TRANSPLANT INTERNATIONAL

Transplant International ISSN 0934-0874

ORIGINAL ARTICLE

Epidemiology of infections in kidney transplant recipients – data miner's approach

Bartosz Wojciuk,[1] Marek Myślak,[2] Krzysztof Pabisiak,[2] Kazimierz Ciechanowski[2] and Stefania Giedrys-Kalemba[1]

1 Department of Microbiology and Immunological Diagnostics, previously Department of Microbiology and Immunology, Pomeranian Medical University, Szczecin, Poland

2 Clinic of Nephrology, Transplantation and Internal Medicine, Pomeranian Medical University, Szczecin, Poland

Keywords
data mining, infection, kidney transplantation.

Correspondence
Bartosz Wojciuk, Clinic of Nephrology, Transplantation and Internal Medicine, Pomeranian Medical University, ul. Powstańców Wlkp. 72, 70–123 Szczecin, Poland.
Tel.: +48609666028;
fax: +48914661197;
e-mail: barciuk@interia.pl

Conflicts of Interests
The authors have declared no conflicts of interest.

Received: 24 March 2014
Revision requested: 25 April 2014
Accepted: 30 January 2015
Published online: 27 March 2015

doi:10.1111/tri.12536

Summary

Infections remain a frequent complication following organ transplantation. Agents present within the general population remain common in recurrent infections among renal transplant recipients. *Data mining* methodology has become a promising source of information about patterns in the organ transplant recipient population. The aim of the study was to use *data mining* to describe the factors influencing single and recurrent infections in kidney transplant recipients. A group of 159 recipients who underwent kidney transplantation between 2005 and 2008 was analysed. RapidMiner and Statistica softwares were used to create decision tree models based on CART Quinlan and C&RT algorithms. There were 171 microbiologically confirmed episodes among 67 recipients (41%), and 191 separate species isolations were performed. Over 50% of the infected patients underwent two or more infectious episodes. Two classification decision tree models were created. The following features were enabled to differentiate the groups with single or recurrent infections: the duration of cold ischaemia, the post-transplant hospitalization period, the cause of chronic kidney disease and pathogens. The post-transplant hospitalization period and the length of cold ischaemia appear to be the principal parameters differentiating the subpopulations analysed. These coexisting factors, connected with recurrent infections in kidney transplant recipients, resemble a network which requires an advanced analysis to support the traditional statistics.

Introduction

The epidemiology of infections in kidney transplant recipients has become an interesting object of research among clinicians. The first complex studies from the early nineties examined the kidney transplant recipient as a typical immunocompromised host [1]. Antibiotic prophylaxis has effectively reduced the risk of some opportunistic infections in kidney transplant recipients. Current guidelines recommend the oral administration of trimethoprim/sulphamethoxazole for prophylaxis against bacterial pathogens but most specifically against *Pneumocystis jirovecii* and ganciclovir/valganciclovir for prophylaxis against cytomegalovirus [2,3]. Despite this, infections in kidney transplant recipient still challenge clinical transplantologists and microbiologists. In the short term, they may create a considerable threat for recipients' life and well-being. In the long term, they become a burden for healthcare system which faces the problem of increasing medication costs. This is especially related to recurrent infections with multidrug-resistant strains. The majority of studies point to urinary tract infections (UTIs) as the most common infection and Gram-negative bacilli *Escherichia coli* as the most frequent pathogen [4–11]. It is worth emphasizing that the majority of publications report a single-centre experience [12–18]. Besides, the studies on risk factors frequently revise a set of previous clinical observations and implement hypothesis-based traditional statistics. These are obviously efficient in

recognition of particular risk factors. However, these fail in simultaneous examination of novel factors. They also appear inefficient in uncovering the hierachy of different attributes coexisting in one patient and this is in fact critical for highly individualized clinical practice. New computing methods based on *data mining* provide researchers with a possibility to gain a deeper insight into the problem. Non-hypothesis-based approach towards data investigation, which underlies the general idea, becomes the key to new opportunities within this methodology.

Starting with a definition, *data mining* consists of finding nontrivial connections within data. First used in nonmathematical data investigation (text mining), because of its flexibility, *data mining* has subsequently used classical statistical methodology to become a highly advanced method dedicated to the investigation of huge databases [19,20]. Contemporary interest in *data mining* has spread among certain scientific disciplines, starting from economics, followed by medicine. Depending on the assumptions and mathematical basis used, there are several sorts of *data mining* methods available: naïve Bayesian classifier, neural networks, association rules, clustering and decision trees [21,22]. These can be divided into three groups: supervised learning decision trees in which output is provided after the analysis of input data [23–25], unsupervised learning where data are analysed without any specific target (clustering) [23,25,26] and reinforcement learning (neural networks where algorithms represent the ability to self-adapt) [23,26,27].

