	<0.001	51.2 ± 14.7	<0.001
Missing	0 (0)	0 (0)		1 (0.1)	
Male	1152 (63.5)	3759 (63.7)	0.88	691 (59.5)	0.03
Ethnicity[b]			<0.001		<0.001
Asian	14 (0.8)	618 (31)		40 (3.4)	
Black	54 (3)	283 (14.2)		268 (23.1)	
Others	1 (0.1)	180 (9)		49 (4.3)	
White	1745 (96.2)	912 (45.8)		804 (69.3)	
Missing	0 (0)	3905 (66.2)		0 (0)	
BMI (kg/m^2), mean ± SD[c]	25.6 ± 4.5	25.5 ± 4.8	0.68	27.4 ± 6.1	<0.001
Missing	0 (0)	892 (15.1)		951 (81.9)	
Pretransplantation donor-specific HLA antibodies	151 (8.3)	NA		NA	
Retransplantation[d]	273 (15)	916 (18.8)		277 (23.9)	
Missing	0 (0)	1028 (17.4)		0 (0)	
Cold ischemia time (hours), mean ± SD	12.7 ± 6.0	13.0 ± 8.1	0.19	8.7 ± 7.5	<0.001
Missing	1 (0.1)	929 (15.8)		61 (5.3)	
Total number of HLA A/B/DR mismatches, mean ± SD	2.7 ± 1.3	3.3 ± 1.5	<0.001	3.8 ± 1.6	<0.001
Missing	0 (0)	230 (3.9)		9 (0.8)	
N Graft failure	255 (14.1)	1009 (17.1)	0.002	171 (14.7)	0.578
Median follow-up time (IQR) (year)	6.6 (3.7–10.6)	5.5 (3.1–8.8)		6.9 (5.3–8.7)	
Biopsy characteristics, *n*	6272	11043		2185	
Days since transplantation, median (interquartile range)	362.0 (89.0–732.0)	222.0 (83.0–630.0)	<0.001	168.0 (95.0–380.0)	<0.001
Indication biopsies	1666 (26.6)	6143 (55.8)	<0.001	854 (39.1)	<0.001
Missing	0 (0)	37 (0.3)		0 (0)	
Banff 22 categories					
No rejection	4610 (73.5)	6738 (61.0)		879 (40.2)	
TCMR	449 (7.2)	1033 (9.4)		621 (28.4)	
Borderline TCMR	431 (6.9)	1027 (9.3)		514 (23.5)	
AMR	343 (5.5)	1175 (10.6)		197 (9.0)	
$MVI_{DSA-/C4d-}$	278 (4.4)	575 (5.2)		101 (4.6)	
PVAN	180 (2.9)	262 (2.4)		76 (3.5)	
Isolated v	152 (2.4)	53 (0.5)		1 (0.0)	
Probable AMR	100 (1.6)	210 (1.9)		17 (0.8)	
Mixed rejection	91 (1.5)	239 (2.2)		119 (5.4)	
Unclassifiable[e]	0 (0.0)	705 (6.4)		17 (0.8)	

AMR antibody-mediated rejection, *BMI* body mass index, *IQR* interquartile range, $MVI_{DSA-/C4d-}$ microvascular inflammation, *PVAN* polyomavirus nephropathy, *SD* standard deviation, *TCMR* T-cell-mediated rejection.

[a]In comparison with the derivation cohort, two-sided Student's *t* test and chi-square test were used, where appropriate.

[b]Ethnicity not available for centers in Amsterdam, Leiden, Lyon, Paris and Vienna.

[c]BMI not available for university hospital Schleswig-Holstein and university of Pittsburgh cohorts.

[d]Retransplantation information not included in the university hospital Schleswig–Holstein cohort.

[e]Biopsies unclassifiable due to missing key Banff lesion scores for definitive diagnosis.

Percentages and statistical tests are derived from complete values.

https://doi.org/10.1038/s41467-025-65153-9

Table 2 | Overview of the continuous indices and their intended use

Index	Calculation	Range	Intended use
AMR/MVI index	0.938 × g + 0.762 × ptc + 0.728 × cg + 2.716 × C4d#	0–10	- Discrimination within the phenotypic spectrum Probable AMR – $MVI_{DSA-/C4d-}$ – AMR - Estimation of AMR/MVI intensity - Evaluation of underlying cause is warranted for interpretation of a high AMR index (DSA and/or C4d positive AMR vs. $MVI_{DSA-/C4d-}$)
TCMR/TI index	0.970 × i + 0.623 × t + 1.540 × v + 0.195 × ct	0–10	- Discrimination within the phenotypic spectrum Borderline TCMR – TCMR – PVAN - Estimation of TCMR intensity - Evaluation of underlying cause is warranted for interpretation of a high TCMR index (TCMR vs. PVAN)
Activity index	t + i + v + g + ptc + 2 × C4d#	0–17	- Global estimation of disease activity/inflammation level, irrespective of the diagnostic category - Independent association with graft failure rates, irrespective of the diagnostic (sub) category
Chronicity index	ci + ct + cv + 2 × cg same as Haas et al.	0–15	- Global estimation of biopsy chronicity level, irrespective of the diagnostic category - Independent association with graft failure rates, irrespective of the diagnostic (sub) category

AMR antibody-mediated rejection, *DSA* donor-specific antibody, $MVI_{DSA-/C4d-}$ microvascular inflammation, DSA negative and C4d negative, *PVAN* polyomavirus nephropathy, *TCMR* T-cell-mediated rejection, *TI* tubulointerstitial inflammation.
i = interstitial inflammation (range 0–3).
t =tubulitis (range 0–3).
v = intimal arteritis (range 0–3).
g = glomerulitis (range 0–3).
ptc = peritubular capillaritis (range 0–3).
ci = interstitial fibrosis (range 0–3).
ct = tubular atrophy (range 0–3).
cv = vascular fibrous intimal thickening (range 0–3).
cg = glomerular basement membrane double contours (range 0–3).
C4d# = C4d deposition in peritubular capillaries (score 0 or 1, as negative/positive according to the Banff definitions).

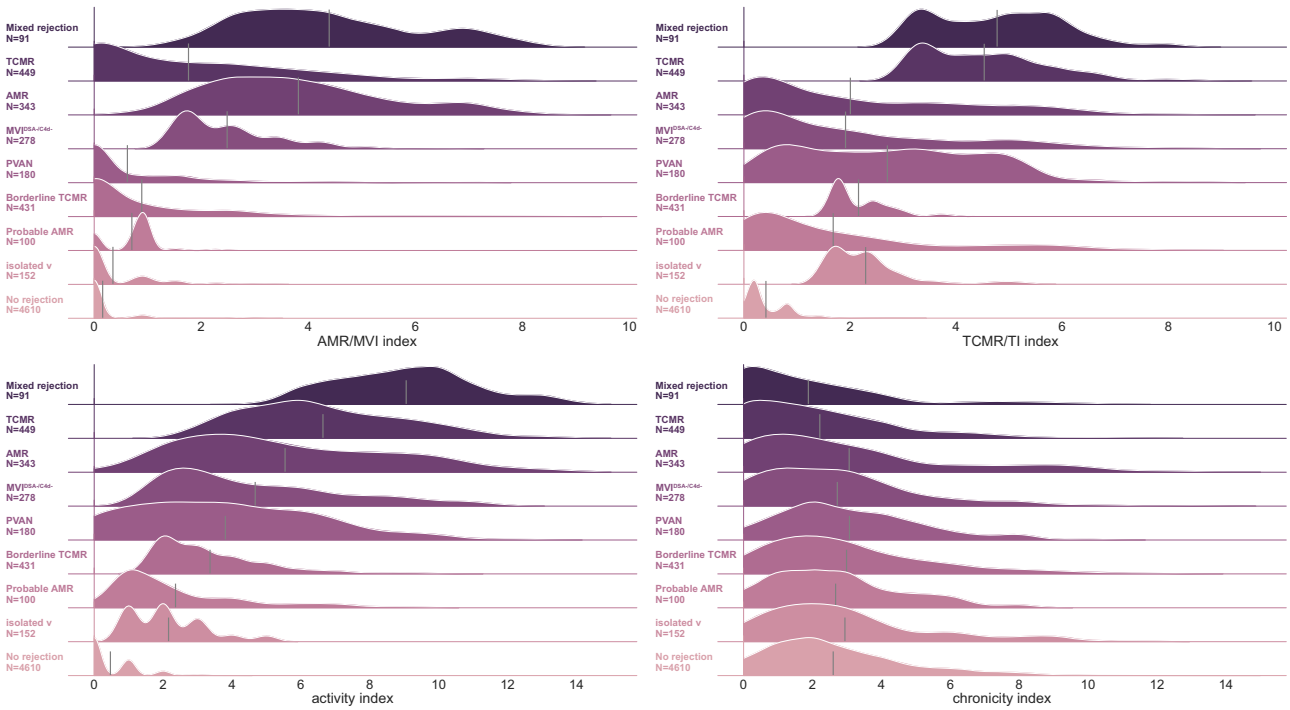

Fig. 1 | Distribution of the indices among the main Banff diagnoses in the derivation cohort (*N* = 6272 biopsies). Antibody-mediated rejection/microvascular inflammation (AMR/MVI) and T-cell-mediated rejection/tubulointerstitial inflammation (TCMR/TI), activity and chronicity indices, ordered by the mean activity index per category (gray vertical line). Both the AMR/MVI and TCMR/TI indices demonstrated logical and hierarchical ordering between No rejection, intermediate (Probable AMR, Borderline TCMR, isolated v) and complete rejection histology, with, in some cases, significant overlap. The activity index demonstrated similar gradual and hierarchical ordering between No rejection biopsies and fully inflamed categories, with intermediate categories (Borderline TCMR, Probable AMR and isolated v) in between. The chronicity index demonstrated low variability among the main diagnostic categories.

