OPEN

Stratification of Kidney Transplant Recipients Into Five Subgroups Based on Temporal Disease Trajectories

Isabella F. Jørgensen, PhD,[1] Victorine P. Muse, PhD,[1] Alejandro Aguayo-Orozco, PhD,[1] Søren Brunak, PhD,[1] and Søren S. Sørensen, MD, DMsc[2]

Background. Kidney transplantation is the treatment of choice for patients with end-stage renal disease. Considerable clinical research has focused on improving graft survival and an increasing number of kidney recipients die with a functioning graft. There is a need to improve patient survival and to better understand the individualized risk of comorbidities and complications. Here, we developed a method to stratify recipients into similar subgroups based on previous comorbidities and subsequently identify complications and for a subpopulation, laboratory test values associated with survival. **Methods.** First, we identified significant disease patterns based on all hospital diagnoses from the Danish National Patient Registry for 5752 kidney transplant recipients from 1977 to 2018. Using hierarchical clustering, these longitudinal patterns of diseases segregate into 3 main clusters of glomerulonephritis, hypertension, and diabetes. As some recipients are diagnosed with diseases from >1 cluster, recipients are further stratified into 5 more fine-grained trajectory subgroups for which survival, stratified complication patterns as well as laboratory test values are analyzed. **Results.** The study replicated known associations indicating that diabetes and low levels of albumin are associated with worse survival when investigating all recipients. However, stratification of recipients by trajectory subgroup showed additional associations. For recipients with glomerulonephritis, higher levels of basophils are significantly associated with poor survival, and these patients are more often diagnosed with bacterial infections. Additional associations were also found. **Conclusions.** This study demonstrates that disease trajectories can confirm known comorbidities and furthermore stratify kidney transplant recipients into clinical subgroups in which we can characterize stratified risk factors. We hope to motivate future studies to stratify recipients into more fine-grained, homogenous subgroups to better discover associations relevant for the individual patient and thereby enable more personalized disease-management and improve long-term outcomes and survival.

(*Transplantation Direct* 2024;10: e1576; doi: 10.1097/TXD.0000000000001576.)

Kidney transplantation improves and prolongs lives of patients with end-stage renal disease compared with other treatments such a dialysis.[1] However, the expected remaining lifetimes of kidney transplant recipients is only around half of the general population dependent on the age at transplantation.[2] In the past decades, clinical research has succeeded in improving graft survival, and therefore, more patients with kidney transplantations die with a functioning graft.[3] Thus, it is important to stratify recipients to identify individualized risk of comorbidities and complications to improve patient survival.[3-5]

Received 13 June 2023. Revision received 2 November 2023.

Accepted 28 November 2023.

[1] Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen N, Denmark.

[2] Department of Nephrology, Rigshospitalet, Copenhagen University Hospital, Copenhagen Ø, Denmark.

I.F.J., S.B., and S.S.S. contributed to designing the study. I.F.J. created disease trajectories, clustering and complication analysis. V.P.M. analyzed the laboratory values and created linear models. A.A.O., S.B., and S.S.S. contributed to the decision-making of the analysis. Furthermore, S.S.S. contributed with clinical expertise. I.F.J. drafted the article and figures, which was edited and approved by all authors.

S.B. has ownerships in Intomics A/S, Hoba Therapeutics Aps, Novo Nordisk A/S, Lundbeck A/S, and ALK Abello A/S and managing board memberships in Proscion A/S and Intomics A/S. The other authors declare no conflicts of interest.

This work was supported by the Novo Nordisk Foundation for providing funding (grant agreement NNF14CC0001) and the Danish Innovation Foundation (BigTempHealth grant no. 5153-00002B) for the study.

The data that support the findings of this study are not publicly available as they contain person sensitive information. To obtain access to data, the study needs to be approved by the Danish Data Protection Agency (www.datatilsynet.dk). All studies should be conducted in compliance with The Danish Act on Processing of Personal Data and all other applicable laws and regulations.

Supplemental digital content (SDC) is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.transplantationdirect.com).

Correspondence: Søren Brunak, PhD, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen N, Denmark; (soren.brunak@cpr.ku.dk); and Søren S. Sørensen, MD, DMsc, Department of Nephrology, Rigshospitalet, Copenhagen University Hospital, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark. (soeren.schwartz.soerensen@regionh.dk).

ISSN: 2373-8731

DOI: 10.1097/TXD.0000000000001576

Several patient characteristics of both donor and recipient are known to influence survival and other long-term outcomes for transplant recipients.[6] High donor age and body weight as well as receiving a kidney from a deceased donor compared with a living donor are associated with poor transplant outcomes.[7,8] Multiple characteristics of the recipient, such as age, recurrence of native kidney disease, HLA compatibility, anti-HLA immunization, ethnic background, and time spent on dialysis before transplantation as well as comorbidities, especially cardiovascular diseases, have shown to impact outcomes of transplantation.[6,9]

The age of transplant recipients is increasing and with this trend also the burden of comorbidities both pre- and posttransplantation.[10-12] Studies show that even patients with increasing age and high comorbidity burden have survival benefit from kidney transplantations, and thus, this group of recipients keeps growing.[13-15] Several studies have also investigated how comorbidities impact survival of kidney transplant recipients. Another study found an increased risk of serious infections in patients with diabetes compared with patients without diabetes, but the 1-y survival was the same for both groups.[16] Conversely, other studies have found poorer long-term survival for recipients with both pre-existing and posttransplant diabetes compared with recipients without diabetes.[17-19] However, there is no evidence of improvement in survival, illustrating the need for new, more personalized strategies to reduce mortality for kidney recipients.[5,20]

Additionally, several studies have reported associations between pretransplant laboratory values and mortality. Two of these found higher levels of pre- and posttransplant serum albumin to be associated with both better graft and patient survival.[21,22] Dahlberg et al[22] also found 1-y posttransplant serum creatinine to be associated with graft failure. Few studies have focused on investigating trajectories of kidney transplantation recipients based on estimated glomerular filtration rate (eGFR),[23-26] measured GFR (mGFR),[27] proteinuria,[28] or body mass index.[29] Moreover, the majority of these have solely focused on 1 or a few laboratory values and its impact on graft survival. To the best of our knowledge, no studies investigate longitudinal disease trajectories for kidney transplant recipients. Furthermore, data-driven approaches could reveal new associations and stratify high-risk subgroups moving toward precision medicine.

Using data from a 40-y period, we have performed a data-driven, longitudinal analysis of hospital diagnoses both before and after kidney transplantation to find statistically significant, temporal disease patterns for recipients. These patterns can be used to stratify patients into different disease progression subgroups and evaluate which variables have an impact on survival as well as the risk of complications in each of the subgroups, possibly enabling individualized follow-up and better monitoring of patients.