For users without a professional mathematical background, decision trees appear outstandingly popular, and this is also true in medicine. As mentioned, they represent supervised learning tools. These require following input data: attributes e.g. a set of clinical parameters present in particular group (objects) and a class value which in fact identifies the exact scientific problem of particular analysis e.g. clinical outcome. In this manner, the researcher decides which features will be analyzed as the attributes and which as a class label. This approach makes decision trees even more useful for clinicians who, in this way, can verify their observations or results obtained from many previous studies based on classical statistics. The parametres used as hypotheses in classical statistics can, in this case, get verified as class labels. [23]. The general idea of the creating decision trees is to extinguish highly homogenous subgroups (leaves) identified with a class label and characterized by a set of attributes. In order to establish the hierarchy of their classificatory potential, the attributes are tested during recurrent operations. These are tending to divide the objects into possibly homogenous groups regarding particular attributes one by one. Once the most valuable attribute has been found, the same algorithm is implemented in the subgroups extinguished on this basis. Subsequently, decision trees present a systemic classification of objects described with a group of attributes and class label. They have been praised for their transparent graphical structures of patterns displayed, illustrated by leaf nodes and non-leaves (internal leaves or test leaves). Non-leaves are labelled by attributes, whereas leaves are labeled according to class value [28–30]. There are two validation rates used to evaluate the correctness of the model: *recall* and *precision. Precision* measures how many examples classified as 'positive' are in fact 'positive' and refers to medical term of *specificity. Recall* assesses how well the classifier can recognize positive samples and corresponds with *sensitivity* [31]. The aim of the study was to investigate patients suffering from infectious complications after kidney transplantation with the use of *data mining* methodology.

Patients and methods

Group

A group of 159 patients was examined. They underwent kidney transplantation between 2005 and 2008 and were observed within 24 months post-transplant (Table 1). Sixty-seven (41%: 33 women, 34 men) suffered at least one infectious episode (all patients were followed within the whole period). Thirty-seven (55.2%) of these were hospitalized longer than 17 days directly post-transplant and in 11 of those 37 (29.7%) first infection episode appeared during this hospitalization period. One hundred and seventy-one infectious episodes were diagnosed, and 191 pathogen isolations were performed; UTI and cytomegalovirus infection (CMV) comprised 79% of all episodes. Besides postoperative wound infections, blood infections and respiratory tract infections were also diagnosed.

Forms of infection

The definitions of particular clinical forms were established as follows: UTI – leukocyturia over 10 in high-power field plus positive urine culture according to the current guidelines, CMV – seroconversion in the case of primarily seronegative recipients combined with fever, diarrhea or decreased graft function and successful ganciclovir therapy, a positive test for CMV DNA in the case of a primarily seropositive recipient combined with fever, diarrhea or decreased graft function and successful ganciclovir therapy, postoperative wound infection-positive culture in material obtained from a wound combined with a local inflammatory reaction or symptoms of an impaired healing process, blood infection, positive blood culture with clinical features of systemic inflammatory reaction, respiratory tract infection-positive BAL culture with clinical features of local or systemic inflammatory reaction.

Table 1. Characteristics of the group.

Parameter	Women $n = 61$	Men $n = 98$	All $n = 159$
Age	46.1 Min. 20 Max. 65 SD = 12.6	43.6 Min. 17 Max. 72 SD = 13	44.9 Min. 17 Max. 72 SD = 13
CMV (+)	54	84	138 (86.8%)
CMV (–)	7	14	21 (13.2%)
First transplantation	51	78	129 (81.1%)
Second and following transplantation	10	20	30 (18.9%)
CIT	18.3 h Min. 5 Max. 42 SD = 7.7	18.6 h Min. 6 Max. 38 SD = 7.7	18.5 h Min. 5 Max. 42 SD = 7.6
Low HLA compatibility – more than two mismatches (including at least one DR antigen)	33	38	71 (44.6%)
Living donor	2	1	3 (1.8%)
Deceased donor	59	97	156 (98.2%)
Induction therapy – basiliximab	8	8	16
Induction therapy – daclizumab	0	2	2
Number of acute rejection episodes	8	7	15
Glomerulonephritis	17	31	48 (30.1%)
Diabetic nephropathy	5	15	20 (12.5%)
Autosomal dominant polycystic kidney disease	6	6	12 (7.5%)
Chronic pyelonephritis	7	4	11 (6.9%)
Congenital urinary tract malformations	3	8	11 (6.9%)
Other (including hypertension associated nephropathy)	23	34	57 (35.8%)

Treatment

Most patients were treated with a standard immunosuppression protocol, predominantly mycophenolate plus steroids and either tacrolimus (n = 122) or cyclosporine (n = 35). Everolimus ($n = 13$), sirolimus ($n = 16$) and azathioprine ($n = 2$) were also used in the protocol. They also received 1.5 g cefuroxime intravenously once directly before transplantation and 1,5 g cefuroxime daily during following 3 days post-transplant, 480 mg thrimetoprime/sulphametoxazole orally in a 6 months of post-transplant period. CMV-positive patients received antiviral prophylaxis with acyclovir 400 mg orally for 6 months, CMV negative who received graft from CMV-positive donors had valganciclovir administered orally, 900 mg for 100 days, and no routine antifungal prophylaxis was used. To be as precise as possible I ought to write as follows: The patients with ADPKD underwent nephrectomy on individual indications.