https://doi.org/10.1038/s41467-025-65153-9

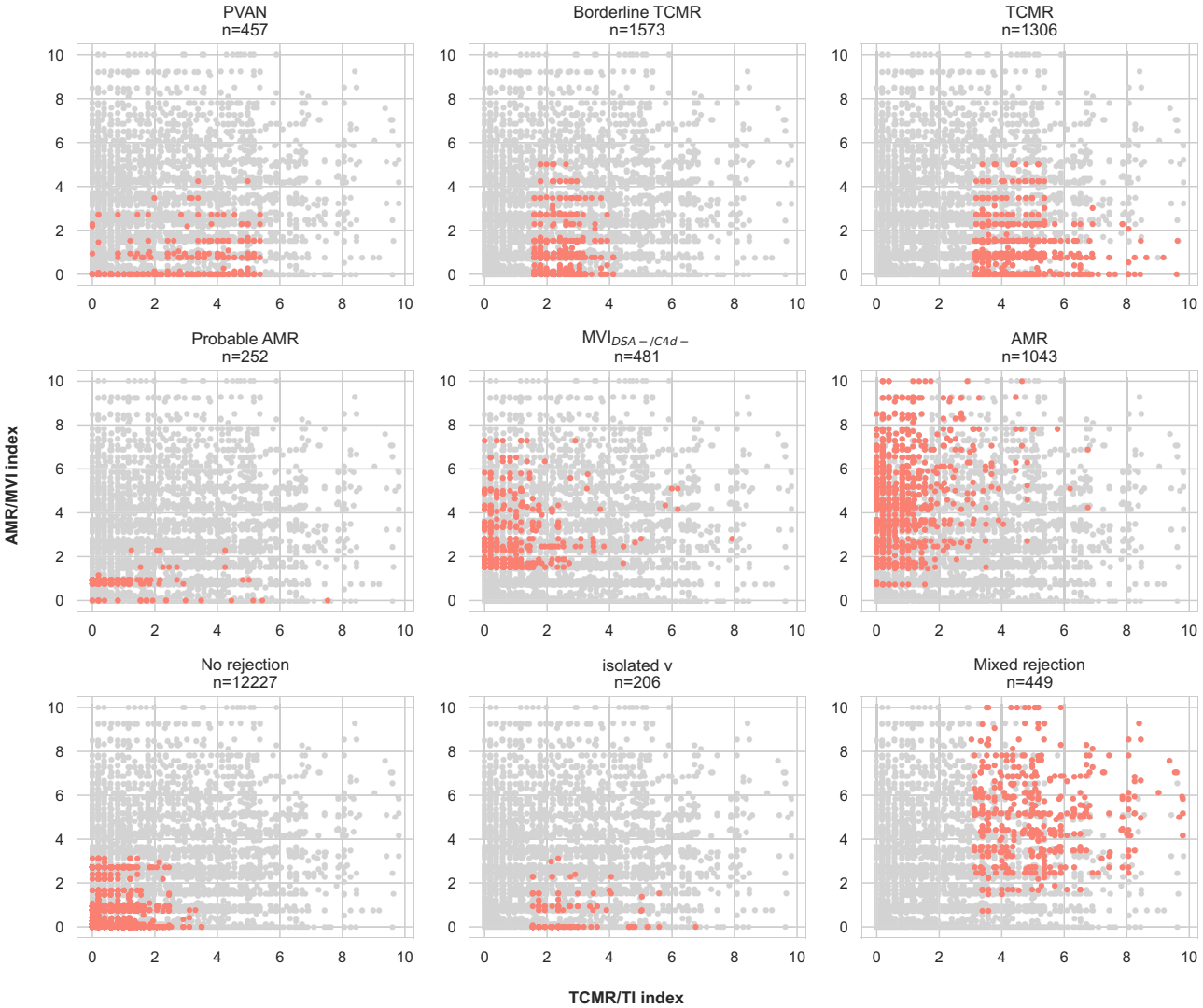

Fig. 2 | Comparison of the continuous rejection indices to Banff diagnostic categories in all classifiable biopsies, *N* = 18,778. Distribution of the individual biopsies per diagnostic category (in red) on the basis of the antibody-mediated rejection/microvascular inflammation (AMR/MVI) (y axis) and T-cell-mediated rejection/tubulointerstitial inflammation (TCMR/TI) indices (x axis). For visualization purposes, only the 'pure' phenotypes with no other concomitant diagnoses are displayed. No rejection biopsies are strictly characterized by both low AMR/MVI and low TCMR/TI indices, whereas Mixed rejection cases have both high AMR/MVI and TCMR/TI indices. High heterogeneity in Mixed rejection cases is, however, observed, resulting from the combination of various levels of intensity of the AMR and TCMR phenotypes within this diagnostic category. TCMR and Borderline TCMR cases are contained within clear boundaries of the TCMR/TI index. Probable AMR cases are restricted by a low AMR/MVI index. We note a significant overlap of the $MVI_{DSA-/C4d-}$ cases with the AMR cases in terms of the AMR/MVI index, although high values of the AMR/MVI index are affected only by biopsies.

the definition of $MVI_{DSA-/C4d-}$, for which C4d negativity is required. TCMR and Borderline TCMR cases fell within well-defined boundaries set by the TCMR/TI index. Overall, the four indices were systematically greater in indication biopsies than in protocol biopsies (Supplementary Fig. 1), albeit with significant overlap across the range, demonstrating that severe inflammation and chronic injury can occur subclinically, with stable graft function.

The activity index exhibited a gradual and hierarchical increase from No rejection to fully inflamed categories (AMR, $MVI_{DSA-/C4d-}$, TCMR and Mixed rejection), with intermediate categories (Borderline TCMR, Probable AMR and isolated v) in between (Fig. 1). The chronicity index demonstrated low variability among the main diagnostic categories (Fig. 1). Despite significantly different mean values of activity and chronicity indices among the acute/active, chronic-active, and chronic subcategories of AMR and TCMR, substantial overlap was observed in these indices across the subcategories (Supplementary Fig. 2). The latter analysis of the subcategories of TCMR was performed on a subset of the derivation cohort ($N = 1635$), for which scoring of critical lesions of chronic active TCMR, namely, ti and i-IFTA, was available. On another subset of 1254 biopsies with available lesion scores, the "i-IFTA" and "t-IFTA" scores correlated more strongly with the chronicity index (Spearman correlation: 0.40, $p < 0.001$ and 0.43, $p < 0.001$, respectively) than with the activity index (Spearman correlation: 0.10, $p < 0.001$ and 0.26, $p < 0.001$) and were therefore not considered for inclusion in the activity index.

The distributions of the four continuous indices across the main diagnostic categories were similar in the validation cohort and the derivation cohort (Supplementary Figs. 3 and 4), although the chronicity index was noticeably greater in both validation cohorts, with a greater spread and gradual increase from no rejection to mixed rejection categories. In the US cohort, TCMR/TI and activity indices were higher in cases of probable AMR, MVI, and AMR compared to the other cohorts. Higher activity indices were also observed in PVAN cases. The correlation between the activity indices and the chronicity

Table 3 | Discrimination performance as AUC with 95% CI of the acute indices (activity, AMR/MVI and TCMR/TI indices) for relevant binary outcomes in all three cohorts

Discrimination	Derivation cohort *n* = 6272	European validation cohort *n* = 10,338	US validation cohort *n* = 2162
AMR/MVI index			
AMR vs. No AMR	0.98 (0.97–0.98)	0.97 (0.96–0.97)	0.96 (0.95–0.97)
MVI/AMR vs. No MVI/No AMR	0.98 (0.98–0.99)	0.98 (0.98–0.98)	0.97 (0.96–0.98)
Probable AMR/MVI/AMR vs. No probable AMR/MVI/AMR	0.96 (0.95–0.96)	0.96 (0.96–0.97)	0.96 (0.95–0.97)
TCMR/TI index			
TCMR vs. No TCMR	0.99 (0.99–0.99)	0.99 (0.99–0.99)	0.98 (0.97–0.98)
(Borderline) TCMR vs. No (Borderline) TCMR	0.98 (0.97–0.98)	0.98 (0.98–0.98)	0.95 (0.94–0.96)
PVAN vs. No PVAN	0.83 (0.80–0.86)	0.80 (0.77–0.83)	0.84 (0.80–0.87)
Activity index			
Mixed rejection vs. No mixed rejection	0.98 (0.98–0.99)	0.98 (0.98–0.99)	0.96 (0.95–0.97)
No rejection vs. Any other category	0.96 (0.95–0.96)	0.97 (0.97–0.97)	0.99 (0.99–0.99)
Any rejection vs. No rejection	0.96 (0.95–0.96)	0.95 (0.95–0.96)	0.96 (0.95–0.97)

AMR antibody-mediated rejection, *AUC* area under the receiver operating characteristic curve, *CI* confidence interval, *MVI* microvascular inflammation (DSA-/C4d-), *PVAN* polyomavirus nephropathy, *TCMR* T-cell-mediated rejection, *TI* tubulointerstitial inflammation.
Biopsies that were unclassifiable (*n* = 722) because key lesions were missing were excluded.

indices was also more pronounced in both the European and the US validation cohorts (correlation coefficient [r]=0.18, $P < 0.001$ and r = 0.42, $P < 0.001$, respectively) than in the derivation cohort (r = −0.08, $P < 0.001$).