MATERIALS AND METHODS

Identifying Kidney Transplantation Recipients

Kidney transplantation recipients were identified in the Danish National Patient Registry (DNPR)[30] using its registration of surgical procedures. Since 1996, Denmark has used the Nordic Medico-Statistical Committee Classification of Surgical Procedures codes[31] to classify operations in the Danish hospital system. Kidney transplantations were identified using the code "Kidney transplantation and related surgery" (code: KKAS). Patients with the subcodes "Allogeneic kidney transplantation with kidney from [deceased donor]" (code: KKAS10) and "Allogeneic kidney transplantation with kidney from living donor" (code: KKAS20) were selected, whereas patients with the subcode "Autologous kidney transplantation" (code: KKAS00) were excluded. Codes are time-stamped at the time of kidney transplantation. Before 1996, Denmark recorded operations using an "operation and treatment classification" system, and here the code "Transplantatio renis (homotransplantation)" (code: 57480) was used to identify kidney transplant recipients and define the time of transplantation. Patients with the code "Transplantio renis (autotransplantation)" (code: 57490) were excluded. We further excluded patients who also had other organs transplanted both pre- and postkidney transplantation as these are relatively few, and, in some respects very different (see **Figure S1A, SDC**, http://links.lww.com/TXD/A606 for the flowchart).

Since 1994, the DNPR has used the International Classification of Diseases 10th revision (ICD-10) terminology to record diagnoses. From 1977 to 1993, the eighth revision (ICD-8) was used. All ICD-8 diagnosis codes were translated to ICD-10 diagnosis codes using a complete mapping between the 2 versions.[32] In the analysis, we merged 2 ICD codes, "Chronic kidney disease" (code: N18) and "Unspecified kidney failure" (code: N19) into 1 diagnosis termed "Kidney failure" (code: N24). Additionally, a new code "Kidney transplantation" (code: N55) was introduced on the date of the kidney transplantation. When a patient received multiple kidney transplants over time, we use the first transplantation to calculate follow-up time, whereas additional transplantations were included as an outcome in Table 1.

Creating Temporal Disease Trajectories

After identifying all kidney transplantation recipients using the criteria mentioned earlier, we extracted diagnosis information from the DNPR to create disease trajectories using a previously published method (see **Figure S1B, SDC**, http://links.lww.com/TXD/A606 for visual explanation).[33] For each patient and each diagnosis code (using the categorical level, the first 3 characters, eg, N18), only the first diagnosis was included. Initially, all diseases both pre- and postkidney transplantation are identified and tested to establish whether they co-occur significantly more together in the population of recipients than expected based on their individual frequencies. To quantify the strength of the correlation the relative risk (RR) was used for all disease co-occurrences (*D1* and *D2*). The number of patients with both *D1* and *D2* were calculated ($C_{exposed}$) and comparison groups of N = 10 000 randomly selected patients were matched to patients with *D1* on birth decade, sex, type of hospital encounter, and discharge week to avoid bias from seasonality. Occurrences of *D2* were calculated in the kidney transplantation group ($C_{exposed}$) and in the matched comparison groups ($C_1 \ldots C_N$), and the RR can be defined as:

$$RR = \frac{C_{exposed}}{\frac{1}{N}\sum_i C_i} \quad (1)$$

A *P* value for the RR was calculated and adjusted for multiple testing using the Bonferroni approach. All disease pairs (*D1* and *D2*) with an RR >1 and a significant

TABLE 1.

Characteristics of the 5 trajectory subgroups of kidney transplanted patients

	Group 1	Group 2	Group 3	Group 4	Group 5
Index diagnosis toward transplantation	Glomerulonephritis	Hypertension	Diabetes	Hypertension and glomerulonephritis	Diabetes and hypertension
Number of patients	543	1270	293	574	454
Males, n (%)	334 (62)	780 (61)	193 (66)	383 (67)	314 (69)
Age at first transplantation operation code, mean (SD)	38.43 (14.98)	48.88 (13.99)	45.83 (10.50)	42.43 (13.99)	47.58 (11.48)
Follow-up time/observational time from first transplantation operation code, mean (SD)	13.55 (8.51)	10.49 (8.44)	7.99 (6.17)	14.44 (9.31)	8.43 (6.54)
Number of patients that died, n (%)	170 (31)	597 (47)	151 (52)	270 (47)	239 (53)
Age at death, mean (SD)	57.50 (13.43)	60.44 (11.55)	52.52 (9.34)	57.72 (11.79)	55.04 (11.30)
Age at first diagnosis in registry, mean (SD)	23.15 (16.25)	33.43 (15.09)	24.77 (12.05)	28.94 (14.21)	26.59 (13.19)
Follow-up time since first diagnosis in registry, mean (SD)	28.84 (8.53)	25.94 (10.60)	29.05 (9.47)	27.93 (9.34)	29.43 (9.14)
Have had additional kidney transplantation operations, n (%)	126 (23)	214 (17)	26 (9)	135 (29)	38 (8)
Number of patients with decease donor, n (%)	224 (41)	528 (42)	146 (50)	203 (35)	234 (52)
Number of patients with living donor, n (%)	184 (34)	247 (19)	83 (28)	96 (17)	110 (24)
Charlson weighted comorbidity score, mean (SD)	3.97 (2.47)	4.19 (2.34)	6.48 (2.05)	4.53 (2.51)	6.99 (1.94)

For *P* values and false discovery rate corrected *P* values for difference between the 5 trajectory subgroups, see Table S1 (**SDC**, http://links.lww.com/TXD/A607).

P value were tested for directionality using a binomial test. The binomial test identifies disease pairs where the first diagnosis of *D1* occurs significantly more before the first diagnosis of *D2*, or vice versa. A direction was considered significant when the Bonferroni corrected *P* value of <0.05. Statistically significant directional disease pairs were then merged into longer trajectories with overlapping diagnoses. A set of linear disease trajectories can be visualized as a disease progression network that summarizes alternative routes in the entire population (**Figure S1B, SDC**, http://links.lww.com/TXD/A606). For additional details and figures on the method, see the original article[33] as well as a recent update.[34] As the DNPR contains 7.2 million patients, using a control group of 10 000 patients for each disease pair strongly reduces or eliminates temporal seasonal variation in disease progression patterns, as well as many other potential confounding factors.[33,34]

Clustering of Kidney Transplantation Trajectories

To discover homogeneous clusters of disease trajectories, we first calculated the Jaccard similarity between all disease trajectories as shown in the following equation. Thus, the similarity between trajectories *A* and *B* was calculated as the number of diagnoses *A* and *B* have in common divided by the union,

$$J(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}. \quad (2)$$

Subsequently, we used hierarchical clustering with Euclidian distance and the ward.D measure to cluster the linear disease trajectories. The R package "NbClust"[35] was used to find the optimal number of clusters. For each of the clusters identified, we performed yet another clustering to discover subclusters (see **Material, SDC**, http://links.lww.com/TXD/A606, for details). As some recipients follow disease trajectories from >1 cluster, they are further stratified into more fine-grained trajectory subgroups for which outcomes are evaluated.

Outcomes Evaluated

To compare mortality in each of the different trajectory subgroups of kidney transplantation recipients identified, Poisson models were used. The Poisson model examines the expected number of deaths per unit of observed time, and thus, populations of different size, observed for different length of time, can be compared.[36] The time at risk was calculated from January 1, 1977, when the DNPR started or when patients were born until the end of data, April 10, 2018, or death, whichever came first. Person-years (PYs) were calculated for each stratum as the time from the first kidney transplantation to the end of follow-up or death. The number of deaths was modeled as a function of the trajectory subgroup, age at transplantation, and sex using the time at risk and PYs as offsets. All-cause mortality rates per 1000 PY were used to visualize mortality.

To evaluate the overall comorbidity burden of kidney recipients in each subgroup, the Charlson comorbidity index[37] was used.