Data mining methodology

A Microsoft Access 2003 database was used to store clinical and demographic data which had been obtained from medical documents. CART Quinlan and C&RT algorithms were used to create two classifying decision tree models. Both algorithms represent the most conventional and frequently applied forms of decision tree modelling classification and enabled the creation and comparison of two sorts of models: multinodal (CART Quinlan) as well as binary (C&RT). Regarding algorithms, a single microbiological isolation of a pathogen was established as a *case*. The case itself was characterized by two attributes: type of pathogen and the moment of isolation post-transplant (in months).

Regarding the structure of the database, every case was also assigned to each particular patient and labelled as either a *single episode* or a *recurrent episodes,* who displayed a certain complex of further data such as: chronic kidney disease (CKD) pretransplant, cold ischaemia time (CIT in hours), human leucocytes antigens (HLA) compatibility, age, gender, length of post-transplant hospitalization (in days). These were chosen from the spectrum of factors previously recognized as risk factors (gender, age, CKD pretransplant, length of post-transplant hospitalization), as well as further factors discussed among clinicians but not yet fully verified.

Human leucocytes antigens compatibility was encoded as follows: *low* – compatibility with single DR class antigens and complete incompatibility in B class, or complete incompatibility in DR loci, *high* – all other cases.

For analytical purposes, the pathogens belonging to Gram-negative bacilli were divided into two subgroups: *Escherichia coli* and *other Gram-negative bacilli.* The species *Enterococcus faecalis* and *Enterococcus faecium* were analysed together as *Enterococcus* spp. RapidMiner and Statistica softwares were used to introduce the algorithms. Accuracy parameters were calculated for every model.

Traditional statistics

Logistic regression models for single episodes and recurrent infections were also created. Odds ratios were calculated for the parameters identical to these used in data mining algorithms, except for pathogens. P value <0.05 was used as the statistical significance level.

The type of pathogen was not considered in logistic regression, as logistic regression is a traditional statistical method when searching for typical causative connections among clinical conditions existing before particular episodes of infection. At the same time, decision trees created in this study represent classification, but not regression models, so that pathogens were included as one of classificatory attributes.

Results

Female gender and post-transplant hospitalization periods were found to increase the risk of recurrent infections in the logistic regression model, whereas CIT appeared to decrease the risk (OR < 1). No statistical significance was found in the group suffering from single episodes, and no statistical significance was found in particular CKD pretransplants, with regard to either single or recurrent episodes (Table 2).

Figure 1 presents the classifying model built by the CART Quinlan algorithm using the RapidMiner application: a multilevel decision tree was created. There are two highly homogenous leaves containing the patients with a single infection (1, 2 marked in blue), six highly homogenous leaves containing the patients with recurrent infections (3, 4, 5, 6, 7, 8 marked in red) and five less homogenous leaves. The combination of post-transplant hospitalization period, CKD pretransplant, moment of diagnosis post-transplant, CIT and HLA compatibility enables homogenous groups of patients to be distinguished. The type of pathogen and gender were not found classificatory in this model. The model presents high-accuracy parameters (Table 3).

The groups suffering from single episodes consisted of the following individuals:

1 Group 1, marked in blue – patients with CIT longer than 31 h

2 Group 2, marked in blue – recipients suffering from chronic pyelonephritis pretransplant with high HLA compatibility

Table 2. Factors related to single and recurrent infections according to logistic regression model.

	Single episode		Recurrent infections	
Parameter	Odds ratio	*P* value	Odds ratio	*P* value
Age	1.02	0.8	0.99	0.6
Female gender	2.16	0.06	**3.53**	**0.0007**
CIT	0.96	0.2	**0.8**	**0.0001**
Hospitalization post-transplant	0.97	0.4	**1.08**	**0.0001**
HLA compatibility	0.98	0.7	0.93	0.1
ADPKD	0.68	0.98	2.7	0.97
Unknown	2.2	0.96	0.76	0.99
Glomerulonephritis	1.04	0.99	0.45	0.98
Diabetic nephropathy	0.75	0.98	3.33	0.97
Hypertension-related nephropathy	2.49	0.95	3.64	0.97
Chronic pyelonephritis	2.50	0.95	1.9	0.98
Congenital urinary tract malformations	0.77	0.98	1.15	0.99

Values in bold represent statistic significance.

The groups suffering from recurrent infections consisted of the following individuals with CIT shorter than 31 h:

1 Group 3, marked in red, placed on the first left branch of the tree – recipients suffering from ADPKD pretransplant

2 Group 4, marked in red, placed on the second left branch of the tree – recipients suffering from chronic pyelonephritis pretransplant with low HLA compatibility

3 Group 5, marked in red, placed on the third left branch of the tree – recipients suffering from diabetic nephropathy pretransplant, hospitalized longer than 18 days post-transplant

4 Group 6, marked in red, placed on the central branch of the tree – recipients suffering from congenital urinary tract malformations pretransplant, diagnosed later than 3.5 months post-transplant

5 Group 7, marked in red, placed on the central branch of the tree – recipients suffering from congenital urinary tract malformations, diagnosed earlier than 3.5 months post-transplant, with low HLA compatibility

6 Group 8, marked in red, placed on the first right branch of the tree – recipients suffering from CKD of unknown origin, hospitalized for longer than 22 days post-transplant.