Discriminative performance of the AMR/MVI and TCMR/TI indices for Banff diagnostic categories

To demonstrate that the AMR/MVI and TCMR/TI indices maintain congruence with the clinically used Banff classification system of AMR and TCMR, we next evaluated the discriminative performance of the AMR/MVI and TCMR/TI indices for the Banff diagnostic categories, analyzing the derivation and validation cohorts separately.

In the derivation cohort (Table 3), the AMR/MVI index discriminated the presence of AMR from the absence of AMR, with an area under the receiver operating characteristic curve (AUC) of 0.98 (95% confidence interval [CI], 0.97 to 0.98). When $MVI_{DSA-/C4d-}$ cases were combined with AMR cases, the AUC was 0.98 (95% CI, 0.98 to 0.99). The TCMR/TI index demonstrated an AUC of 0.99 (95% CI, 0.99 to 0.99) for discriminating between the presence and absence of TCMR. This variable performed equally well if Borderline TCMR cases were considered TCMR cases (AUC 0.98; 95% CI, 0.97 to 0.98). Compared with the other diagnostic categories, the TCMR/TI index underperformed for the discrimination of PVAN, with an AUC of 0.83 (95% CI, 0.80–0.86), which can be explained by the absence of specific polyomavirus markers (e.g., SV40 staining results) in the TCMR/TI index formulation. The activity index discriminated No rejection from all other categories, with an AUC of 0.96 (95% CI, 0.95–0.96), and Mixed rejection from No Mixed rejection, with an AUC of 0.98 (95% CI, 0.98–0.99), clearly outperforming both the AMR/MVI and TCMR/TI indices on those two tasks. Overall, the area under the precision-recall curve (AUPRC), which is less impacted by significant class imbalance, demonstrated similar ordering of the variable performances among the different Banff categories as the AUC results (Supplementary Tables 1, 2 and 3). Compared to individual Banff lesions, the continuous indices provided smoother and more discriminative ROC curves and consistently outperformed single lesions in distinguishing between key diagnostic categories (Fig. 3).

As shown in Table 3 and Supplementary Tables 1, 2 and 3, the performances of all indices were similar in both validation cohorts for all discrimination tasks.

Additionally, sensitivity analyses revealed similar discrimination performance of the three acute indices (activity, AMR/MVI and TCMR/TI indices) across biopsy status (protocol vs. indication), recipient sex, recipient ethnicity, or biopsy timing (early vs. late post-transplant) (Supplementary Tables 4 to 7). The net benefit of the AMR/MVI and TCMR/TI indices was demonstrated for the discrimination of AMR and TCMR (Supplementary Results and Supplementary Fig. 5). The classification performances of the AMR/MVI and TCMR/TI indices across their ranges are detailed in the Supplementary Results: AMR/MVI and TCMR/TI indices ≥3 had 99.2% and 96.8% specificity for $AMR/MVI_{DSA-/C4d-}$ and TCMR, respectively (Supplementary Tables 8, 9 and 10). All Probable AMR ($N = 327$) and Borderline TCMR ($N = 1972$) cases had AMR/MVI and TCMR/TI indices, respectively, that were strictly inferior to 3 and strictly superior to 0 for Borderline TCMR cases (Supplementary Tables 11 and 12). For illustrative purposes, though we do not recommend discretizing the continuous indices, the AMR/MVI and TCMR/TI indices were divided into four categories using simple cut-off values (1, 3, and 6) (Supplementary Table 13). Most AMR and TCMR diagnoses fell into the high or severe categories, while the vast majority of "No Rejection" cases were classified in the low category. The few cases of No rejection with high indices were explained by peculiar lesion combinations (e.g. cg3, ptc0, g0, C4d0). Notably, intermediate phenotypes such as Borderline rejection and $MVI_{DSA-/C4d-}$ predominantly clustered in the medium index range, highlighting the ability of the indices to reflect a biologically meaningful continuum.

Associations of the indices with graft failure

To assess the relationships between different disease stages and outcomes, we analyzed the association of the four indices with long-term graft failure (Fig. 4; Supplementary Tables 14-15).

In the derivation cohort, the four indices were associated with graft failure in Cox models adjusted for time posttransplantation, donor age, recipient age and recipient sex, with adjusted hazard ratios (HRs) of 1.30 (95% CI, 1.25–1.36) for the activity index, 1.29 (95% CI, 1.23–1.36) for the chronicity index, 1.49 (95% CI, 1.39–1.60) for the AMR/MVI index and 1.50 (95% CI, 1.39–1.61) for the TCMR/TI index, (based on the last biopsy per patient). Similar associations were observed in the two validation cohorts. In all three cohorts, the associations of the indices with graft failure were independent of the

https://doi.org/10.1038/s41467-025-65153-9

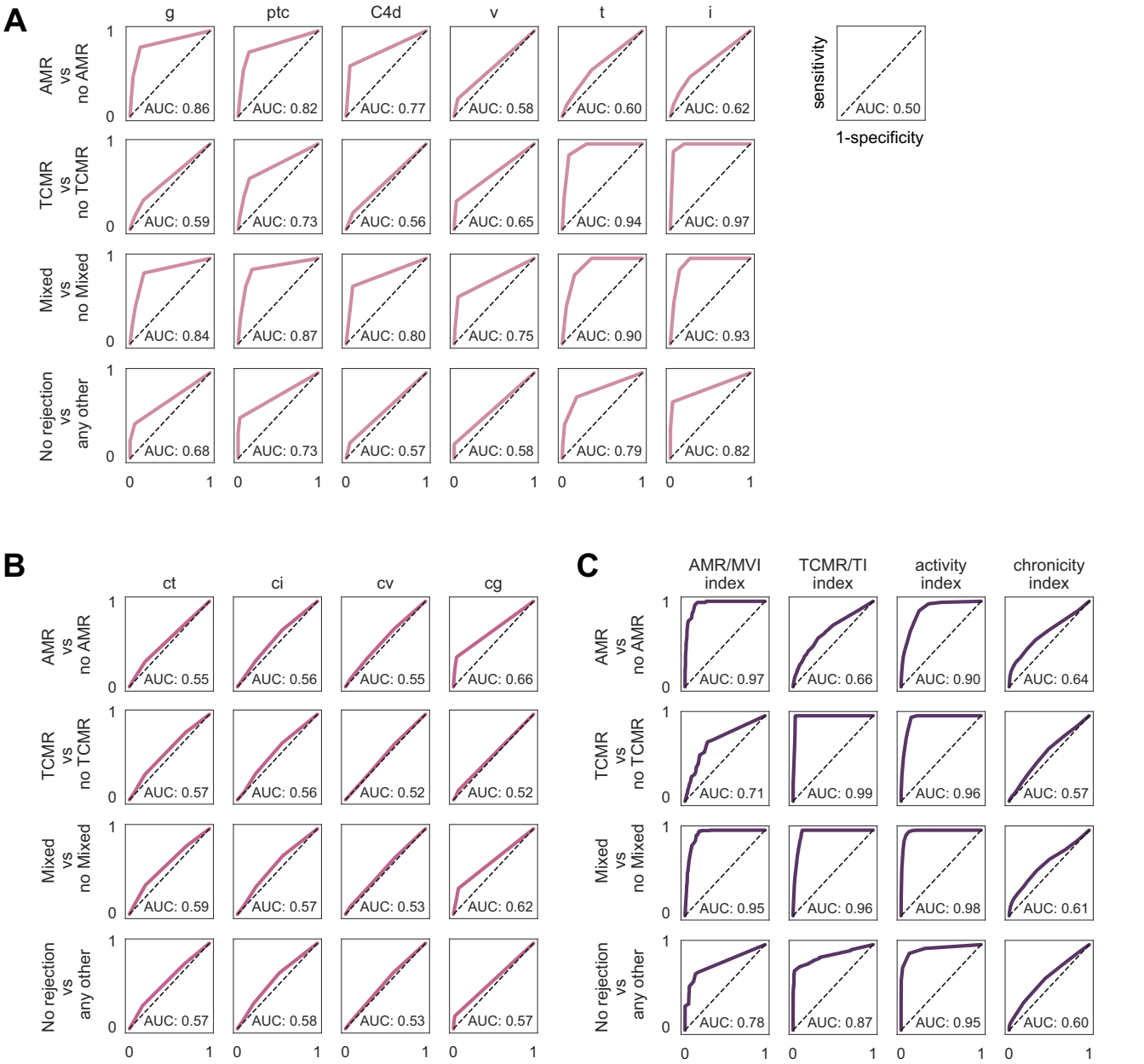

Fig. 3 | Comparison of the discriminative performance of individual Banff lesions and continuous indices across relevant Banff diagnostic categories. Panel **A** shows ROC curves for individual acute Banff lesions, panel **B** for chronic lesions, and panel **C** for the derived continuous indices. ROC curves from individual Banff lesions appear stepwise due to the ordinal scoring system (four discrete levels), which limits granularity. In contrast, continuous indices produce smoother and more discriminative curves. Across all diagnostic comparisons, the three acute indices—AMR/MVI, TCMR/TI, and the activity index—consistently outperformed individual lesion scores. As expected, the chronicity index did not discriminate among rejection phenotypes. The lesions selected for each index typically had the highest AUCs within their diagnostic category, supporting their biological and diagnostic relevance.