Laboratory Values

The Clinical Laboratory Information System (LABKA)[38] and the B-Data Clinical Chemistry Laboratory System (BCC) database record results from routine tests performed at Danish hospital laboratories. We used LABKA and BCC data from 2 regions in Denmark (Capital Region of Denmark and Region Zealand, respectively) covering around half the Danish population from the period 2009 to 2016. Thus, the laboratory test values were only available for a subpopulation in the population-wide registry data. The biochemical databases were cleaned systematically for completeness and suitability.[39] A series of lookup dictionaries for identifying incomplete tests or qualitative results was used to separate nonquantitative data from the remainder of the data sets. Additional cleaning was conducted to conform numerical results to the same format, for example, by aligning all decimal separators. Furthermore, character signs, such as "=," or typos, such as extra spaces and decimal separators, were removed to allow for smooth quantitative processing of data.

We identified patients with kidney transplantations in the LABKA and BCC data and filtered the cohort to only include patients with laboratory values available within 100 d before transplantation to capture the patient's biochemical heath profile more accurately at the time of surgery. We ranked the laboratory tests according to frequency for the entire cohort. Missing values were imputed using missForest[40] for each of the trajectory subgroups identified.

Linear Laboratory Value Models

Laboratory values were tested for significant correlation with time to death (<4 y after transplantation or >4 y after transplantation), using linear models controlling for age at time of transplantation. Linear models were made for the entire cohort as well as for each of the trajectory subgroups identified. All *P* values were corrected for multiple testing using the false discovery rate (FDR).

Statistical Tests

To discover significantly different characteristics between the trajectory subgroups of kidney transplant recipients, a Fisher test or a nonparametric Wilcoxon–Mann-Whitney test was used depending on the type of variable. Two-sided FDR corrected *P* values of <0.01 were considered statistically significant. Follow-up time was calculated for each of the trajectory subgroups from the date of kidney transplantation until the end of data, April 10, 2018, or death, whichever came first. To find differences in the occurrence of complications after transplantation, the total number of patients diagnosed with each complication after kidney transplantation was calculated. To find significant differences between the groups, a Fisher test with an FDR corrected *p* value of <0.01 was considered significant.

All statistical analyses were performed in RStudio version 1.1.383.

Data and Materials Approval

This study was approved by The Danish Data Protection Agency (ref: 514-0255/18-3000, 514-0254/18-3000, SUND-2016-50), The Danish Health Data Authority (ref: FSEID-00003724 and FSEID-00003092) and The Danish Patient Safety Authority (3-3013-1731/1/).

RESULTS

Demographics

In total, 5896 kidney transplant recipients were identified in the DNPR (2277 from 1977 to 1995 and 3619 from 1996 to 2018, respectively) with 7152 transplantations in all. This number is similar to kidney transplantation recipients identified in the joint Nordic Scandiatransplant database (http://www.scandiatransplant.org/; Figure 1A). In the period from 1977 to 1995, the large majority of patients had homotransplantations (code: 57480), whereas only a few had autotransplantations (code: 57490; Figure 1B). The same pattern was observed from 1996 to 2018 where the homotransplantations or allogeneic transplantations dominate. Allogenic transplants from 1996 and onward split into kidneys from deceased and living donors. Most donations are from deceased donors (code: KKAS10) as opposed to transplantations from living donors (code: KKAS20). A minority are autologous kidney transplants (code: KKAS00). Patients with autologous kidney transplantations (code: 57490 or KKAS00) were excluded (81 patients) from the analysis as well as 63 patients who also had other organs transplanted, leaving 5752 kidney transplant recipients (**Figure S1, SDC**, http://links.lww.com/TXD/A606). We investigated at which hospital patients had their kidney transplantation performed (**Figure S2, SDC**, http://links.lww.com/TXD/A606).

Disease Trajectories Reveal High Heterogeneity and Disease Burden

Using a previously published method,[33] we identified significant, temporal disease trajectories for kidney transplantation patients (see Methods). Sets of linear disease progression trajectories were summarized in a network showing alternative routes pre- and postkidney transplantation (Figure 2). In Figure 2, the network consists of disease trajectories made with 4 consecutive diseases where a minimum of 5% (287 patients) are diagnosed with all 4 diseases in the order specified by the trajectory. Here, 13 different diseases occur before kidney transplantation with ticker edges indicating that more patients follow the specified path. After kidney transplantation 24 diseases appear with fewer patients following each path. Lowering the minimum number of patients to 50 (instead of 287) reveals more diverse disease progression patterns, more precisely 702 different disease trajectories for kidney transplantation recipients that are too many to visualize effectively in a network. This highlights the heterogeneity and substantial disease burden for this group of patients in relation to both the severe disease history before transplantation but also the many complications that may present posttransplantation.

Stratification of Kidney Transplant Recipients Into 5 Major Subgroups

To illustrate similarities and discover clusters across all 702 disease trajectories, we used a Jaccard similarity score and hierarchical clustering and found 3 major clusters of frequent disease trajectories of kidney transplantation recipients (**Figure S3, SDC**, http://links.lww.com/TXD/A606). The 3 major clusters are in particular driven by nephritic syndrome (codes: N00 or N03) hereafter referred to as glomerulonephritis, hypertension (code: I10), and diabetes (codes: E10, E11, or E14). For each of the major clusters, we have identified further clinical subclusters (see **Results and Figures S4–S7, SDC**, http://links.lww.com/TXD/A606 for details). The majority of the kidney transplantation recipients follow at least 1 of the major trajectories in the identified clusters (Figure 3). From Figure 3, we identified 5 trajectory subgroups: group 1 consisting of 543 patients following a trajectory with glomerulonephritis only, group 2 consisting of 1270 patients only following a hypertension trajectory, and group 3 consisting of 293 patients following diabetes trajectories only. Furthermore, we defined 2 groups of patients following more than 1 type of trajectory: group 4 consisting of 574 patients following trajectories with both hypertension and glomerulonephritis and group 5 with 454 patients following trajectories with both diabetes and hypertension. The remaining 2 groups consisting of 28 and 14 patients, respectively, had too few patients for them to be included in subsequent analyses.

Characteristics of the 5 Clinical Subgroups

Comparison of the 5 trajectory subgroups shows that patients which follow a disease trajectory starting with

FIGURE 1. Kidney transplantations in DNPR. A, Comparison of kidney transplantations identified by the operation codes in the DNPR and the Scandiatransplant database, which registers all transplantations in Finland, Sweden, Denmark, Norway, Iceland, and Estonia from 1969 and onward (http://www.scandiatransplant.org/). The data from DNPR used here end in April 2018. We identified 5896 unique patients with 7152 transplantations. B, The different kidney transplantation operation codes given according to discharge year. Only combinations of code and year with 5 or more patients are included in the figure because of disclosure restrictions in the data permissions. DNPR, Danish National Patient Registry.

glomerulonephritis are the youngest at kidney transplantation (mean age 38.4 y), whereas patients following a trajectory starting with hypertension, or, hypertension and diabetes are the oldest at transplantation (mean age 48.9 and 47.6 y, respectively) (Table 1; **Table S1, SDC**, http://links.lww.com/TXD/A607). Furthermore, patients following trajectories with diabetes die earlier than in the other subgroups (mean age 52.5 and 55.0 y compared with 60.4, 57.5, and 57.7 y for hypertension, glomerulonephritis, and hypertension and glomerulonephritis, respectively). Moreover, patients following