Figure 2 presents the classifying model, built with a C&RT algorithm, using a Statistica application to create a 6-levelled, 6-leaved, binary decision tree. To improve the work of the algorithm, those patients lacking single data were excluded. This model presents high-accuracy parameters (Table 4). Gender and HLA compatibility were not found classificatory in this model.

A combination of CIT, post-transplant hospitalization, the time of diagnosis post-transplant, the age and the type of pathogen distinguished six homogenous groups of patients (marked with colours and numbers), three groups suffering from a single episode of infection (groups 1, 4, 6, marked in blue) and three suffering from recurrent infections (groups 2, 3, 5 marked in red).

The groups with single episodes consisted of the following individuals:

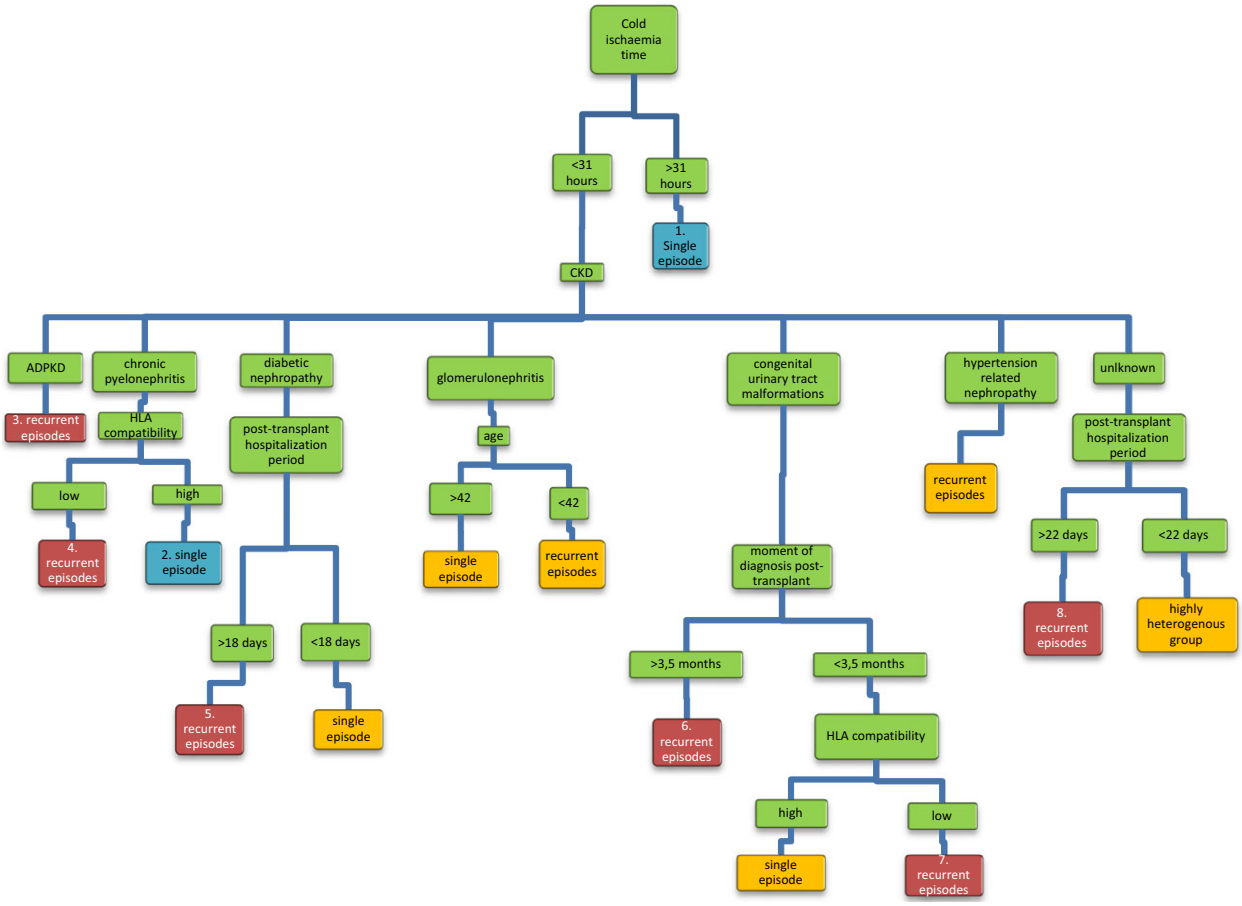
1 Group 1, marked in blue – patients with CIT longer than 31 h

2 Group 4, marked in blue – patients older than 41, diagnosed earlier than 4 months post-transplant, hospitalized shorter than 17 days post-transplant, with CIT shorter than 31 h,

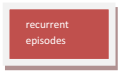
3 Group 6, marked in blue – patients younger than 41, infected with *Enterococcus* spp., diagnosed earlier than 4 months post-transplant, hospitalized shorter than 17 days post-transplant, with CIT shorter than 31 h.

The groups suffering from recurrent infections consisted of the following individuals:

1 Group 2 marked in red – patients with CIT shorter than 31 h, hospitalized longer than 17 days post-transplant



Legend:



Homogenous leaf consisting of recurrent infections cases.