biopsy selected per patient. Notably, in a subset of the derivation cohort with available data (N = 3724), replacing the "i" score in the activity with "ti" in the activity index did not alter the results (Supplementary Fig. 6). In the overall cohort, the AMR/MVI and TCMR/TI indices yielded strata of increasing severity, as indicated by the significant associations of the AMR/MVI and TCMR/TI indices with graft failure (Fig. 5, Supplementary Fig. 7). Additionally, the indices effectively discriminated significantly different survival trends within the main Banff diagnostic categories. For example, AMR cases with a higher AMR/MVI index presented a greater risk of graft failure than did those with a lower AMR/MVI index. The same was true for $MVI_{DSA-/C4d-}$ cases stratified by the AMR/MVI index, as well as for (Borderline) TCMR cases stratified by the TCMR/TI index (Supplementary Fig. 8).

The activity, chronicity and TCMR/TI indices demonstrated linear relationships with graft failure rates (i.e., a proportional increase in event rate with increasing index values, Fig. 5E). The AMR/MVI index demonstrated slight departures from a strict linear association ($p = 0.004$), although its overall relationship with graft failure remained stable and clinically interpretable (Fig. 4E). For each unit increase in the activity and chronicity indices, the relative hazard for graft failure increased by 15.0% (95% CI, 13.6 to 17.0%), 20.6% (95% CI, 18.0 to 22.9%) and 33.0% (95% CI, 28.0 to 38.2), respectively (computed on the whole cohort on the basis of a random biopsy per patient). These numbers, as reflected by HR, are also similar across the different cohorts (Fig. 3, *all biopsies*; Supplementary Tables 14 and 15). Overall, each unit increase in the AMR/MVI index corresponds roughly to a

https://doi.org/10.1038/s41467-025-65153-9

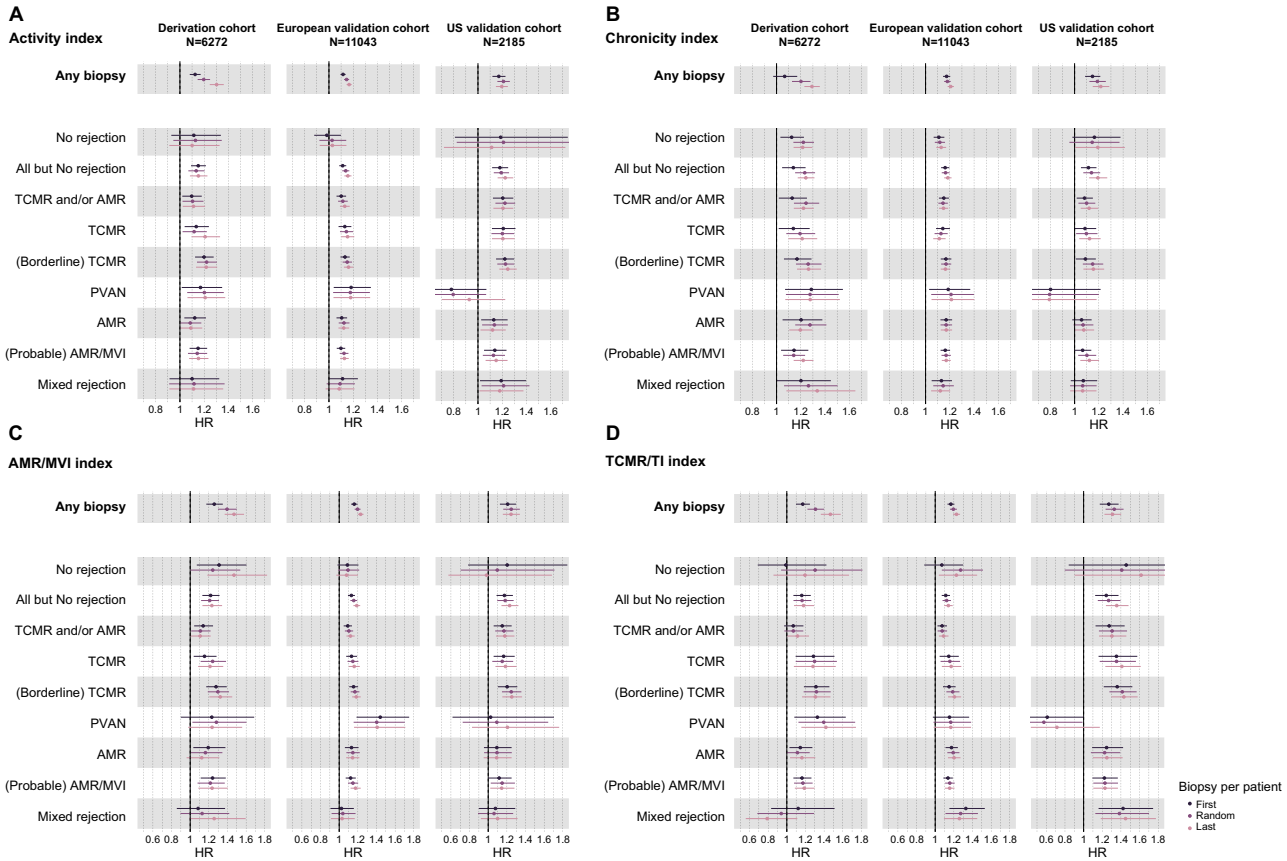

Fig. 4 | Associations between continuous indices and graft failure across Banff categories and cohorts. Associations with graft failure are shown for the activity (**A**), chronicity (**B**), AMR/MVI (**C**) and TCMR/TI (**D**) indices. Dots represent hazard ratios (HR) as the measure of centre, and error bars indicate the 95% confidence intervals (CIs). HR are adjusted for time posttransplantation, donor age, recipient age and recipient sex in the three cohorts. HR refers to a one-unit increase in the indices. The corresponding numbers, as well as the subsets sample sizes are reported in Table S14-S15.

22.4% (95% CI, 19.8 to 25.0%) increase in the relative hazard for graft failure (under a linear association).

These indices also maintained an association with graft outcome within most of the main Banff-defined diagnostic categories (Fig. 3, Supplementary Tables 14-15). Only in the No rejection and in Mixed rejection cases, which are less heterogeneous in terms of histological lesion scores, did the acutes indices (activity, AMR/MVI and TCMR/TI indices) not stratify outcomes further. Overall, similar associations of the indices with graft outcome were observed in both validation cohorts, except in the US cohort where AMR and chronicity indices were not associated with outcome in AMR cases, and no indices were associated in PVAN cases. In contrast to the derivation cohort, the activity and TCMR/TI indices even stratified outcomes within Mixed rejection cases in both validation cohorts.

Globally, the four indices associated with graft outcome irrespective of recipient sex, recipient ethnicity, timing of biopsy or indication vs. protocol status (Supplementary Fig. 9, 'all biopsies'). These associations were maintained within most of the Banff categories with a few exceptions. The chronicity index was not associated with graft failure in the subset of protocol biopsies classified as TCMR. While the indices demonstrated similar associations with outcomes in Black and White recipients, these associations appeared attenuated in Asian recipients for all four indices. Whether this reflects a true biological difference or is due to limited statistical power in this subgroup remains unclear and warrants further investigation in larger datasets. Finally, we observed slightly higher associations with graft failure in late, biopsies compared to early biopsy for all indices for all categories.

The indices remained associated with graft failure within the AMR active, chronic-active, and chronic subcategories (Supplementary Fig. 10, Supplementary Table 11). Similar analyses of the TCMR/TI index within TCMR subcategories could not be performed because of the low prevalence of chronic active TCMR in the derivation subset and the lack of key Banff lesion scores, specifically ti and i-IFTA in the validation cohorts.