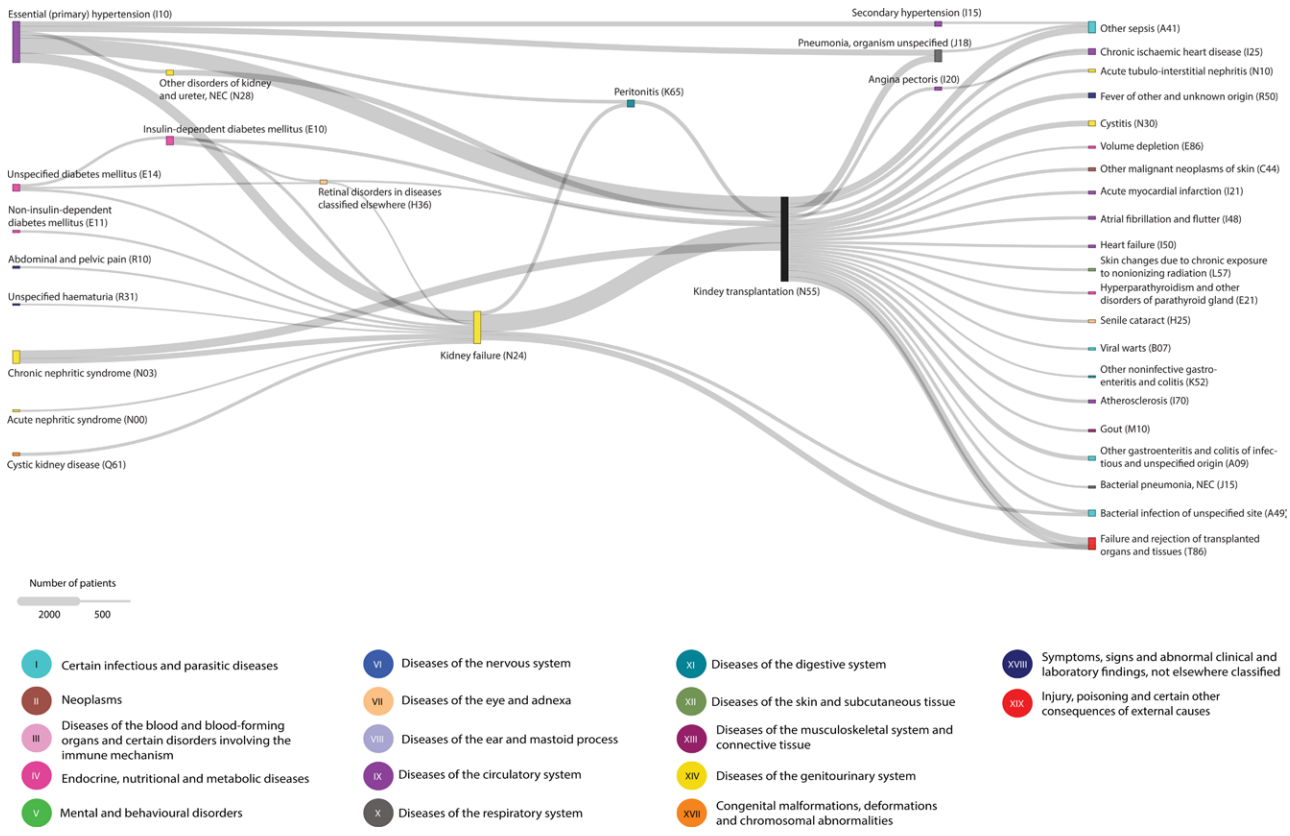

FIGURE 2. Temporal disease trajectory network. Significant temporal disease trajectories were made for the 5752 kidney transplantation recipients. Diabetes-, hypertension-, and kidney-related diagnoses manifest with significant direction before transplantation, whereas many different complications appear after kidney transplantation. The trajectory network is based on trajectories of 4 consecutive diseases that a minimum of 5% (287 patients) follow-up. The thickness of the arrows between diagnoses represents the number of patients following the disease path in the network. Diagnoses are color coded according to ICD-10-chapter legend included below. ICD-10, International Classification of Diseases 10th revision.

trajectories with diabetes also show a higher Charlson comorbidity index (mean score 6.5 and 7.0 for the diabetes groups compared with 4.2, 4.0 and 4.5 for the groups not following the diabetes trajectories, respectively; Table 1; **Table S1, SDC**, http://links.lww.com/TXD/A607).

The increased mortality of patients with a pretransplant diagnosis of diabetes was also observed using a Poisson model (Table 2; Figure 4). Following disease trajectories starting with diabetes and hypertension or only diabetes corresponds to relative risks of 1.47 and 1.67, respectively (Table 2). Furthermore, an increasing age at transplantation is associated with higher risk (RR = 1.02).

Stratified Risk of Complications

To determine whether particular subgroups are at higher risk of some complications, we visualized the disease trajectories from each of the 3 clusters (**Figures S8–S10, SDC**, http://links.lww.com/TXD/A606). The first glomerulonephritis diagnosis typically occurs around the age of 30 y, whereas the kidney transplantation occurs between age 30 and 50 y (**Figure S8A, SDC**, http://links.lww.com/TXD/A606). The first hypertension diagnosis appears around the age of 35 y (**Figure S9A, SDC**, http://links.lww.com/TXD/A606), whereas the first diabetes diagnosis typically occurs between the ages of 20 and 30 y (**Figure S10A, SDC**, http://links.lww.com/TXD/A606). The time from each of the 3 diagnoses to kidney transplantation is generally longer for the diabetes patients compared with glomerulonephritis or hypertension patients. Furthermore, we performed a frequency analysis and evaluated statistical differences between the subgroups with a Fisher test (**Table S2, SDC**, http://links.lww.com/TXD/A608). The diabetes subgroups are less often diagnosed with "Gout" (code: M10) compared with the other subgroups. Furthermore, there is a tendency that diabetes patients are diagnosed less often with "Other malignant neoplasms of skin" (code: C44), but more often with "Senile cataract" (code: H25) and "Atherosclerosis" (code: I70). The hypertension patients are more often diagnosed with diseases from the cardiovascular chapter such as "Angina pectoris" (code: I20), "Acute myocardial infarction" (code: I21), "Chronic ischemic heart disease" (code: I25), "Atrial fibrillation and flutter" (code: I48), and "Heart failure" (code: I50). The patients with glomerulonephritis are more often diagnosed with infections such as "Other gastroenteritis and colitis of infectious and unspecified origin" (code: A09) and "Viral warts" (code: B07). The group of patients with both glomerulonephritis and hypertension has a particularly high frequency (38%) of "Failure and rejection of transplanted organs and tissues" (code: T86). For overview of all comorbidities with increased risk for each subgroup, see Figure S11 (**SDC**, http://links.lww.com/TXD/A606).