Homogenous leaf consisting of single infection cases.

Heterogenous leaf consisting of both single and recurrent episodes cases with a prevalence of single episode cases.

Heterogenous leaf consisting of both single and recurrent episodes cases with a prevalence of recurrent episodes cases.

Highly heterogenous leaf consisting of both single and recurrent episodes cases.

Attributes and non-leafs.

Figure 1 Classifying decision tree for patients suffering from post-transplant infections based on CART Quinlan algorithm.

Table 3. Accuracy parameters for CART Quinlan model.

Accuracy 88.5%	Correct as *single episode*	Correct as *recurrent episodes*	Precision
Predicted as *single episode*	28	10	73.68%
Predicted as *recurrent episodes*	12	141	92.16%
Recall	70.00%	93.38%	

2 Group 5 marked in red – patients with CIT shorter than 31 h, hospitalized shorter than 17 days post-transplant, diagnosed later than 4 months post-transplant

3 Group 6 marked in red – patients infected with Gram-negative bacilli other than *Escherichia coli*, younger than 41, hospitalized less than 17 days post-transplant, with CIT shorter than 31 h.

The significant variety of pathogens isolated from the patients hospitalized longer than 17 days post-transplant is displayed (Fig. 3). This group of micro-organisms is characterized by the strong domination of *Enterobacter cloacae* and the presence of other *Enterobacteriaceae* spp. (*Escherichia coli, Klebsiella pneumoniae, Citrobacter freundii, Proteus mirabilis)*, as well as other Gram-negative bacilli (*Pseudomonas aeruginosa, Acinetobacter baumanii)* and fungi. The spectrum of species isolated from patients hospitalized shorter than 17 days post-transplant appears to be less differentiated. *Escherichia coli, Klebsiella pneumoniae* and *Enterococcus faecalis* dominate in this group, whereas *Enterobacter cloacae* and fungi isolates are almost absent.

Discussion

Data mining seems to be a promising method supporting the researchers. Since its invention, this has been widely known and used among economists. However, the application of particular algorithms has also become more frequent in medicine. Decision trees appear to be an evident example. The overall aim of new computational methods used in clinical research is the maximal personalization of treatment regarding the network of factors coexisting in one patient. Their advantage is the opportunity of nonhypothesis-based investigation, but on the other hand, the implementation of particular algorithms in clinical practice requires strong evidence. In the medical field, data mining still coexists with classical statistics, and its role is to sketch new directions of research, rather than to provide complete answers.

This study is one of the few implementing *data mining* in clinical transplantation [23,32] and highlights several issues which are significant in the authors' opinion. These are high-accuracy parameters of both models, similar classifiers used in both trees, differences between the logistic regression model and decision trees. Whilst the absence of gender in decision trees and the presence of CKD and HLA had been found insignificant in logistic regression, the role of CIT and post-transplant hospitalization was revealed in every model.

In terms of decision trees, the influence of CIT duration and post-transplant hospitalization period on the classification, and the classification itself, is identical. The binary characteristics of the second tree create a clear border of CIT values at 31 h, which (as far as the model is concerned) classifies the cases as recurrent infections. The same value appears in the first model, but in connection to its multinodal character, the strength of this attribute reflects a lower significance. In fact, discussing the influence of CIT on this particular group is problematic, as there are no data suggesting a connection between recurrent infections in kidney recipients and CIT duration. Furthermore, CIT duration may differ between different transplantation centres, so that with regard to the single-centre character of this analysis, the role of CIT in this context needs further multicentre studies.

The impact of post-transplant hospitalization period has been clearly displayed in the both decision tree models, and logistic regression model. The role of the length of hospitalization in the context of post-transplant infections has been outlined in a single study as a potential risk factor [18]. The association between the post-transplant hospitalization period and the frequency of recurrent infections is supported in the epidemiological analysis displayed in Fig. 3. Remarkably, the spectrum of pathogens isolated from those patients who left hospital earlier than on the 17th day of hospitalization resembles the aetiological agents of UTI in the general population, demonstrating the prevalence of *Escherichia coli.* In contrast, the spectrum isolated from those who stayed over 17 days at the clinic displays a significant impact of other Gram-negative bacilli such as *Enterobacter cloacae, Citrobacter freundii* and *Proteus* spp. According to the research conducted during the last 10 years at the centre that the patients come from, the species mentioned above had previously mostly been diagnosed as alert pathogens presenting the phenotype of AmpC, or extended spectrum beta-lactamase (ESBL)-mediated multidrug resistance [6]. These data make it highly reasonable to suspect that, regarding the post-transplant hospitalization period, the pathogen spectrum in the analysed group had evolved from the image of common UTI aetiology [33], to the typical nosocomial infections profile as shown mostly by patients with recurrent infections. At the same time, group 5 in the C&RT model contains patients who left hospital before the 17th day post-transplant, but had acquired similar, probably nosocomial flora, and further suffered from recurrent infections. What also needs mentioning is that in one-third cases of prolonged post-transplant hospitalization, the first infection episode