Discussion

This study indicates that four continuous measures of kidney transplant rejection, derived from standard histological lesion scores, capture much of the histological spectrum and severity of rejection while closely aligning with the current Banff classification. The AMR/MVI and TCMR/TI indices directly reflect the two distinct yet potentially concurrent inflammation patterns of kidney allografts: microvascular and tubulointerstitial, independent of the underlying disease cause. The activity and chronicity indices provide a global evaluation of the level of inflammation and the level of chronic damage, respectively. The AMR/MVI and TCMR/TI indices effectively discriminate between their corresponding Banff-defined rejection categories and constitute simple yet meaningful proxies for continuous rejection processes, especially when the TCMR and AMR processes are present simultaneously in mixed rejection phenotypes. The AMR/MVI and TCMR/TI indices correlate with graft outcomes even within the Banff diagnostic categories. The activity index helps distinguish between rejection and no rejection and assesses inflammation severity, with the latter being associated with graft failure independent of the underlying disease and even within the Banff AMR acute/active, chronic-active, and chronic

https://doi.org/10.1038/s41467-025-65153-9

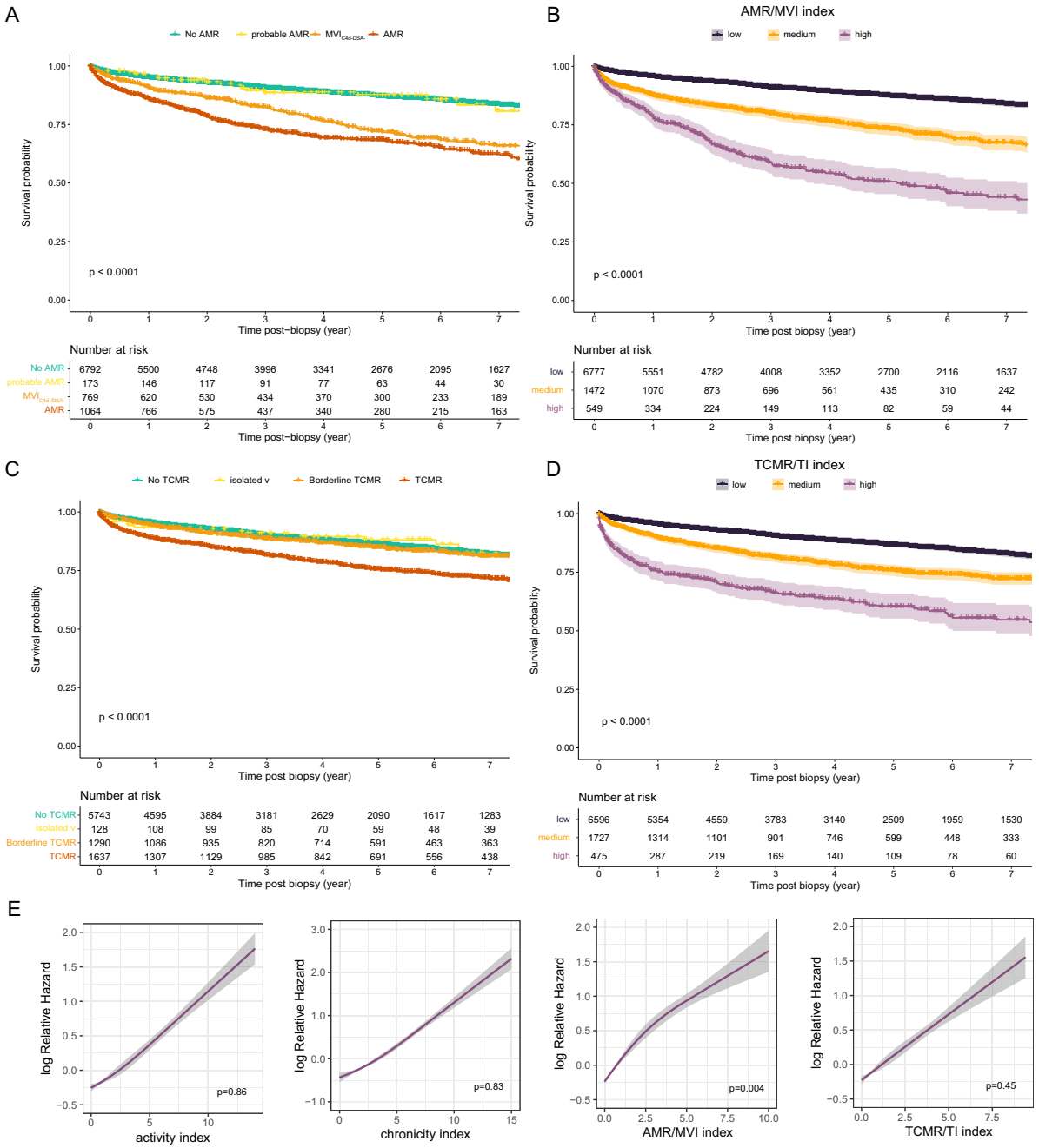

Fig. 5 | Risk of graft failure across diagnostic categories and index-based stratifications and functional form of index associations. Kaplan-Meier curves of the AMR-related (**A**) and TCMR-related (**C**) Banff diagnostic categories and stratification of the AMR/MVI (**B**) and TCMR/TI (**D**) indices in three strata based on arbitrary cut-offs (all cohorts, $N = 19{,}500$). *No AMR* refers to all biopsies that are not AMR, Probable AMR or $MVI_{DSA-/C4d-}$. Similarly, *No TCMR* refers to all biopsies that are not TCMR, Borderline TCMR or isolated v. Note that beyond those non-overlapping categories, concomitant diagnoses are possible (e.g. AMR cases can co-exist with TCMR and Borderline). Due to the negative exponential aspect of the index distributions, the discretization into low, medium and high index, was based on the index value at the 75th and 95th percentiles, corresponding to the following pairs of thresholds: 0.9 and 4.6 and; 1.6 and 4.5, and for the ARM/MVI index and TCMR/TI index, respectively. All Kaplan-Meier curves use the last biopsy per individual. *P*-values refer to log-rank tests. Shaded areas in panels B and D represent the 95% confidence interval for the Kaplan–Meier survival estimates. **E** Association of the four indices with the hazard of graft failure. Shaded areas represent the 95% confidence intervals around the model fit. The activity, chronicity and TCMR/TI indices linearly associate with graft outcome (*p*-value from Wald test for non-linear effect with restricted cubic splines with 3 knots: 0.86, 0.83 and 0.45, respectively). The AMR/MVI index demonstrated slight departures from a strict linear association ($p = 0.004$). Models based on a random biopsy per patient, all cohorts ($N = 19{,}500$).

subcategories. The chronicity index provides prognostic information independent of the diagnostic category and disease activity. For every one-unit increase in the activity and chronicity indices, we observed linear increases of 15.0% and 20.6%, respectively, in the rate of graft failure.

The AMR/MVI and TCMR/TI indices discriminate full rejection phenotypes, providing a detailed view of the disease spectrum; moreover, these indices are increased in cases borderline/suspicious for TCMR, with a diagnosis of probable AMR, or with isolated v lesions, albeit below the index values of fully developed phenotypes. These

findings support their clinical interpretation of such cases as intermediate phenotypes rather than distinct diagnostic categories[15,16]. Moreover, the relationship between AMR/MVI and TCMR/TI indices and graft outcomes, even within these intermediate phenotypes, highlights the additional information regarding disease severity that the continuous scores provide compared with the Banff categories.

On the other hand, the large histological heterogeneity of the mixed rejection cases is particularly apparent when the combination of AMRMVI and TCMR/TI indices is studied: in addition to high AMR/TI-high TCMR/TI indices, combinations such as low AMR/MVI-high TCMR/TI indices or high AMR/MVI-low TCMR/TI indices are all labelled similar to Mixed rejection. The AMR/MVI and TCMR/TI indices provide a more detailed evaluation of the relative contributions of the AMR versus TCMR components in such mixed cases.

While there is generally strong concordance between the continuous AMR/MVI and TCMR/TI indices and the Banff classification, further discussion of their relationship with other phenotypes is warranted. The histology of $MVI_{DSA-/C4d-}$ resembles that of AMR, with glomerulitis and/or peritubular capillaritis, resulting in high AMR scores for this phenotype as well. The only difference in the Banff classification of these cases and Banff-defined AMR is that the underlying disease cause is not directly detected via routine histology. Similarly, the histology of PVAN often resembles that of TCMR, exhibiting tubulointerstitial inflammation[17]. Therefore, it is not surprising that the TCMR/TI index is often high in such PVAN cases, as would be the case for other diseases with a tubulointerstitial nephritis phenotype due to other causes (e.g., drug toxicity). The underperformance of the TCMR/TI index in discriminating PVAN cases is explained by the lack of polyomavirus-specific markers in the index formulation (e.g., SV40-positive staining) and the considerable heterogeneity in PVAN presentations, which can range from noninflamed to heavily inflamed. Overall, the associations of the AMR/MVI and TCMR/TI indices with $MVI_{DSA-/C4d-}$ and PVAN, respectively, underscore the necessity of interpreting the histological picture in conjunction with additional testing for underlying causes. Specifically, HLA-DSA testing is crucial for distinguishing AMR from $MVI_{DSA-/C4d-}$, and SV40 staining (and peripheral blood PCR) is essential for differentiating between TCMR and PVAN diagnoses. As these indices do not contain information on disease causes, they should be interpreted as descriptors of the histological picture, requiring ancillary testing to discriminate rejection from potential other injury processes leading to a similar inflammatory pattern.