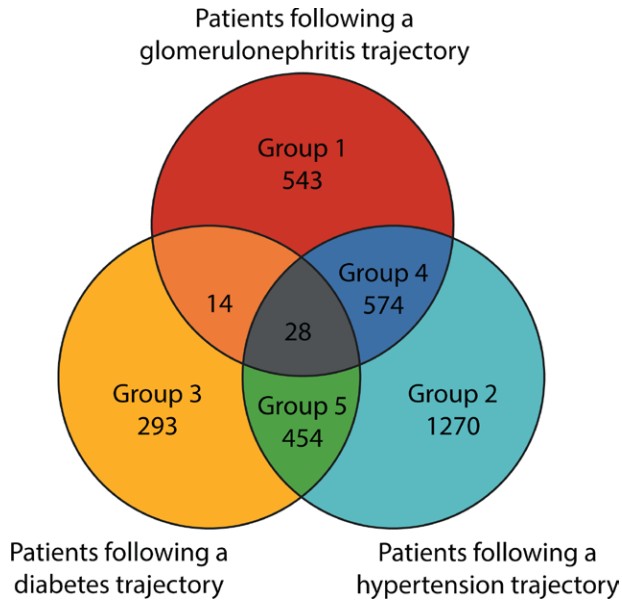

FIGURE 3. Venn diagram of patients in the 3 clusters. Some patients follow trajectories from >1 cluster. Thus, we identified 5 trajectory subgroups with a reasonable number of patients. Trajectory subgroup 1 with patients only following trajectories starting with glomerulonephritis (543 patients); trajectory subgroup 2 with patients only following trajectories starting with hypertension (1270 patients); trajectory subgroup 3 with patients only following trajectories starting with diabetes (293 patients); trajectory subgroup 4 with patients following trajectories starting with both hypertension and glomerulonephritis (574 patients); and trajectory subgroup 5 with patients following trajectories starting with both diabetes and hypertension (454 patients). The remaining 2 groups consisting of 28 and 14 patients, respectively, had too few patients for them to be included in subsequent analyses.

Laboratory Test Values for a Subset of Recipients

In total, 522 kidney transplant recipients had laboratory value results within 100 d before kidney transplantation in 1 of the 2 Danish regions included in the laboratory test databases. The top 24 laboratory test values and the number of patients that have had at least 1 test result within 100 d before transplantation are shown in Table S3 (**SDC**, http://links.lww.com/TXD/A606). All 24 tests had reported values for at least 485 (93%) patients. The a few missing laboratory values were imputed using missForest as described in Methods section.

Creating linear models for each of the 24 different laboratory tests for all 522 patients showed the trend that albumin and CRP levels have an impact on survival (FDR adjusted *P* values: 0.082 and 0.082, respectively) with higher albumin levels and lower CRP before transplantation being associated to better survival (Figure 5; **Table S4, SDC**, http://links.lww.com/TXD/A609). Similar trends were seen for the 288 patients following any of the trajectories with a significant impact from CRP and the same trend for albumin (FDR adjusted *P* values: 0.018 and 0.085, respectively). Stratifying patients into the 5 trajectory subgroups showed that lower levels of basophils before transplantation is significantly associated with a better survival (FDR adjusted *P* value: 5.09e-04) for patients in the glomerulonephritis subgroup (Figure 5; **Table S4, SDC**, http://links.lww.com/TXD/A609). Moreover, lower levels of leukocytes before transplantation are significantly associated with better survival (FDR adjusted *P* value: 0.016) for patients in the diabetes subgroup. In total, 234 kidney transplantation recipients do not follow any of the trajectories and for this group, higher levels of glucose showed a significant association with poor survival (FDR adjusted *P* value: 0.044). For overview of all laboratory values associated with survival for each subgroup, see Figure S11 (**SDC**, http://links.lww.com/TXD/A606).

DISCUSSION

This is the first study to investigate disease-agnostic trajectory patterns both pre- and postkidney transplantation. The systematic, data-driven method identified 3 clusters of longitudinal disease progression patterns leading up to kidney transplantation including glomerulonephritis, diabetes, and hypertension, which confirm existing knowledge. As patients obviously can follow several trajectories >3 frequent longitudinal patterns may exist overall. Only a few patients (0.5%) follow disease trajectories from all 3 clusters indicating that the 3 clusters represent relatively different phenotypes with little overlap. More than 700 different disease trajectories were identified, which highlights the significant diversity within this patient group, with a severe disease history before transplantation and the numerous potential complications that may arise subsequently.

An association between low levels of albumin before kidney transplantation and poor survival outcomes has previously been reported in the literature.[21,22,41,42] We confirm this trend when investigating our subpopulation of kidney transplantation recipients. However, when stratifying patients into the 5 identified disease progression subgroups, we observed

TABLE 2.

Poisson model for survival after first kidney transplantation

	Poisson model	
	Relative risk (95% CI)	***P***
Intercept	2.57e-07 (2.00e-07–3.38e-07)	<2.2e-16*
Sex (male)	0.98 (0.87–1.09)	0.659
Trajectory subgroup (hypertension)	1.12 (0.94–1.33)	0.211
Trajectory subgroup (hypertension + glomerulonephritis)	1.02 (0.84–1.23)	0.873
Trajectory subgroup (diabetes)	1.67 (1.34–2.08)	4.28e-06*
Trajectory subgroup (diabetes + hypertension)	1.47 (1.21–1.79)	1.29e-04*
Age	1.02 (1.02–1.03)	<2.2e-16*

A Poisson model was fitted to estimate differences in survival after kidney transplantation for the 5 trajectory subgroups of kidney transplantation recipients. The Poisson model includes age at the kidney transplantation split into 5 age groups, sex and which trajectory subgroup patients belong to.
*The reference groups are patients that follow a trajectory starting with glomerulonephritis and females. *P* values <0.05 were considered statistically significant.