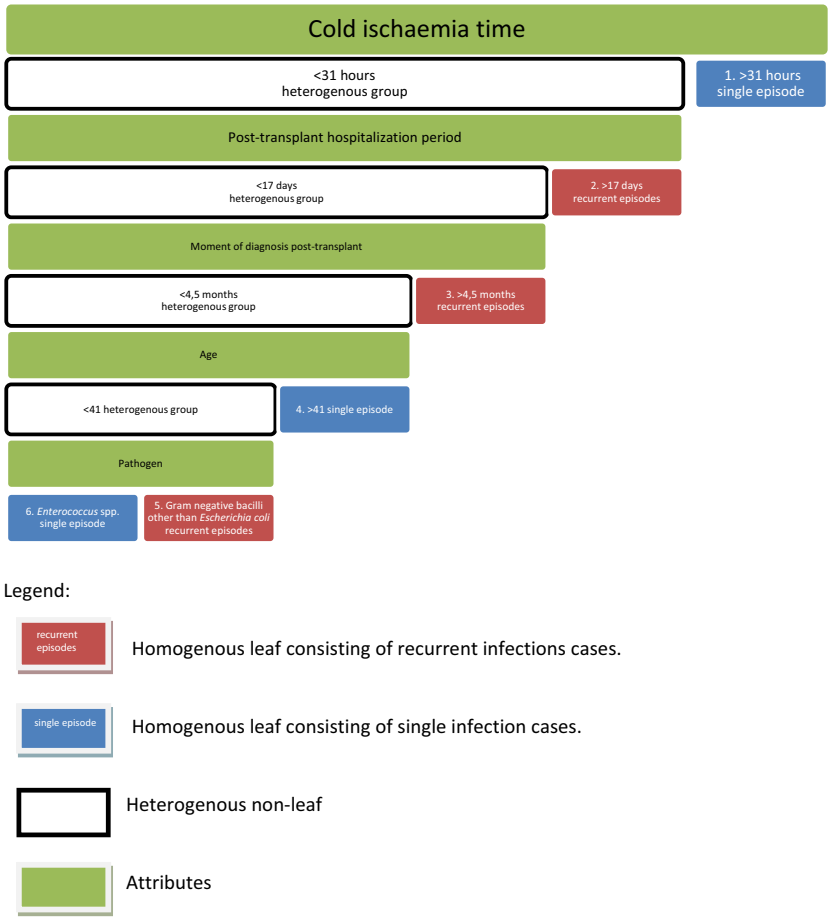


Figure 2 Binary classifying decision tree for patients suffering from post-transplant infections based on C&RT algorithm.

Table 4. Accuracy parameters for C&RT model.

Accuracy 90%	Correct as *single episode*	Correct as *recurrent episodes*	Precision
Predicted as *single episode*	20	2	91%
Predicted as *recurrent episodes*	13	114	90%
Recall	61%	98%	

appeared during this period which influenced duration of this period. However, the data from Fig. 3 (nosocomial flora isolated from patients after long post-transplant hospitalization), as well as attributes characterizing group 5 from C&RT model (as above), make it hypothetic that the exposure to nosocomial flora itself, in the majority but not all cases connected with prolonged hospitalization, influences long-term history of kidney recipients.

A Multinodal CART Quinlan model presents the cause of CKD as the second most important classificatory factor. However, apart from ADPKD, all causes of CKD have coexisted with other attributes.

Nevertheless, what seems worth emphasizing is that in logistic regression model, ADPKD pretransplant achieved high OR value: 2.7, but this has not been found statistically significant. Similar, diabetic nephropathy (OR = 3.33) has not been found significant in logistic regression and has been found classificatory within CART Quinlan algorithm in combination with the post-transplant hospitalization period. Post-transplant hospitalization period has been found a significant attribute in binary C&RT model, but in multinodal, CART Quinlan algorithm coexisted with other attributes such as diabetic nephropathy (as mentioned above) but also with unknown CKD pretransplant. At the same time, chronic pyelonephritis and urinary tract malformations have classified the cases as recurrent infections in combination with low HLA compatibility. This may possibly indicate the intensity of immune suppression, postulated as an independent risk factor, and remarkably, both sorts of CKD are connected to infections pretransplant. This has been noted by other authors using traditional methodology; however, deeper insight into the system of particular associations displayed

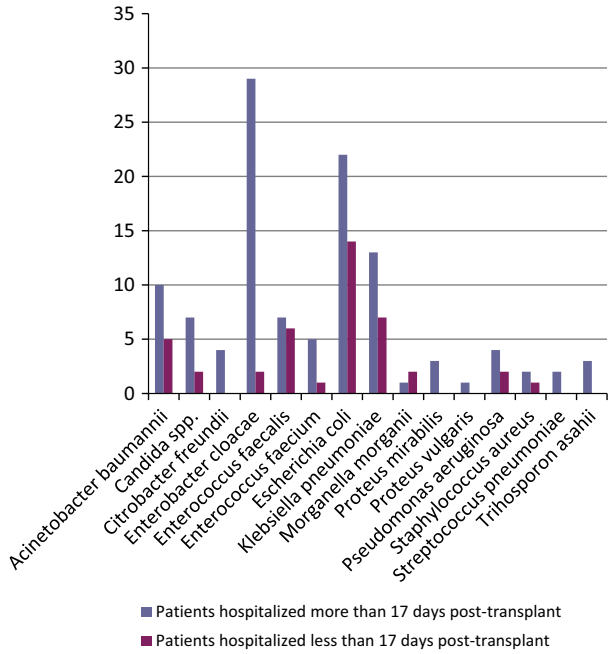


Figure 3 Spectrum of pathogens isolated dependently on post-transplant hospitalization period.

on the model remains unavailable in conventional statistics [34–36].

