

Analyzing Association Rules for Graft Failure Following Deceased and Live Donor Kidney Transplantation

Syed Asil Ali NAQVI^{a,1}, Karthik TENNANKORE^b, Amanda VINSON^b and Syed Sibte Raza ABIDI^a

^a*Department of Computer Science, Dalhousie University, Halifax, NS, Canada*

^b*Department of Nephrology, Dalhousie University, Halifax, NS, Canada*

Abstract. This paper investigates the clinical attributes that contribute to kidney graft failure following live and deceased donor transplantation using an association rule mining approach. The generated rules are used to analyze the distinctive co-occurrence of attributes for those with or without all-cause graft failure. Analysis of a kidney transplantation dataset acquired from the Scientific Registry of Transplant Recipients that included over 95000 deceased and live donor recipients over 5-years was performed. Using an association rule mining approach, we were able to confirm established risk factors for graft loss after live and deceased donor transplantation and identify novel combinations of factors that may have implications for clinical care and risk prediction post kidney transplantation. Using lift as the metric to evaluate association rules, our results indicate that advanced recipient age (i.e. over 60 years), end stage kidney disease due to diabetes, the presence of recipient peripheral vascular disease and recipient coronary artery disease have a high likelihood of graft failure within 5 years after transplantation.

Keywords. Kidney Transplantation, Graft Survival, Association Rule Mining

1. Introduction

Outcomes following kidney transplantation are influenced by both donor and recipient factors [1]. These factors may span across clinical and socio-demographic domains [2], and have been shown to be predictors of kidney transplant allograft survival. Survival analysis is an established method to determine outcomes of a transplantation, informing both the occurrence of an adverse event (i.e. graft failure) and the time to occurrence.

Association Rule Mining (ARM) algorithms [3] are widely used to conduct exploratory data analysis in medicine and healthcare [4]. The rules generated by ARM denote the correlation of attributes, indicating their co-occurrence in a dataset as opposed to their causality, towards an outcome. The outcome variable is usually fixed to determine the most frequent co-occurring predictors for that particular outcome. ARM has been applied to discover the most frequent predictors for liver transplant and breast cancer datasets [5][6]. In these studies, the co-occurring attributes were later used to build a machine learning (ML) classifier. Although, the use of association rules is rigorous in

¹ Syed Asil Ali Naqvi, Department of Computer Science, Dalhousie University, Halifax NS, B3H1W5, Canada; E-mail: a.naqvi@dal.ca.

bioinformatics and healthcare, the literature on association rules to analyze survival outcomes is lacking. In one of the few studies to utilize association rules, Pinheiro et al. [7] performed ARM to determine factors associated with survival for liver cancer patients. To the best of our knowledge, ARM has not been applied in a kidney transplantation dataset to ascertain factors associated with graft survival. Identifying novel combinations of factors associated with graft survival may inform recipient selection and expected prognosis for both deceased and live donor transplantation.

In this paper, we present our analysis of an established kidney transplant dataset by applying ARM to explore the complex relationship of donor and recipient attributes to help improve the outcome of transplant graft survival. A dataset of living and deceased donor kidney transplants was derived from the Scientific Registry of Transplant Recipients (SRTR) spanning a 5 years period post transplantation.

2. Methods

2.1. Data

The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the US, submitted by the members of the Organ Procurement and Transplantation Network (OPTN). The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors.

The overall cohort consisted of 277,316 kidney transplant recipients from 2000-2017 and the dataset contained recipient and donor characteristics prior to transplant and the outcome of all-cause graft failure (inclusive of death with graft function). We used only complete cases (i.e., no attributes had missing values), totalling 97,485 deceased and live donor kidney transplants. Table 1 provides a list of key clinical attributes.