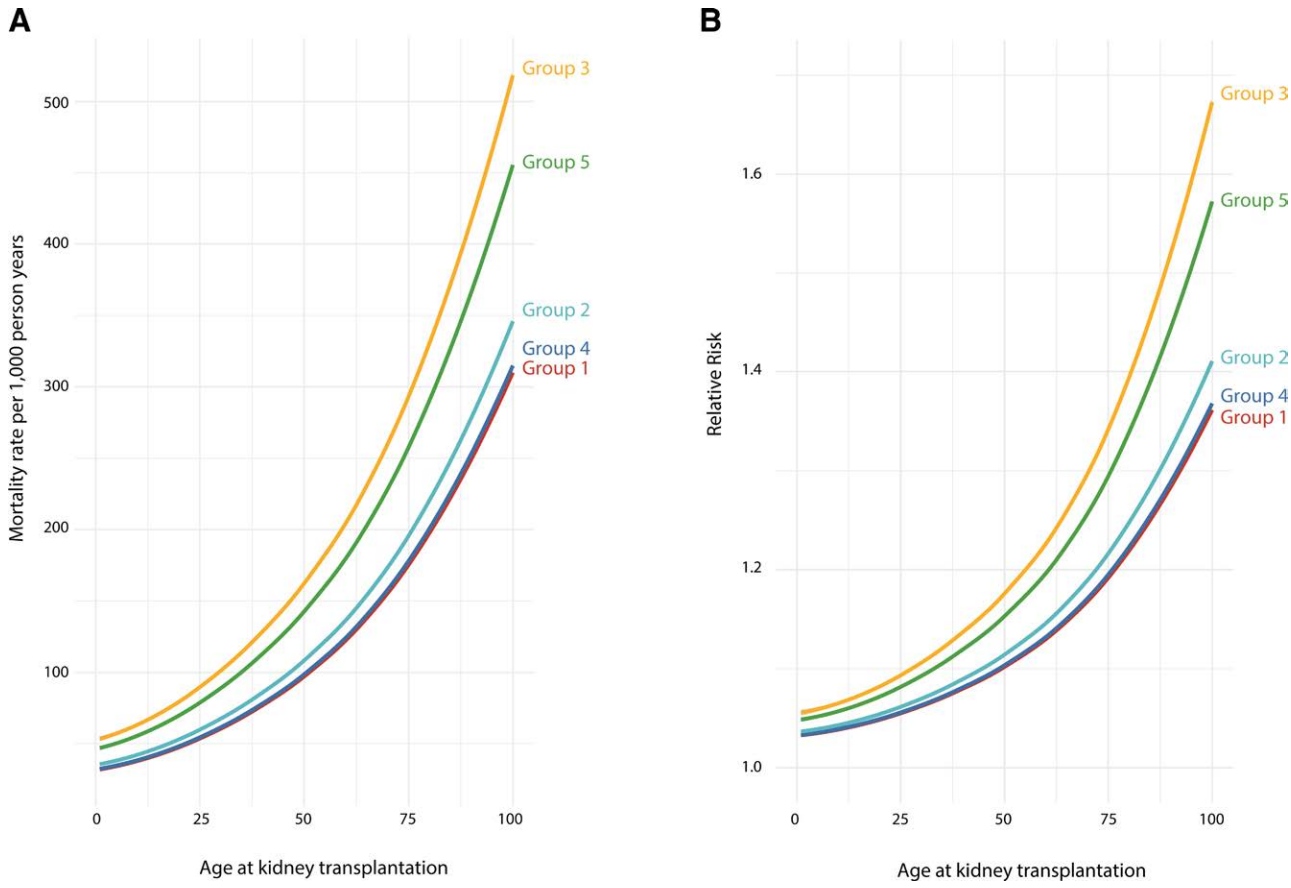

FIGURE 4. Poisson model-based predictions of mortality. Number of deaths were modeled using a Poisson model including the 5 trajectory subgroups, age at transplantation and sex and using the time at risk and PY as offsets. A, All-cause mortality rates per 1000 PY as a function of age at kidney transplantation for each trajectory subgroup. B, Relative risk of death as a function of age at kidney transplantation for each trajectory subgroup. The 2 groups diagnosed with diabetes (groups 3 and 5) have highest mortality rates and relative risks. Group 1 = glomerulonephritis (543 patients), group 2 = hypertension (1270 patients), group 3 = diabetes (293 patients), group 4 = hypertension and glomerulonephritis (574 patients), and group 5 = diabetes and hypertension (454 patients). For color coding of the 5 trajectory subgroups, see Figure 3. PY, person years.

that other laboratory values were more important for survival with high levels of basophils and leukocytes being associated with worse survival for the subgroup of glomerulonephritis and diabetes, respectively. This indicates the need to stratify patients into more homogeneous subgroups to discover more individualized survival patterns and subgroups with higher risks of certain complications.

Kidney transplant recipients diagnosed with diabetes have a significantly increased risk of dying earlier than patients without diabetes, which confirm findings from previous studies.[17-19] Additionally, higher levels of glucose showed an association with poor survival for the 234 kidney transplantation recipients that do not follow any of the trajectories, which could indicate underdiagnosis of prediabetes in the kidney transplantation recipients. Previous studies have identified underdiagnosed diabetes in candidates for kidney transplantations[43] and recommend screening for diabetes in patients on waiting lists for a new kidney for better risk-adjusted treatment and optimal care.[44,45]

Higher levels of leukocytes are significantly correlated with poor survival for patients with diabetes. High levels of leukocytes often indicate severe infection and inflammation, which could be one of the reasons behind the higher mortality seen in this group of patients. Interestingly, the diabetes subgroups are less often diagnosed with "Other malignant neoplasms of skin" (code: C44) than the other subgroups. However, this could also be because of the fact that the diabetes patients die earlier (mean age at death for the 2 trajectory subgroups with diabetes 52.5 and 55.0 y), whereas skin cancers typically appear around age 55–60 y (**Figures S8–S10, SDC**, http://links.lww.com/TXD/A606).

Higher blood levels of basophils are significantly correlated with poor survival for patients with glomerulonephritis. High basophil levels often indicate chronic inflammation and can contribute to the generation and maintenance of high levels of specific Ig autoantibodies, resulting in kidney damage.[46,47] Patients with glomerulonephritis were more often diagnosed with different types of bacterial infections ("Viral warts" and "Other gastroenteritis and colitis of infectious and unspecified origin") that could cause the increased level of basophils. A possible explanation could be that this group of patients are treated with immunosuppressive drugs before transplantation, but additional studies are needed to confirm this.

Cardiovascular disease is the primary reason for kidney transplantation recipient death.[4,14,48] We found that patients with pretransplant hypertension have increased risk of additional cardiovascular diseases including angina pectoris, chronic ischemic heart disease and heart failure, and therefore, this subgroup of patients might benefit from close monitoring of cardiovascular disease.