What appears particularly surprising is the absence of gender in the models, which is still regarded as an independent risk factor in terms of post-transplant UTIs and has been revealed in logistic regression model for recurrent infections [33,37]. Immunosuppressive regimen (including induction of immunosuppression) has been taken into account in the analysis. However, the models containing the data on immunosuppressive regimens have displayed unsatisfactory accuracy parameters (not shown). This is the reason why diabetes post-transplant has not been included in the analysis, as whether its impact on post-transplant infection is independent or CNI dependent is still in question [38]. This is also the reason why the question about the relation of changes in immunosuppression and the onset of infections may not have been addressed properly.

There are two interesting groups of patients, whose attributes are difficult to discuss at this stage and who probably need further investigation. These are group 6 in the CART Quinlan model and group 5 in C&RT model: patients with a relatively late onset of infection suffering from congenital urinary tract disorders pretransplant (according to CART Quinlan) and those hospitalized shorter than 17 days post-transplant (according to C&RT, already mentioned). In particular, the onset of recurrent infections, despite short post-transplant hospitalization, suggests the presence of other coexisting factors that ought to be taken into account in future analyses.

The role of data mining analyses in clinical studies is rather to point to new hypotheses to be tested further, than to provide complete answers. Regarding the classificatory, but not predictive character, of presented models, it appears challenging to judge particular risk factors only on the basis of these. The study reveals several weaknesses of which the authors are conscious. These are its single-centre character and the absence of certain factors such as urological disorders post-transplant. Nevertheless, the presence of the post-transplant hospitalization period in every model and the absence of gender in decision trees make it reasonable to hypothesize that the prolonged post-transplant hospitalization period may be a risk factor for recurrent infections, possibly connected to the exposure to nosocomial flora. Further *data mining* regarding other coexisting factors in this group, as well as molecular studies of the pathogens isolated from these patients, is required.

At the same time, the absence of gender as a classifier in decision trees does not detract from its role as a risk factor for post-transplant infections, as this has been strongly documented before. This fact, as well as unsatisfactory accuracy parameters displayed by the models that included immunesuppressive treatment, has challenged the authors. On this basis, the authors suggest that enclosing immune-suppression regimens into the analysis points to the need for studies using data mining and classical statistics in a complementary way, based on larger cohorts of predefined groups: men and women separately, patients stratified with medication protocols themselves, CKD pretransplant, post-transplant hospitalization or other factors such as these present in groups 5 and 6 described above.

Authorship

BW: participated in study design, data acquisition, data analysis and drafting the article. MM, KP: participated in data acquisition and analysis. SGK: participated in study design and drafting the article. KC: revised the article critically for important intelectual content.

Funding

The authors have declared no funding.

References

1. Rubin RH. Infectious disease complications of renal transplantation. *Kidney Int* 1993; **44**: 221.
2. Prophylaxis of infection in solid organ transplantation. Available at: http://www.update.com/contents/prophylaxis-of-infection-in-solid-organ-transplantation (accessed on 14.03.2013).
3. Subramanian AK. Antibiotic prophylaxis regimen following transplantation. *Curr Opin Infect Dis* 2011; **24**: 344.