Table 1. List of clinical attributes used to generate the association rules

Feature Description	Abbreviation
Peak Panel Reactive Antibody – Discrete	PKPRAGROUP
Donor Age – Continuous	DAGE
Expanded Criteria Donor – Discrete (Deceased Only)	ECD
Donor Weight - Continuous	DWT
Recipient Weight - Continuous	RWT
Donor Creatinine – Continuous (Deceased Only)	DONCREAT
Donor Height - Continuous	DHT100
Recipient Height - Continuous	RHT100
Donor Hypertension – Discrete	DHTN
Recipient Hypertension – Discrete	RHTN
Cold Ischemia Time – Continuous	CIT
Recipient Age – Continuous	RAGETX
Donor Recipient Race Pair - Discrete	DRRACE
Number of Human Leukocyte Antigen mismatches – Discrete	HLAMM
Functional Status of the Recipient – Discrete	FUNCTSTAT
Recipient Cardiovascular Disease	RCVD
Recipient Peripheral Vascular Disease – Discrete	RPVD
Recipient Malignancy – Discrete	RMALIG
Years on dialysis pre-transplant – Continuous	VINTAGE
Pre-emptive Transplant – Discrete	PREEMPTIVE
Recipient Diabetes – Discrete	RDM
Recipient Coronary Artery Disease – Discrete	RCAD
End-Stage Kidney Disease (ESKD) diagnosis – Discrete	ESRDDX

2.2. Data Preprocessing

We included all graft failures (inclusive of death with graft function also known as all-cause graft failure) for a 5 year time-period and any individuals that did not experience graft failure within 5 years were deemed as “successful grafts”. Patients who were censored prior to completing 5 years of follow-up were excluded from the data set. Kazim et al. [8] also used a similar approach for discarding right-censored data whereby the censored cases prior to 7-years (the end point of their study) from the date of transplant were discarded during the data preprocessing stage. The ARM algorithms require the predictors to be categorical; therefore, nine attributes were binned into smaller categories based on expert advice and at approximate lower, median and upper quartiles.

2.3. Apriori Algorithm

The ARM algorithm was first proposed by Agrawal et al. [3] which showed that purchased items in transactional databases have hidden relationships. A rule is defined as an association between two frequent item sets of a transactional database and is represented as $X \rightarrow Y$, where X is the antecedent and Y is consequent. The evaluation of a rule is usually based on two measures namely, support (Supp) and confidence (Conf). The support of the rule $X \rightarrow Y$ is the percentage of the transactions in a database that contain both X and Y. Confidence is defined for the rule $X \rightarrow Y$ as the percentage of transactions containing X that also contains Y. The arules [9] R package was used for implementing the algorithm. Several combinations of algorithm parameters (minimum rule length, minimum support, and minimum confidence) were tried for the graft outcomes. The minimum support (minsup) and minimum confidence (minconf) for graft failure was selected as 1% and 60%, respectively, whereas we increased the minsup to 3% for successful grafts. The minconf for successful grafts was set to 80%. We found that the selected hyperparameters assured the interestingness of the rules.

Discovery of redundant rules is a known problem with this algorithm hence we used a two stage removal wherein we first used the *is.redundant()* and *is.maximal()* functions in the package and later took the guidance of domain expert for further refinement.

3. Results

Table 2 and 3 provide the association rules for failed and successful grafts for deceased and live donors, respectively. After pruning uninteresting rules, the total number of rules in for each outcome was below 100 from which we selected important and most interesting rules based on maximum lift values and nephrology expert’s clinical opinion.

Table 2. Rules for failed grafts for living (L) and deceased (D) donor types—the consequent is graft failure.