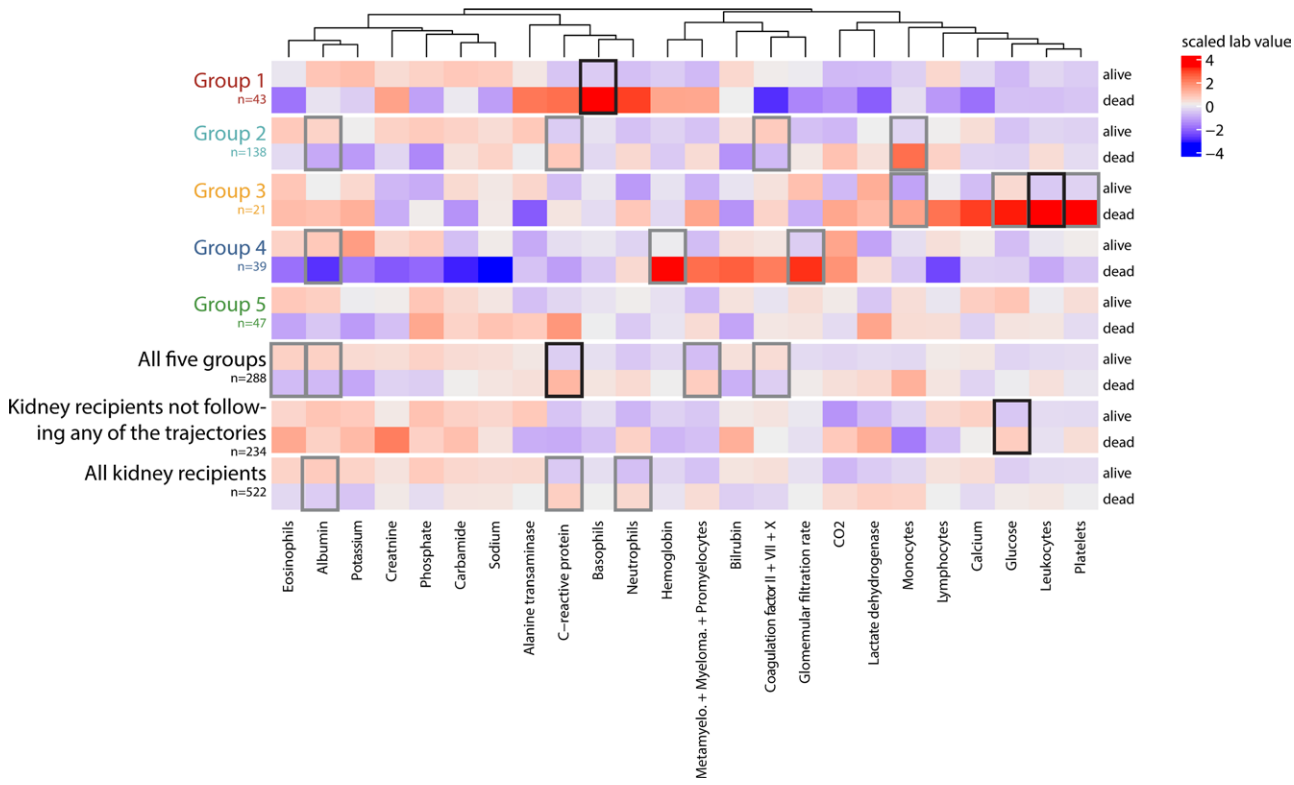

FIGURE 5. Heatmap of laboratory test value results. The heatmap compares scaled laboratory test value results for patient survival in each of the 5 trajectory subgroups. Laboratory values correlated with time to death (<4 y after transplantation or >4 y after transplantation) were calculated using linear models. Linear models were made for each of the 24 most frequent laboratory tests for all kidney transplant recipients in the database. Laboratory values marked with a gray square show significant difference before multiple correction. Black squares indicate significant differences between alive and deceased patients after multiple-testing FDR correction. Different laboratory values are relevant for patient survival for the 5 different trajectory subgroups. For color coding of the trajectory subgroups, see Figure 3. FDR, false discovery rate.

Other studies have investigated kidney transplantation trajectories for nondiagnosis-related data. However, they often investigate trajectories of 1 laboratory value only. For example, trajectories of eGFR have been found to associate with progression to end-stage kidney disease after transplantation[24] and proteinuria levels could to some degree predict graft failure.[28] Additionally, studies have investigated eGFR trajectories and their clinical implications but for patients with chronic kidney disease and not kidney transplantation recipients.[49-51] The majority of previous clinical research aim to improve short-term kidney transplantation outcomes and graft survival[23,28,29,52-54]; in complement we here focus on long-term outcomes as individualized risk of comorbidities, complications and patient survival. Disease progression and complications can be challenging to predict, hence stratifying patients into more fine-grained, homogenous subgroups that show higher risk of certain progression paths is 1 step toward an increased level of personalized medicine in this domain.

Strength and Limitations

This population-based, data-driven study enabled the possibility to include all comorbidities and laboratory values available for patients, and thus, fewer predefined selection criteria were needed. All patients with allogeneic kidney transplantations in Denmark from 1977 and onward are included in the study and all their hospital-recorded comorbidities, more than 40 y are included in the initial, longitudinal analysis. However, only a few basic characteristics such as age and sex were available for recipients and no features regarding the actual transplantation. Additionally, information about kidney donors is limited as only the donor status (alive or deceased) is available and only since 1996. As kidney transplantation recipients are often followed closely at hospitals both pre- and posttransplantation, comorbidity information and laboratory test results can be considered relatively complete. We utilized 2 diverse datasets, one comprising a population-wide health registry and a phenotypically deeper one based on electronic health records and laboratory test results for a smaller fraction of the patients. Electronic health records are rarely available for extended periods, and hence, fewer patients are available for the laboratory test analysis. Only 2% of the top 24 laboratory test values were missing. The few laboratory test values missing were imputed using missForest that has shown to be highly accurate and outcompete other techniques when imputing missing laboratory data with up to 30% missingness.[55] We have imputed laboratory test values for each trajectory subgroup as diseases such as diabetes and hypertension might influence the values and therefore could be important to take into account when imputing. We also imputed values using missForst for all groups simultaneously as input, which showed similar results. It has not been examined how accurate kidney transplant operation codes are in the DNPR. However, studies have investigated other operation codes and the positive predictive value here is typically between 90% and 100% depending on the type of operation.[30] Here, we show high agreement between kidney transplantations in DNPR and Scandiatransplant data, confirming the validity of the operation codes. The validity of ICD-10 diagnosis codes in the DNPR is continuously evaluated. A

comprehensive review show that the positive predictive values of diagnoses vary widely (<15%–100%) depending on the study setting, calendar year as well as nature of the disease.[30]

In conclusion, the method presented stratifies patients into clinically meaningful subgroups based on comorbidity profiles and subsequently analyses patterns of complications, laboratory test values and survival for each subgroup. This indicated that the comorbidities and disease patterns pretransplantation have an important impact on the risk of mortality and posttransplantation complications. Even though some of the clinical subgroups might be known, studies rarely stratify recipients into more homogenous subgroups. Furthermore, we identified potential undiagnosed diabetes in a subgroup of patients, highlighting the need for screening for diabetes before transplantation. Additionally, close monitoring of cardiovascular disease in patients with prehypertension is important. This study highlights that stratification of kidney transplant recipients is needed for individualized and optimal management, early detection, and prevention of comorbidities to improve long-term outcomes and survival.