The AMR/MVI and TCMR/TI indices provide a detailed view of the disease spectrum, avoiding consensus-based thresholds between intermediate diagnostic categories such as Borderline TCMR, isolated v, $MVI_{DSA-/C4d-}$ and Probable AMR[14,15,18]. However, as we demonstrated that the relationship between the diagnostic categories and the indices is very strong, reporting on the continuous variables, e.g., in pathology reports, can be easily performed in parallel with the dichotomized diagnostic categories without affecting their interpretation. In most cases, our data confirm that the indices align well with the diagnostic categories. Discrepancies are mostly found in those expected cases where the Banff system potentially overemphasizes the relevance of certain markers (e.g., HLA-DSA-positive cases with minimal inflammation considered AMR; underestimation of the relevance of t3 + i1 or i3 + t1 in cases classified as Borderline TCMR). From this perspective, it can even be expected that increasing experience with the continuous variables calculated from Banff lesion scores, and perhaps in the future with biopsy-based transcript analyses, will eventually lead to the refinement of the Banff definitions for the diagnostic categories per se.

In this study, the activity and chronicity indices were associated with graft failure risk and showed high heterogeneity and relationships with outcomes within the AMR subcategories (active, chronic-active, and chronic). These findings indicate that these subcategories do not fully capture disease stage, as intended. Replacing subcategories with activity and chronicity indices in biopsy reports could improve clinical decision-making and offer more granularity for clinical trials. Highly active disease could benefit from treatment with anti-inflammatory or targeted therapies aimed at immune activation and its underlying etiology. More chronic disease presentations with less activity could be less reversible by such etiological therapies. Although intuitive, the clinical value and benefit of grading activity and chronicity alongside diagnostic categories should be tested in prospective trials. Finally, although the strong associations with graft failure reported here highlight the clinical relevance of the indices, they should not be regarded as standalone prognostic tools. These indices are derived solely from histological lesion scores and do not incorporate other prognostically significant factors such as kidney function (eGFR), proteinuria, DSA status, or time since transplantation[19]. The primary value of the indices lies in the detailed characterization of phenotype severity and disease stage, which is crucial for guiding therapeutic decisions. However, this should not be mistaken for a comprehensive prognostication system.

Our activity and chronicity indices were constructed on the basis of previous mathematical modeling, with further refinement on the basis of discussions within the Banff Working Group. However, other algorithms could be considered. For example, Haas et al. proposed an alternative activity index (g+ptc+v+C4d) that is specific for AMR cases[20]. However, they could not demonstrate an association with graft failure in this specific setting. In addition, more recently defined Banff lesion scores, such as ti, i-IFTA and t-IFTA, could be included in the models. However, we demonstrated that replacing the Banff i score with the ti score does not affect the results and that i-IFTA and t-IFTA are highly collinear with chronic lesions, potentially confounding activity and chronicity if these parameters are included in an activity index. Moreover, the lack of large datasets in which these lesions have been evaluated systematically hinders robust statistical analysis of the added value of including these lesions in the models. Further improvements in the activity and chronicity score definitions proposed herein should be tested in future studies.

This study has several limitations. Histological Banff lesion scores are required to determine the continuous indices; however, Banff lesion scores remain inherently pathologist dependent. Emerging techniques such as whole-slide image analysis could increase the reproducibility of lesion scores and derived indices. Similarly, the Banff diagnostic categories, used as ground truth here, have evolved during the study, and the Banff 2022 classification has been attributed retrospectively. Additionally, in the future, the Banff classification may evolve, although the distinction between essentially tubulointerstitial disease and microvascular disease has been stable over two decades, reflecting distinct biological mechanisms. The latter supports the long-term validity of the TCMR/TI and AMR/MVI indices, as well as future updates of the Banff classification. Nevertheless, future refinements of the indices should be explored. In particular, the integration of newly introduced or less routinely reported Banff lesions (e.g. i-IFTA, ti, pvl) may offer improved performance, although their limited availability in historical datasets remains a constraint. Information related to the treatment given based on the biopsy results was not available. Therapeutic approaches differ across centers, relate to the histological phenotypes and clinical presentation, and treatment options and protocols have evolved over the course of the study, potentially affecting the relationship between histology and outcomes. While this impact of therapeutic choices on outcomes cannot be assessed in this observational cohort study, it can be concluded that associations between histological phenotypes and worse outcomes were observed despite such therapeutic interventions, indicating an unmet need. Although a small proportion of biopsies, mostly in the European validation cohort, were unclassifiable due to missing lesion scores, the missingness appeared unrelated to any specific phenotype and unlikely to affect the generalizability of the results. In this study,

stratifications were for illustration purposes only, and continuous variables should ideally not be dichotomized[21]. Additionally, grouping biopsies into broader rejection categories may mask heterogeneity. Finally, although our study is not a true multicenter study by design, the robustness of the indices was confirmed through validation in diverse and heterogeneous datasets, reflecting real-world clinical data and ensuring the generalizability of our findings beyond the controlled conditions of a traditional multicenter study. Likewise, although follow-up duration may limit the capture of very late events, the available data remain representative of real-world clinical timelines while maintaining meaningful biopsy–outcome associations.

In conclusion, we showed that continuous indices calculated from Banff lesion scores offer a straightforward and interpretable evaluation of kidney transplant biopsy histology that is generalizable across centers in Europe and the United States. The formulation of the indices is transparent and based on routinely assessed Banff lesion scores, enabling automated calculation and potential integration into structured pathology workflows and electronic health records to support both clinical decision-making and research. The AMR/MVI and TCMR/TI indices differentiate histological phenotypes and quantify the spectrum of inflammation, irrespective of its underlying cause. Additionally, the activity and chronicity indices hold significant clinical relevance, as each one-unit increase is associated with an overall 15–20% higher rate of graft failure. Overall, these indices offer a nuanced understanding of the continuous rejection process and support probabilistic reasoning in the diagnosis of kidney transplant rejection.

Online methods

The study was approved by the Ethics Committee Research of the University Hospitals Leuven (S64006), with data transfer agreements with the contributing centers. Written informed consent for patient inclusion was waived by the ethics committee as this is not legally mandated for studies on data collected as part of routine clinical practice, since these do not fall under the scope of the Belgian law of May 7, 2004. The research was performed in accordance with the guidelines of the Declaration of Helsinki and the Guidelines for Good Clinical Practice.

Study population and clinical data

A total of 8977 adult kidney transplant recipients from 10 European and US transplantation centers were included in this multicenter observational cohort study.

The derivation cohort consisted of all adult kidney transplantations performed at the University Hospitals Leuven, Belgium, between March 2004 and May 2021 ($N = 1891$; ClinicalTrials.gov ID NCT06505200). The transplantations were performed with negative complement-dependent cytotoxicity crossmatches. Patients who received combined transplants or kidney transplants following a different solid organ transplant were excluded. The standard immunosuppressive maintenance protocol consisted of tacrolimus, mycophenolate, and corticosteroids[22]. Posttransplantation biopsies were conducted on the basis of medical indications (indication biopsies at the time of graft dysfunction) or as part of an established follow-up protocol (protocol biopsies), primarily at 3, 12, and 24 months posttransplantation. All clinical data for the derivation cohort were prospectively gathered from electronic health records. Data inclusion concluded on June 11, 2022. Clinical data included donor demographics (type of donor, age, sex, and diabetes status), recipient demographics (age, self-reported sex, self-reported ethnicity, and body mass index), and transplant characteristics (HLA-DSAs, HLA antigen mismatches, cold ischemia time, graft failure, and mortality). Information on ethnic background was gathered to facilitate discussions on the generalizability of the findings. Graft failure was defined as a return to dialysis or retransplantation. In the case of death with a functioning graft, we censored graft failure at the time of death.

We performed validation in two external cohorts: a European validation cohort consisting of 5898 transplants and a US validation cohort of 1161 transplants. In these validation cohorts, we obtained the same clinical and biological data as those collected for the derivation cohort according to routine clinical care, in accordance with local and national regulations, and submitted them to the central research database in Leuven (Supplementary Methods).

Kidney transplant biopsy evaluation

All allograft biopsies in the derivation cohort were prospectively cataloged in the electronic health records of the patients. Pre- and post-transplantation HLA-DSA status was monitored in one histocompatibility laboratory (HILA – Red Cross Flanders), as reported previously for this cohort[23].

In both the derivation and the validation cohorts, all transplant biopsies were evaluated via semiquantitative scoring of Banff lesion scores according to the Reference Guide to the Banff 2022 classification (https://banfffoundation.org/central-repository-for-banff-classification-resources-3/). Biopsies that did not reach the threshold for minimal sample requirements were excluded[24].

Development of continuous histological indices

AMR/MVI and TCMR/TI indices. The two indices aimed at distinguishing microvascular inflammation (MVI) from tubulointerstitial inflammation (TI) patterns were constructed using a latent variable framework in which a continuous unobserved variable (representing the underlying intensity of rejection) is inferred from observed Banff lesion scores. We used a probit model, in which the binary diagnostic categories (AMR vs. no AMR and TCMR vs. no TCMR) are treated as a threshold response to a continuous latent variable. This approach assumes that the categorical diagnosis reflects whether the latent severity has crossed a threshold, which is conceptually aligned with the assumption that AMR and TCMR are not binary phenomena, but rather the discretization of an underlying continuous rejection process. The model was implemented using the OrderedModel function from the Python statsmodels package[1] (version 0.13.5), specifying the probit as the distributional family.