Type	Antecedent	Supp	Conf	Lift
D	{ragetx=>60yrs^rpvd=Yes^preemptive=>No}	1%	60%	1.65
D	{ecd=Yes^rcad=Yes^ragetx=>60yrs}	1%	60%	1.64
D	{rpvd=Yes^cit=>6hours^ragetx=>60yrs}	1%	60%	1.63
D	{rpvd=Yes ^doncreat=>1.5mg/dL^ragetx=>60yrs}	1%	60%	1.63
D	{dage=>60yrs^ragetx=>60yrs^vintage=>4yrs}	1%	60%	1.62
D	{dage=>60yrs^dhtn=Yes^esrdxx=Diabetes}	1%	60%	1.61

L	{functstat= 100% ^ragetx=> 60yrs ^cit< 6hours ^rdm2= Yes }	2%	72%	1.93
L	{functstat= 100% ^dage=> 50yrs ^ragetx=> 60yrs ^preemptive= No }	2%	68%	1.83
L	{esrddx= Diabetes ^rcad= Yes ^preemptive= No }	2%	61%	1.65
L	{rhtn= Yes ^rpvd= Yes ^rmalig= No } {esrddx= Diabetes ^drace= Donor White/Recipient White	2%	61%	1.64
L	^ragetx=> 60yrs ^preemptive= No }	2%	61%	1.61
L	{esrddx= Diabetes ^cit< 6hours ^rmalig= No ^rcad= Yes }	2%	62%	1.61

The results show that a number of attributes such as an advanced donor (DAGE) and recipient age (RAGETX), End Stage Kidney Disease due to diabetes (ESRDDX=diabetes), presence of recipient coronary disease (RCAD=yes) and recipient peripheral vascular disease (RPVD=yes) commonly occur among both type of donors who experience graft failure. A long time on dialysis before transplantation (i.e. greater than 4 years; VINTAGE) and cold ischemia time (CIT) greater than 6 hours were observed particularly in deceased donor recipients.

Table 3. Rules for successful grafts for living (L) and deceased (D) donor types. Consequent is graft success

Type	Antecedent	Supp	Conf	Lift
D	{ragetx=< 60yrs ^pkpragroup=< 20 ^esrddx= PCKD }	3%	81%	1.25
D	{dhtn= No ^esrddx= PCKD ^vintage=< 2yrs }	3%	81%	1.25
D	{ecd= No ^esrddx= PCKD ^vintage=< 2yrs }	3%	81%	1.25
D	{pkpragroup=< 20 ^esrddx= PCKD ^vintage=< 2yrs }	3%	80%	1.24
D	{ragetx=< 60yrs ^preemptive= Yes ^rdm2= No }	3%	80%	1.24
L	{esrddx= PCKD ^rhtn= Yes ^ragetx=< 60yrs ^preemptive= Yes }	3%	86%	1.36
L	{esrddx= PCKD ^ragetx=< 60yrs ^rpvd= No ^preemptive= Yes }	4%	85%	1.33
L	{rrace= White ^esrddx= PCKD ^vintage=< 2 yrs ^preemptive= Yes }	4%	83%	1.31
L	{functstat= 90% ^ragetx=< 60yrs ^rpvd= No ^preemptive= Yes }	4%	83%	1.31
L	{functstat= 80% ^ragetx=< 60yrs ^preemptive= Yes ^rcvd= No }	3%	82%	1.30

The rules for successful grafts were mostly similar for both deceased and living donor transplants. Table 3 shows that most of the rules were highly predictive of graft survival including the presence of ESKD due to polycystic kidney disease and pre-emptive transplantation (i.e. no pre-transplant dialysis time) the latter of which is consistent with what has been observed in the literature [10].

4. Discussion and Conclusion

In this explorative study, we used association rules to identify important predictive factors in a cohort of patients who did and did not experience graft failure in the first 5 years following transplantation. We identified that some factors were common for both live and deceased donor recipients (for example, recipient peripheral vascular disease, coronary artery disease and end-stage kidney disease due to diabetes). Whereas some factors (i.e. prolonged dialysis vintage) were of more importance in deceased donors. These findings align with what is often observed in clinical care and the literature. For example, recipient coronary artery disease and peripheral vascular disease have been identified as risk factors for poor graft outcomes following transplantation in deceased and live donor recipients [11]. Dialysis vintage would be expected to be a more important finding for deceased donor recipients, a population in whom one would expect extremes of pre transplant dialysis exposure [12] due to factors such as increased immune