REFERENCES

1. Suthanthiran M, Strom TB. Renal transplantation. *N Engl J Med*. 1994;331:365–376.
2. *ERA Registry: ERA Registry Annual Report 2020*. Amsterdam UMC, Location AMC, Department of Medical Informatics, Amsterdam, the Netherlands, 2022.
3. Howard RJ, Patton PR, Reed AI, et al. The changing causes of graft loss and death after kidney transplantation. *Transplantation*. 2002;73:1923–1928.
4. Awan AA, Niu J, Pan JS, et al. Trends in the causes of death among kidney transplant recipients in the United States (1996-2014). *Am J Nephrol*. 2018;48:472–481.
5. Lamb KE, Lodhi S, Meier-Kriesche HU. Long-term renal allograft survival in the United States: a critical reappraisal. *Am J Transplant*. 2011;11:450–462.
6. Legendre C, Canaud G, Martinez F. Factors influencing long-term outcome after kidney transplantation. *Transpl Int*. 2014;27:19–27.
7. Watson CJ, Johnson RJ, Birch R, et al. A simplified donor risk index for predicting outcome after deceased donor kidney transplantation. *Transplantation*. 2012;93:314–318.
8. Molnar MZ, Kovesdy CP, Mucsi I, et al. Higher recipient body mass index is associated with post-transplant delayed kidney graft function. *Kidney Int*. 2011;80:218–224.
9. Díez-Sanmartín C, Sarasa Cabezuelo A. Application of artificial intelligence techniques to predict survival in kidney transplantation: a review. *J Clin Med*. 2020;9:572.
10. Coemans M, Süsal C, Döhler B, et al. Analyses of the short- and long-term graft survival after kidney transplantation in Europe between 1986 and 2015. *Kidney Int*. 2018;94:964–973.
11. Scientific Registry of Transplant Recipients. Annual data report: kidney. Available at https://srtr.transplant.hrsa.gov/annual_reports/2019/Kidney.aspx. Accessed March 22, 2023.
12. Hemmersbach-Miller M, Alexander BD, Sudan DL, et al. Infections after kidney transplantation does age matter? *Clin Transplant*. 2019;33:e13516.
13. Rømming Sørensen V, Heaf J, Wehberg S, et al. Survival benefit in renal transplantation despite high comorbidity. *Transplantation*. 2016;100:2160–2167.
14. Stoumpos S, Jardine AG, Mark PB. Cardiovascular morbidity and mortality after kidney transplantation. *Transpl Int*. 2015;28:10–21.
15. Strohmaier S, Wallisch C, Kammer M, et al. Survival benefit of first single-organ deceased donor kidney transplantation compared with long-term dialysis across ages in transplant-eligible patients with kidney failure. *JAMA Netw Open*. 2022;5:e2234971.
16. Maamoun HAH, Soliman AR, Fathy A, et al. Diabetes mellitus as predictor of patient and graft survival after kidney transplantation. *Transplant Proc*. 2013;45:3245–3248.
17. Kim H, Cheigh JS. Kidney transplantation in patients with type 1 diabetes mellitus: long-term prognosis for patients and grafts. *Korean J Intern Med*. 2001;16:98–104.
18. Revanur VK, Jardine AG, Kingsmore DB, et al. Influence of diabetes mellitus on patient and graft survival in recipients of kidney transplantation. *Clin Transplant*. 2001;15:89–94.
19. Sørensen VR, Sørensen SS, Feldt-Rasmussen B. Long-term graft and patient survival following renal transplantation in diabetic patients. *Scand J Urol Nephrol*. 2006;40:247–251.
20. Jeon JY, Kim SJ, Ha KH, et al. Trends in the effects of pre-transplant diabetes on mortality and cardiovascular events after kidney transplantation. *J Diabetes Investig*. 2021;12:811–818.
21. Molnar MZ, Kovesdy CP, Bunnapradist S, et al. Associations of pretransplant serum albumin with post-transplant outcomes in kidney transplant recipients. *Am J Transplant*. 2011;11:1006–1015.
22. Dahlberg R, Muth B, Samaniego M, et al. One-year serum albumin is an independent predictor of outcomes in kidney transplant recipients. *J Ren Nutr*. 2010;20:392–397.
23. Law JP, Borrows R, McNulty D, et al. Early renal function trajectories, cytomegalovirus serostatus and long-term graft outcomes in kidney transplant recipients. *BMC Nephrol*. 2021;22:1–11.
24. Raynaud M, Aubert O, Reese PP, et al. Trajectories of glomerular filtration rate and progression to end stage kidney disease after kidney transplantation. *Kidney Int*. 2021;99:186–197.
25. Van Loon E, Zhang W, Coemans M, et al. Forecasting of patient-specific kidney transplant function with a sequence-to-sequence deep learning model. *JAMA Netw Open*. 2021;4:e2141617.
26. Dong JJ, Wang L, Gill J, et al. Functional principal component analysis of glomerular filtration rate curves after kidney transplant. *Stat Methods Med Res*. 2018;27:3785–3796.
27. Boucquemont J, Loubère L, Metzger M, et al; NephroTest Study Group. Identifying subgroups of renal function trajectories. *Nephrol Dial Transplant*. 2017;32(suppl_2):ii185–ii193.
28. Naesens M, Lerut E, Emonds MP, et al. Proteinuria as a noninvasive marker for renal allograft histology and failure: an observational cohort study. *J Am Soc Nephrol*. 2016;27:281–292.
29. Liu Y, Bendersky VA, Chen X, et al. Post-kidney transplant body mass index trajectories are associated with graft loss and mortality. *Clin Transplant*. 2023;2022:1–13.
30. Schmidt M, Schmidt SAJ, Sandegaard JL, et al. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449–490.
31. Nordic Centre for Classifications in Health Care. *NOMESCO Classification of Surgical Procedures (NCSP), Version 1.15*. Vol 93. Nordic Centre for Classifications in Health Care; 2010.
32. Pedersen MK, Eriksson R, Reguant R, et al. A unidirectional mapping of ICD-8 to ICD-10 codes, for harmonized longitudinal analysis of diseases. *Eur J Epidemiol*. 2023;38:1043–1052.
33. Jensen AB, Moseley PL, Oprea TI, et al. Temporal disease trajectories condensed from population-wide registry data covering 62 million patients. *Nat Commun*. 2014;5:4022.
34. Siggaard T, Reguant R, Jørgensen IF, et al. Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 72 million Danish patients. *Nat Commun*. 2020;11:1–10.
35. Charrad M, Ghazzali N, Boiteau V, et al. Nbclust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw*. 2014;61:1–36.
36. Dalgaard P. Rates and Poisson regression. In: Härdle W.K., editor. *Introductory Statistics with R*. Springer; 2008:259–274.
37. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40:373–383.
38. Grann AF, Erichsen R, Nielsen AG, et al. Existing data sources for clinical epidemiology: the clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. *Clin Epidemiol*. 2011;3:133–138.
39. Muse VP, Aguayo-Orozco A, Balaganeshan SB, et al. Population-wide analysis of hospital laboratory tests to assess seasonal variation and temporal reference interval modification. *Patterns (N Y)*. 2023;4:100778.
40. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28:112–118.
41. Shabir S, Halimi JM, Cherukuri A, et al. Predicting 5-year risk of kidney transplant failure: a prediction instrument using data available at 1 year posttransplantation. *Am J Kidney Dis*. 2014;63:643–651.
42. Moore J, He X, Shabir S, et al. Development and evaluation of a composite risk score to predict kidney transplant failure. *Am J Kidney Dis*. 2011;57:744–751.

43. Dos Santos Q, Hornum M, Terrones-Campos C, et al. Posttransplantation diabetes mellitus among solid organ recipients in a Danish Cohort. *Transpl Int*. 2022;35:10352.
44. Hergesell O, Zeier M. Underdiagnosis of diabetes mellitus in chronic dialysis patients on the renal transplant waiting list. *Transplant Proc*. 2003;35:1287–1289.
45. Bergrem HA, Valderhaug TG, Hartmann A, et al. Undiagnosed diabetes in kidney transplant candidates: a case-finding strategy. *Clin J Am Soc Nephrol*. 2010;5:616–622.
46. Bosch X, Lozano F, Cervera R, et al. Basophils, IgE, and autoantibody-mediated kidney disease. *J Immunol*. 2011;186:6083–6090.
47. Aljadi Z, Nopp A, Winqvist O, et al. Altered basophil function in patients with chronic kidney disease on hemodialysis. *Clin Nephrol*. 2017;88:86–96.
48. Ying T, Shi B, Kelly PJ, et al. Death after kidney transplantation: an analysis by era and time post-transplant. *J Am Soc Nephrol*. 2020;31:2887–2899.
49. Sumida K, Kovesdy CP. Disease trajectories before ESRD: implications for clinical management. *Semin Nephrol*. 2017;37:132–143.
50. Kovesdy CP, Coresh J, Ballew SH, et al; CKD Prognosis Consortium. Past decline versus current eGFR and subsequent ESRD risk. *J Am Soc Nephrol*. 2016;27:2447–2455.
51. Li L, Astor BC, Lewis J, et al. Longitudinal progression trajectory of GFR among patients with CKD. *Am J Kidney Dis*. 2012;59:504–512.
52. Loupy A, Aubert O, Orandi BJ, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. *BMJ*. 2019;366:l4923.
53. Raynaud M, Aubert O, Divard G, et al. Dynamic prediction of renal survival among deeply phenotyped kidney transplant recipients using artificial intelligence: an observational, international, multicohort study. *Lancet Digit Health*. 2021;3:e795–e805.
54. Kaboré R, Haller MC, Harambat J, et al. Risk prediction models for graft failure in kidney transplantation: a systematic review. *Nephrol Dial Transplant*. 2017;32(suppl_2):iii68–iii76.
55. Waljee AK, Mukherjee A, Singal AG, et al. Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*. 2013;3:e002847.