On the basis of the Banff classification, we established a list of potential variable candidates for the indices: g, ptc, v, C4d, v, i, t, ci, ct, cv and cg. As part of the modeling process, a feature selection step was performed to remove lesion scores that did not significantly contribute to the latent variable: backward selection was used to sequentially eliminate the Banff lesions that were not significant (at a p value threshold of 0.05) in the latent variable. Finally, for ease of clinical use, we scaled both latent variables to the [0,10] interval to produce the final AMR/MVI and TCMR/TI indices: the theoretical maximum value of the latent variable, $\max_v$, is achieved when all individual lesions reached their maximum score (i.e. 3 for ordinal lesions or 1 for C4d). We therefore rescaled the initial coefficients of the model by a factor $10/\max_v$, so to reach a theoretical maximum index value of 10.

The indices were derived from models treating all biopsies as independent, as the point estimates used to construct the indices remain unaffected by clustering adjustments for repeated measurements. Model fitting was conducted via maximum likelihood estimation, and performance was assessed with the AUC for the relevant diagnostic categories, ensuring that the inferred indices reliably reflected the associations between the lesions and the diagnostic outcomes.

Activity and chronicity indices. Previously, we established activity and chronicity indices, as the sum of the active and chronic Banff lesion scores were reweighted with the normalized coefficient's z score of univariate Cox models for graft failure and scaled to the unit interval (from 0 to 1; https://rejectionclass.eu.pythonanywhere.com)[10,11]:

- Activity index= 0.049*t + 0.061*i + 0.066*v + 0.062*g + 0.062*ptc + 0.050*thrombi +0.052*C4d + 0.169*DSA
- Chronicity index= 0.100*cg+ 0.074*ci+ 0.050*ct+ 0.038*cv+ 0.061*mm+ 0.052*ah+ 0.049*gs

For ease of clinical implementation, the coefficients were rounded to integers. We simplified the activity index as t + i + v + g + ptc + 2 × C4d (score 0–17), where C4d is a binary indicator (0 or 1) of C4d positivity, which avoids the issue of heterogeneity in C4d staining methods between centers. As a result, to maintain its relative weight compared to other lesions, C4d coefficient was doubled. In addition, HLA-DSA status, a nonhistological parameter, was removed from the original formulation after discussion with the Banff Activity and Chronicity Indices Working Group (L.C., M.H., M.N., and S.S.). For the chronicity index, the simplification of the initial formula led to the same chronicity index as that defined by Haas et al.[20], i.e., 2 × cg + ci + ct + cv (score 0–15), with cg having a greater association with graft outcome, as reported in our original chronicity index formulation[3]. To maximize clinical implementation across centers, we retained only the most clearly defined, reproducible and frequently reported Banff lesions for establishment of the continuous indices. The lesion score for thrombi was therefore dropped out from the initial formulation, as it is not commonly reported in many centers. However, more recently defined Banff lesions, representing both activity and chronicity (total inflammation score "ti"; interstitial inflammation in the scarred area "i-IFTA"; tubulitis in the scarred area "t-IFTA"), were also considered in the updated formulas.

Missing data

Missing data were imputed with multiple imputation by chained equations (MICE)[25], with predictive mean matching to generate 20 imputed datasets. Statistical models were fitted separately within each imputed dataset, and the resulting estimates were then combined using Rubin's Rules to obtain pooled estimates. Data visualizations, which are not directly aggregable according to Rubin's rule, were produced on the basis of the average linear predictors from those 20 sets. The validation datasets were imputed independently of the derivation cohort. Missing histological lesion scores and confounding variables (mostly donor age) were assumed to be missing at random. Along with the lesion scores, the diagnostic categories and the indication status were used as additional proxy binary variables to guide the imputation. The proportions of missing histological lesion scores per cohort are reported in Table S19.

Outcome parameters

The main outcomes of interest were the discriminative performances of the four continuous indices (activity, chronicity, AMR/MVI and, TCMR/TI indices) for Banff-defined diagnostic categories and the associations of these indices with graft outcomes.

Statistical analysis

We followed the STROBE guidelines for reporting on observational studies (Supplementary Appendix).

Pairwise two-tailed *t*-tests were used to compare index distributions across mutually exclusive Banff diagnostic categories (No rejection, probable AMR, $MVI_{DSA-/C4d-}$ and AMR; No rejection, isolated v, Borderline TCMR and TCMR; Indication vs Protocol; active AMR, chronic-active AMR and chronic AMR; acute TCMR, acute+chronic active TCMR and chronic active TCMR). The discriminative performance of the indices was assessed on the basis of the presence or absence of Banff-defined phenotypes, with intermediate phenotypes first considered in the no disease category and subsequently taken together with complete phenotypes. Unclassifiable biopsies (due to key Banff lesions missing in the validation cohort) were excluded from the discrimination analyses. The AUC and AUPRC were used as discrimination metrics. We reported the raw AUPRC and AUPRC adjusted for the positive rate of each category in the cohort for comparison purposes. All biopsies were treated as independent for assessing discrimination performance, as both indices and diagnoses are determined solely by the histological lesion patterns, regardless of any external factors. To evaluate the clinical potential of an index beyond the rank-based metrics, decision curve analysis (DCA) was applied[26]. In DCA, the classification over a broad range of dichotomization thresholds is evaluated and compared with the benefit to two extreme decision policies[26]: classifying *none* of the biopsies as AMR (or TCMR) or *every* biopsy as AMR (or TCMR) and classifying all biopsies as *positive* or *negative* for a given phenotype. The AMR/MVI and TCMR/TI indices were compared with the AMR and TCMR regression models (Model 2 without HLA-DSAs and panel-reactive antibodies), as proposed earlier[12], in terms of the net benefit. The indices were transformed to the same numerical range by using individual logistic regression models trained on a given binary outcome (e.g., AMR versus No AMR). The correlation between a pair of indices was assessed via repeated measures correlation[27].

Graft failure was defined as the return to dialysis or retransplantation. Patients were censored at the last follow-up date or at the time of death. The associations with graft failure rates were evaluated with Cox models for all indices. To assess non-linearity in the association between the indices and the log hazard, we modeled the indices using restricted cubic splines (3 knots at default values) and tested the significance of the non-linear component using the Wald test provided by the anova() function in the rms package[28]. As a sensitivity analysis, the first, last or a random biopsy per patient was selected to establish the models to avoid overinflation of event rates and biases related to repeated measures. The overall associations are reported with HRs, adjusted for time posttransplantation, donor age, recipient age and recipient sex.

The main analyses were performed in R version 4.2.0 (R Foundation for Statistical Computing). The following packages were used: the MICE package[29] for multiple imputation, the RMS package[28] for discrimination performance, the survival package[30] for survival analyses, the dcurves package[31] for DCA and the rmcorr package[27] for the correlation adjusted for repeated measures. The figures were plotted with the ggplot2[32], survminer[33] and ggforestplot[34] packages. The remaining figures were produced in Python 3.9.7 via matlplotlib[35] and seaborn[36]. The latent variable models were produced via statsmodels 0.13.5[37].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset used in this study contains sensitive clinical information and cannot be shared openly due to patient confidentiality constraints. The validation data supporting this study were obtained from the participating centres under data transfer agreements. De-identified data from the derivation and validation cohorts are available for academic research purposes upon request to the corresponding author. Access requires approval from the participating centers' representatives, institutional and ethical clearance, and a data transfer agreement. Requests will receive an initial response within two weeks. The timeframes vary, but depending on the complexity of the request, it may take up to one year from the initial submission to share the requested data or documents. For all requests, please contact the corresponding author, M.N., directly by email at maarten.naesens@kuleuven.be. Source Data are provided with this paper for all the figures. A minimal working example dataset is provided to reproduce the index computation at github.com/ktxLeuven/Continuous_Indices Source data are provided with this paper.