sensitization. In addition to associations with graft loss, we identified factors associated with successful grafts i.e. survival to 5 years without death/graft loss. Polycystic kidney disease as the cause of ESKD and preemptive status (i.e. no prior dialysis exposure) were both associated with graft survival. Of interest, polycystic kidney disease has been shown to associate with better outcomes in some [13], but not all prior studies [14]. Importantly, we found novel combinations of factors that are not typically studied, suggesting that there may be novel phenotypes that can be further explored. For example, it is well known that prolonged cold ischemia time is a risk factor for graft loss following deceased donor transplantation [15]. However, our study identified that recipient peripheral vascular disease and prolonged cold ischemia time often associated with graft loss. A downstream clinical impact of this finding may be to emphasize the need to reduce cold ischemia time among deceased donor transplant recipients with pre-existing vascular disease. Therefore, our novel finding of association rules comprising three or more donor-recipient predictors in combination may have direct clinical application.

References

- [1] Naderi G, Azadfar A, Yahyazadeh SR, et al (2020) Impact of the donor-recipient gender matching on the graft survival from live donors. *BMC Nephrol* 21:5.
- [2] Oztekin A, Delen D, Kong Z (James) (2009) Predicting the graft survival for heart-lung transplantation patients: An integrated data mining methodology. *Int J Med Inform* 78:.
- [3] Agrawal R, Mannila H, Srikant R, et al (1996) Fast discovery of association rules. *Adv. Knowl. Discov. data Min.* 12:307–328
- [4] Altaf W, Shahbaz M, Guergachi A (2017) Applications of association rule mining in health informatics: a survey. *Artif Intell Rev* 47:313–340.
- [5] Raji CG, Vinod Chandra SS (2017) Long-Term Forecasting the Survival in Liver Transplantation Using Multilayer Perceptron Networks. *IEEE Trans Syst Man, Cybern Syst* 47:2318–2329.
- [6] Karabatak M, Ince MC (2009) An expert system for detection of breast cancer based on association rules and neural network. *Expert Syst Appl* 36:3465–3469.
- [7] Pinheiro F, Mu-Hsing Kuo, Thomo A, Barnett J (2013) Extracting association rules from liver cancer data using the FP-growth algorithm. In: 2013 IEEE 3rd International Conference on Computational Advances in Bio and medical Sciences (ICCABS). IEEE, pp 1–1
- [8] Topuz K, Zengul FD, Dag A, et al (2018) Predicting graft survival among kidney transplant recipients: A Bayesian decision support model. *Decis Support Syst* 106:97–109.
- [9] Hahsler M, Grün B, Hornik K (2005) arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. *J Stat Softw* 14:.
- [10] Haller MC, Kammer M, Oberbauer R (2019) Dialysis vintage and outcomes in renal transplantation. *Nephrol Dial Transplant* 34:555–560.
- [11] Kleinstaub A, Halleck F, Khadzhyrov D, et al (2018) Impact of Pre-existing Comorbidities on Long-term Outcomes in Kidney Transplant Recipients. *Transplant Proc* 50:3232–3241.
- [12] Goldfarb-Rumyantzev A, Hurdle JF, Scandling J, et al (2005) Duration of end-stage renal disease and kidney transplant outcome. *Nephrol Dial Transplant* 20:167–175.
- [13] Pruthi R, McClure M, Casula A, et al (2016) Long-term graft outcomes and patient survival are lower posttransplant in patients with a primary renal diagnosis of glomerulonephritis. *Kidney Int.*
- [14] Hadimeri H, Norden G, Friman S, Nyberg G (1997) Autosomal dominant polycystic kidney disease in a kidney transplant population. *Nephrol Dial Transplant* 12:1431–1436.
- [15] Peters-Sengers H, Houtzager JHE, Idu MM, et al (2019) Impact of Cold Ischemia Time on Outcomes of Deceased Donor Kidney Transplantation: An Analysis of a National Registry. *Transplant Direct* 5:e448.