Noname manuscript No.
(will be inserted by the editor)

Stochastic modelling and simulation of a kidney transplant waiting list

Katarína Cechlárová · Martina Hančová · Diana Plačková · Tatiana Baltesová

Received: date / Accepted: date

Abstract We created a dynamic stochastic model to evaluate the performance of a kidney transplantation system. Our model is applicable in the context of a small country where the legislation requires that a kidney from a deceased donor should be used whenever available. Using a systematic design of simulation experiments, we performed a complex simulation study based on real medical data to explore the impact of factors representing different rates of deceased kidneys harvesting, the proportion of patients with a willing living donor and different allocation policies. On the basis of careful statistical analysis carried out by two different statistical methodologies, ANOVA and bootstrap, we draw some important conclusions about the effects of these factors and recommendation for the medical community. The results of the study clearly demonstrate that in addition to increasing the numbers of kidney donors, deceased as well as living, the introduction of a kidney exchange

The work has been supported by VEGA grant 1/0311/18, APVV grant APVV-17-0568 and COST action CA15210 ENCKEP

Katarína Cechlárová
Institute of Mathematics, Faculty of Science, P.J. Šafárik University, Jesenná 5, Košice, Slovakia
E-mail: katarina.cechlarova@upjs.sk

Martina Hančová
Institute of Mathematics, Faculty of Science, P.J. Šafárik University, Jesenná 5, Košice, Slovakia
E-mail: martina.hancova@upjs.sk

Diana Plačková
Institute of Mathematics, Faculty of Science, P.J. Šafárik University, Jesenná 5, Košice, Slovakia
E-mail: diana.plackova@student.upjs.sk

Tatiana Baltesová
L. Pasteur University Hospital, Trieda SNP 1, Košice, Slovakia
E-mail: tatiana.baltesova@unlp.sk

program leads to further expansion of the numbers of donations and to shortening of waiting times for transplantations.

Keywords Transplantation · Waiting list · Kidney exchange · Stochastic modelling · Systematic design of simulation experiments

1 Introduction

Chronic kidney disease is a world-wide problem. It is estimated that approximately 0,1% of the world population suffer from end stage renal disease (ESRD for short). For ESRD patients, two treatment options are available: dialysis or transplantation. The country of the authors has the population of about 5.5 million and the evolution of the number of patients receiving regular dialysis treatment during the last 10 years, according to the data of National Health Information Center (2019) and National Transplant Organization (2019) is shown in Table 1.

While dialysis is associated with a low life quality and many undesired side effects, a patient after a successful kidney transplantation can lead a practically normal life, apart from life-long immunosupressive treatment. However, transplantation is impossible without donors. An organ from a deceased donor (DD for short) can be used, but availability of such an organ is very much unpredictable. Patients are listed in a transplant waiting list for deceased donors (DDWL for short) and the national authority (National Transplant Organization, 2019) is responsible for its management and allocation of kidneys. Further, as it is proved that one kidney is perfectly sufficient for life, a healthy person can donate one of his/her kidneys to a patient in need. Moreover, the long term function of a kidney from a living donor is significantly better compared to cadaveric kidneys Terasaki et al. (1995). With the improvement of surgical techniques that minimize the risk for the donor, transplantation from living donors (LD for short) has become a treatment of choice in many countries.

Table 1 shows the numbers of patients registered in the DDWL and the numbers of transplantations in Slovakia for the last 10 years as recorded by National Transplant Organization (2019).

Table 1 Statistics on the numbers of patients in the dialysis treatment, DDWL and transplantations in Slovakia (KT: Kidney Transplantation)

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Dialysis	4005	4288	4052	4254	4228	4302	4472	4424	4500	4628
DDWL	589	493	478	508	471	504	426	386	377	257
KT DDs	153	156	116	130	108	110	165	124	142	135
KT LDs	19	7	13	3	10	15	19	19	11	11

To prevent graft rejection, some immunological prerequisites have to be fulfilled. First is the AB0 blood type compatibility: a donor of blood group 0 can donate to anybody, a donor with blood type A only to recipients with blood group A or AB, a B-donor only to B or AB recipients and an AB-donor only to recipients of the same blood group. The second condition for transplantation is a negative crossmatch. This test confirms that there are no clinically relevant antibodies against donor's Human Leukocyte Antigens (denoted by HLA) in the recipient's serum. Although various desensitization protocols allowing transplantation across the blood group and HLA barriers are available, they are very expensive and the transplantation results are less favourable for patients (Halloran, 2004).