Code availability

The code to reproduce the computation of the indices, along with a minimal working dataset, is publicly available at github.com/ktxLeuven/Continuous_Indices

References

1. Cornell, L. D. & Helanterä, I. Exploring Microvascular Inflammation and the Spectrum of Antibody-Mediated Rejection. *Am. J. Transplant.* S1600613524005379 (2024) https://doi.org/10.1016/j.ajt.2024.08.028.
2. Callemeyn, J. et al. Allorecognition and the spectrum of kidney transplant rejection. *Kidney Int* **101**, 692–710 (2022).
3. Naesens, M. et al. The Banff 2022 Kidney Meeting Report: Reappraisal of microvascular inflammation and the role of biopsy-based transcript diagnostics. *Am. J. Transplant. J. Am. Soc. Transplant. Am. Soc. Transpl. Surg.* **S1600-6135**, 00818–3 (2023).
4. Yoo, D. et al. An automated histological classification system for precision diagnostics of kidney allografts. *Nat. Med.* **29**, 1211–1220 (2023).
5. Vickers, A. J., Basch, E. & Kattan, M. W. Against diagnosis. *Ann. Intern. Med.* **149**, 200–203 (2008).
6. Roufosse, C. et al. The Banff 2022 Kidney Meeting Work Plan: Data-driven refinement of the Banff Classification for renal allografts. *Am. J. Transplant. J. Am. Soc. Transplant. Am. Soc. Transpl. Surg.* **S1600-6135**, 00855–00859 (2023).
7. Naesens, M. Embracing the Wisdom of Ancient Greece in the Era of Personalized Medicine-Uncertainty, Probabilistic Reasoning, and Democratic Consensus. *Transpl. Int. J. Eur. Soc. Organ Transplant.* **36**, 12178 (2023).
8. Simpkin, A. L. & Schwartzstein, R. M. Tolerating Uncertainty - The Next Medical Revolution?. *N. Engl. J. Med.* **375**, 1713–1715 (2016).
9. Goodman, K. E., Rodman, A. M. & Morgan, D. J. Preparing Physicians for the Clinical Algorithm Era. *N. Engl. J. Med.* **389**, 483–487 (2023).
10. Vaulet, T. et al. Data-driven derivation and validation of novel phenotypes for acute kidney transplant rejection using semi-supervised clustering. *J. Am. Soc. Nephrol.* (2021).
11. Vaulet, T. et al. Data-Driven Chronic Allograft Phenotypes: A Novel and Validated Complement for Histologic Assessment of Kidney Transplant Biopsies. *J. Am. Soc. Nephrol.* (2022) https://doi.org/10.1681/ASN.2022030290.
12. Sikosana, M. L. N., Reeve, J., Madill-Thomsen, K. S., Halloran, P. F. & Investigators, the I. Using Regression Equations to Enhance Interpretation of Histology Lesions of Kidney Transplant Rejection. *Transplantation* https://doi.org/10.1097/TP.0000000000004783.
13. Nankivell, B. J., Shingde, M. & P'Ng, C. H. The Pathological and Clinical Diversity of Acute Vascular Rejection in Kidney Transplantation. *Transplantation* **106**, 1666–1676 (2022).
14. Wellekens, K. et al. The Impact of the Banff v-Lesion on Rejection Classification and Outcomes: Insights from a Multicenter Study. *Am. J. Transplant.* (2025) 10.1016/j.ajt.2025.04.023.
15. Wellekens, K. et al. Probable antibody-mediated rejection in kidney transplantation is a rare and challenging phenotype to define: Findings from a single-center study. *Am. J. Transplant.* S1600613524004374 (2024) 10.1016/j.ajt.2024.07.014.
16. Nankivell, B. J. The meaning of borderline rejection in kidney transplantation. *Kidney Int* **98**, 278–280 (2020).
17. Cleenders, E. et al. An observational cohort study of histological screening for BK polyomavirus nephropathy following viral replication in plasma. *Kidney Int* **104**, 1018–1034 (2023).
18. Sablik, M. et al. Microvascular Inflammation of Kidney Allografts and Clinical Outcomes. *N. Engl. J. Med.* **392**, 763–776 (2025).
19. Loupy, A. et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. *BMJ* **366**, l4923 (2019).
20. Haas, M. et al. A Banff-based histologic chronicity index is associated with graft loss in patients with a kidney transplant and antibody-mediated rejection. *Kidney Int* **103**, 187–195 (2023).
21. Harrell, F. E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. (Springer, 2015).
22. Senev, A. et al. Histological picture of antibody-mediated rejection without donor-specific anti-HLA antibodies: Clinical presentation and implications for outcome. *Am. J. Transplant. J. Am. Soc. Transplant. Am. Soc. Transpl. Surg.* **19**, 763–780 (2019).
23. Senev, A. et al. Specificity, strength, and evolution of pretransplant donor-specific HLA antibodies determine outcome after kidney transplantation. *Am. J. Transplant.* **19**, 3100–3113 (2019).
24. Racusen, L. C. et al. The Banff 97 working classification of renal allograft pathology. *Kidney Int* **55**, 713–723 (1999).
25. van Buuren. *Flexible Imputation of Missing Data, Second Edition*. (CRC Press, 2018).
26. Vickers, A. J. & Elkin, E. B. Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak.* **26**, 565–574 (2006).
27. Bakdash, J. Z. & Marusich, L. R. *Rmcorr: Repeated Measures Correlation*. (2023).
28. Harrell, F. E. J. *Rms: Regression Modeling Strategies*. (2022).
29. Buuren, van, S., Groothuis-Oudshoorn & mice:, K. Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
30. Therneau, T. M. *A Package for Survival Analysis in R*. (2023).
31. Sjoberg, D. D. *Dcurves: Decision Curve Analysis for Model Evaluation*. (2022).
32. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
33. Kassambara, A., Kosinski, M. & Biecek, P. *Survminer: Drawing Survival Curves Using 'Ggplot2'*. (2021).
34. Scheinin, I. et al. *Ggforestplot: Forestplots of Measures of Effects and Their Confidence Intervals*. (2023).
35. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
36. Waskom, M. L. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
37. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with python. in *9th Python in Science Conference* (2010).

Acknowledgements

K.W., A.P., and M.C. hold a The Research Foundation Flanders (FWO) fellowship grant (11P1524N, 1S93023N and 12D6423N respectively), E.V.L. and J.C. held fellowship grants (1143919 N and 1196119 N, respectively) from FWO. M.N. is a senior clinical investigator of FWO (1844024 N). C.B., C.R. and M.W. are supported by the NIHR Imperial Biomedical Research Centre (BRC). This study is supported by the FWO with a project grant (G038024N) and by a grant from the KU Leuven Research Council (C2M/24/057).

Author contributions

TV and M.N. designed the study. A.P., DK, EC, E.V.L., J.C., K.W., M.C., M.N., M.P.E., P.K., T.V., T.Van., were involved in data collection and curation. A.C., A.dV., A.L., C.B., C.R., F.vS.S., G.A.B., G.D., J.K., M.C., M.W., O.A., O.T., S.F., S.S., S.V.S., T.M. provided the validation data. T.V. performed all analysis with input from M.N. C.R., M.H., O.T., A.L., L.C. and P.H., provided supervision, guidance, and critical review of the manuscript. T.V. and M.N. wrote the manuscript. All co-authors revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-65153-9.

Correspondence and requests for materials should be addressed to Maarten Naesens.

Peer review information *Nature Communications* thanks [Zoltan Laszik and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. [A peer review file is available.]

[1]Nephrology and Renal Transplantation Research Group, Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium. [2]Department of Imaging and Pathology, Translational Cell and Tissue Research, KU Leuven, Leuven, Belgium. [3]Department of Nephrology and Renal Transplantation, University Hospitals Leuven, Leuven, Belgium. [4]Université Paris Cité, Inserm UMR1151, Necker Enfants Malades Institute (INEM), Paris, France. [5]Department of Kidney and Metabolic Diseases, Transplantation and Clinical Immunology, Necker Hospital, Assistance Publique - Hôpitaux de Paris, Université Paris Cité, Paris, France. [6]Department of Immunology and Inflammation, Imperial College, London, United Kingdom. [7]Leuven Biostatistics and Statistical Bioinformatics Centre, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium. [8]Department of Pathology & Laboratory Medicine and Department of Medicine at the University of Calgary, Calgary, Alberta, Canada. [9]Division of Nephrology, Department of Medicine, Leiden University Medical Center, Leiden, the Netherlands. [10]Leiden Transplant Center, Leiden University Medical Center, Leiden, the Netherlands. [11]Université Paris Cité, INSERM U970 PARCC, Paris Institute for Transplantation and Organ Regeneration, Paris, France. [12]Histocompatibility and Immunogenetics Laboratory, Belgian Red Cross-Flanders, Mechelen, Belgium. [13]Amsterdam UMC, University of Amsterdam, Dept. of Pathology, Amsterdam, The Netherlands. [14]Amsterdam Institute for Infection and Immunology, Amsterdam, The Netherlands. [15]Department of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA. [16]Department of Medicine, University of Alberta, Edmonton, AB, Canada. [17]Department of Pathology, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands. [18]Department of Pathology, Leiden Transplant Center, Leiden University Medical Center, Leiden, Netherlands. [19]Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, Netherlands. [20]Division of Nephrology and Hypertension, Weill Cornell Medicine, New York, NY, USA. [21]Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA. [22]CIRI, INSERM U1111, Université Claude Bernard Lyon I, CNRS UMR5308, Ecole Normale Supérieure de Lyon, Univ. Lyon, Lyon, France. [23]Department of Transplantation, Nephrology, and Clinical Immunology, Hospices Civils de Lyon, Edouard Herriot Hospital, Lyon, France. [24]Division of Nephrology and Dialysis, Department of Medicine III, Medical University of Vienna, Vienna, Austria. [25]Department of Nephrology and Hypertension, University Hospital Schleswig-Holstein—Campus Kiel, Kiel, Germany. [26]Thomas E. Starzl Transplantation Institute, University of Pittsburgh, Pittsburgh, PA, USA.
✉e-mail: maarten.naesens@kuleuven.be