4. Takai K, Aoki A, Suga A, *et al.* Urinary tract infections following renal transplantation. *Transplant Proc* 1998; **30**: 3140.
5. Chuang P, Parikh CR, Langone A. Urinary tract infections after renal transplantation: a retrospective review at two US transplant centers. *Clin Transplant* 2005; **19**: 230.
6. Wojciuk B, Jursa-Kulesza J, Giedrys-Kalemba S. Seven years' retrospective analysis of microbiological data collected from renal transplant recipients. *Nefrol Dial Pol* 2011; **3**: 379.
7. Martinez-Marcos F, Cisneros J, Gentil M, *et al.* Prospective study of renal transplant infections in 50 consecutive patients. *Eur J Clin Microbiol Infect Dis* 1994; **13**: 1023.
8. Veroux M, Giulfrida G, Corona D, *et al.* Infective complications in renal allograft recipients: epidemiology and outcome. *Transplant Proc* 2008; **40**: 1873.
9. Martin-Pena A, Cordero E, Fijo J, Sanchez-Moreno A, Martin-Govantes J, Torrubia F. Prospective study of infectious complications in a cohort of pediatric renal transplant recipients. *Pediatr Transplant* 2009; **13**: 457.
10. Grimaldi A, Barletta A, Rascente M, *et al.* Infectious complications in the renal transplant recipient. *Transplant Proc* 2005; **37**: 2502.
11. Khoury JA, Brennan DC. Infectious complications in kidney transplant recipients: review of the literature. *Saudi J Kid Dis Transplant* 2005; **16**: 453.
12. Pourmand G, Salem S, Mehrsai A, Taherimahmoudi M, Ebrahimi R, Pourmand MR. Infectious complications after kidney transplantation: a single centre experience. *Transpl Infect Dis* 2007; **9**: 302.
13. Charfeddine K, Zaghden S, Kharrat M, Kamoun K, Jarraya F, Hachicha J. Infectious complications in kidney transplant recipients: a single-center experience. *Transplant Proc* 2005; **37**: 2823.
14. Lyerová L, Viklický O, Nemcová D, Teplan V. The incidence of infectious diseases after renal transplantation: a single-centre experience. *Int J Antimicrob Agents* 2008; **31**(Suppl 1): S58.
15. Ranganathan M, Akbar M, Ilham MA, Chavez R, Kumar N, Asderakis A. Infective complications associated with ureteral stents in renal transplant recipients. *Transplant Proc* 2009; **41**: 162.
16. Giakoustidis D, Diplaris K, Antoniadis N, *et al.* Impact of double-j ureteric stent in kidney transplantation: single-center experience. *Transplant Proc* 2008; **40**: 3173.
17. Alangalden GJ, Thyagarajan R, Gruber SA, Morawski K, Garnick J, El-Amm JM. Infectious complications after kidney transplantation: current epidemiology and associated risk factors. *Clin Transplant* 2006; **20**: 401.
18. Dantas SRPE, Kuboyama RH, Muzzali M, Moretti ML. Nosocomial infections in renal transplant patients: risk factors and treatment implications associated with urinary tract and surgical site infections. *J Hosp Infect* 2006; **63**: 117.
19. Yoo I, Alafaireet D, Marinov M, *et al.* Data mining in healthcare and biomedicine: a survey of literature. *J Med Syst* 2012; **36**: 2431.
20. Iavindrasana I, Cohen G, Depeursinge A, Műller H, Meyer R, Geissbuhler A. Clinical data mining: a review. *Yearb Med Inform* 2009: 121.
21. Belazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *Int J Med Inf* 2008; **77**: 81.
22. Lavrac N, Zupan B. Data mining in medicine. In: OZ Maimon, L Rokach, eds. *Data Mining and Knowledge Discovery Handbook*. 2005, s. 1107. Available at: http://books.google.pl (accessed: 15.03.2013).
23. Greco R, Papalia T, Lofaro D, Maestripieri S, Mancuso D, Bonofiglio R. Decisional trees in renal transplant follow-up. *Transplant Proc* 2010; **42**: 1134.
24. Mitchell TM. Does machine learning really work? *Al Mag.* 1997; **18**: 11.
25. Quinlan JR. *C4.5 Programs for Machine Learning*. Morgan: Kauffman, 1992. 3.
26. Akl A, Amani MI, Ghoneim M. Prediction of graft survival of living donor kidney transplantation: nomograms or artificial neural networks? *Transplantation* 2008; **86**: 1401.
27. Forrest S. Genetic algorithms: principles of natural selection applied to computation. *Science* 1993; **261**: 872.
28. Maimon OZ, Rokach L. Decision trees. In: OZ Maimon, L Rokach, eds. *Data Mining and Knowledge Discovery Handbook*. 2005, s. 165 1107. Available at: http://books.google.pl (accessed on 15.03.2013).
29. Hammann F, Drewe J. Decision tree models for data mining in hit discovery. *Expert Opin Drug Discov* 2012; **7**: 342.
30. Castellani B, Castellani J. Data mining: qualitative analysis with health informatics data. *Qual Health Res* 2003; **13**: 1005.
31. Rokach L, Maimon O. Evaluation of classification trees in data mining with decision trees. *Theory Appl.* Series in machine perception artificial intelligence. 2008; **69**: 25. Available at: http://books.google.pl (accessed on 15.03.2013).
32. Lofaro D, Maestripieri S, Greco R, *et al.* Prediction of chronic allograft nephropathy using classification trees. *Transplant Proc* 2010; **42**: 1130.
33. Hummers-Pradiers E, Kochen MM. Urinary tract infections in adult general practice patients. *Br J Gen Pract* 2002; **52**: 752.
34. Papasotitriou M, Sawidaki E, Kalliakmani P, *et al.* Predisposing factors to the development of urinary tract infections in renal transplant recipients and the impact on long term graft function. *Ren Fail* 2011; **33**: 405.
35. Igbal T, Ngavi R, Akhter SF. Frequency of urinary tract infections in renal transplant recipients and effect on graft function. *J Pak Med Assoc* 2010; **60**: 826.
36. Sorto R, Irinar SS, Delgadillo G, Alben J, Correa-Rotter R, Morales-Buenrostro LE. Risk factors for urinary tract infections during the first year after kidney transplantation. *Transplant Proc* 2010; **42**: 280.
37. Vidal E, Torre-Cisneros J, Blanes M, *et al.* Bacterial urinary tract infection after solid organ transplantation in the RESITRA cohort. *Transpl Infect Dis* 2012; **14**: 595.
38. Säemann M, Hörl WH. Urinary tract infection in renal transplant recipients. *Eur J Clin Invest* 2008; **38**(Suppl 2): 58.