The aim of this paper is to model how the transplantation process could evolve in time under various assumptions. We model populations of patients and deceased donors that arrive randomly in time. We assume that the means of the considered stochastic processes do not change with time, but we consider two different intensities of deceased donors arrivals and two different proportions of patients bringing their willing donors – a genetic or emotional relative. In addition to the standard 'first-come-first-transplanted' allocation policy we also consider its modification aimed at helping highly sensitized patients. To increase the utilization of living donors, we also employ the idea of kidney paired donation (KPD for short, also known as kidney exchange), see Rapaport (1986). We also perform extensive statistical evaluation of the obtained results to help to estimate the influence of various parameters and policies on the outcomes of the donation process. Our results are applicable in the context of a small country where the legislation requires that a kidney from a deceased donor should be used whenever available even in the case when the patient has a willing living donor.

2 Related work

There are many scientific papers that analyze the dynamic process of organ donation. We would like to draw the reader's attention only to a selection of them. An overview of the features of the existing simulation studies is given in Table 2.

The aim of the paper by Zenios et al. (1999) was to compare alternative strategies for allocation of deceased donors' kidneys using Monte-Carlo simulation. Four allocation strategies were compared: the first-come first-transplanted (FCFT for short) policy; the point system used by the United Network of Organ Sharing (UNOS for short); an efficiency-based algorithm that incorporated correlates of patient and graft survival; and a distributive efficiency algorithm, which had an additional goal of promoting equitable allocation among African-American and other candidates. 10-year computer simulation was performed. Simulations identified the efficiency-based policy to be superior to the UNOS algorithm in terms of quality-adjusted life expectancy, median waiting time to transplantation among those who were transplanted and an increase in the

overall likelihood of transplantation. However, the median waiting time for African-American patients was increased compared with UNOS, whereas the median waiting time for other patients was substantially decreased. The distributive efficiency model yielded moderate decreases in median waiting times for both races.

The queueing model developed in Abellán et al. (2006) combines the information on the arrival of patients to the DDWL and the process of donation, while the authors assumed that each deceased donor can provide either one or two kidneys, and the proportion of donors donating two kidneys was θ. They assume that the number of daily arrivals and the number of daily donations are two independent stochastic processes which follow Poisson distributions of rates λ and μ, respectively. They considered the parameters θ, λ, μ of the model as unknown and used Bayesian inference for their estimation. Then they took 2,500 values sampled from the posterior distributions of these parameters and simulated the evolution of the queue for the next year and a half. The obtained results obtained by the modelling were compared with the real evolution of the DDWL in País Valencià, Spain. Finally, they simulated the behaviour of the queue under three different increments in the donation rates and compared the predicted mean size of the waiting list.

Other papers try to evaluate the efficiency of kidney paired donation (KPD for short) programs. Segev et al. (2005a) simulated a cohort of incompatible donor-recipient pairs. Initially, 250 pairs were simulated. Each month an optimized match based on Edmond's algorithm was performed and 250 new pairs were added to the incompatible pool. The second option was the interval of 4 months, when 1000 pairs were allowed to accumulate between matches. To compare median waiting times for patients of different races and different blood groups, three different scenarios have been considered. In the first scenario, patients were entered into an optimized match only within their race. In the second scenario, patients of all races were entered into a combined optimized match. In the final scenario, bonus points were given in an effort to balance matching inequalities for minority races. The simulations suggest that the most cost-efficient modality for transplanting incompatible patient-donor pairs is a national kidney paired donation program. A minority bonus can improve matching opportunities for the treated minority, but at the expense of the majority population.

In a recent work, Santos et al. (2017) took into consideration various policies of KPD programs found in practice (incompatible pairs, altruistic donors, and compatible pairs) and various matching policies including longer cycle and chain lengths. The authors developed a modular simulator enabling 72 different configurations for each instance and performed extensive numerical simulations. Their main conclusion is that policies should encourage compatible pairs to enter the KPD pool, as this leads to remarkable improvements on the number of transplants. They also found that shorter time intervals between matches lead to higher number of effective transplants and to shorter waiting times for patients. Moreover, the inclusion of compatible pairs can lead to greater benefits for 0-blood type patients.

Table 2 Overview of simulation studies (KT: kidney transplantation, NDD: nondirected living donor)

Publication	Research question	Donors	Policies/Scenarios	Outcome variables	Methods
Zenios et al. (1999)	Equity and efficiency of alternative organ allocation strategies	DD only	1. FCFT, 2. UNOS point system, 3. maximizing QALE, 4. distributive efficiency	patient survival, QALE, median waiting time, KT probability	95% confidence intervals for performance measures, Kaplan-Meier techniques for survival estimate
Segev et al. (2005a)	Compare waiting times of various patient groups (blood group, race)	LD, incompatible pairs	time between match runs, matching within race groups, minority bonus	median waiting time, KT probability	Wilcoxon rank - sum test
Segev et al. (2005b)	Estimate the impact of improved matching schemes	LD, incompatible pairs	first accept scheme, optimal matching algorithm, national vs. regional matching	number of kidneys matched, HLA mismatch, 5-year graft survival, cost	Wilcoxon paired sign-rank test
Abellán et al. (2006)	Estimate the parameters of the DDWL so as to predict its size	DD with 1 or 2 kidneys	donation rates	mean size of the DDWL	queueing theory, Bayesian statistics
Santos et al. (2017)	Develop flexible simulation tool	incompatible/compatible pairs, NDD	matching frequency, maximum cycle/chain, types of pairs	number of transplants, waiting times, pool composition	integer linear programming, descriptive statistics
Ashlagi et al. (2018)	Estimate the effect of the frequency of match-runs	patient-donor pairs and NDD	weights for PRA, no-delay vs. delay model, arrival and/departure rates, cycle lengths	fraction of patient-donor pairs transplanted, average waiting time to transplants	descriptive statistics

Simulation studies are complemented by publications that develop theoretical models to analyse the processes connected with waiting for transplantation. As this problem has various different aspects that have to be taken into account, it is difficult to propose an exact model capturing all its features and possibilities. Still, the existing publications may provide some explanations for the phenomena observed in reality.

Zenios (1999) proposes a queueing model for the DDWL. The author assumes that there are several classes of patients, several classes of organs, and patients reneging due to death. Focusing on randomized organ allocation policies he developed closed-form asymptotic expressions for the stationary waiting time, stationary waiting time until transplantation, and fraction of patients who receive transplantation for each patient class. These expressions helped to explain that the main reasons of longer waiting times of African-Americans compared to those of Caucasians might be not the allocation policy alone, but the mortality rates for the various patient classes as well.

Stanford et al. (2014) address the 'blood type 0 problem', which means that recipients of blood type 0 experience longer waiting times than those of other blood types. This is partly due to cross-transplantation of too many organs of type 0 to compatible recipients of other blood types. The authors constructed a special structure that they called an 'array of idealised transplant queues' (AITQ). This structure consists of several queues that are linked together so as to model that donor organs of a particular type can be used for recipients of specified compatible blood groups. The authors show that AB0 identical transplantation cannot achieve equity either. The policy they propose allows for a small fraction of type 0 organs to be transplanted into type B recipients, and another small fraction of type A organs to be transplanted into type AB recipients and achieves comparable waiting times for all blood types.

Similar conclusion has been obtained by Perlman et al. (2018). The authors consider two distinct random streams of discrete objects and two distinct types of resources that arrive stochastically over time. If a resource unit is not allocated immediately, it is lost. In the context of kidney transplantations, the two queues of objects and resources, respectively, correspond to candidates and kidneys of blood group 0 and B. When a kidney arrives, it is allocated to the queue of a particular blood type. For each candidate in this queue the number of HLA mismatches is randomly generated and the kidney is allocated to the candidate with the best fit, independently of his position in line.

The authors define a new measure of system effectiveness, called Expected Value of Transplantation (EVT), that takes into account the extent to which the candidates and the kidneys they receive are compatible in terms of their HLA. As expected, long queues and long waiting times are associated with higher EVT values, as they increase the likelihood that an incoming kidney will find a well-matched candidate. On the other hand, long waiting times are expected to lead to deterioration in candidates' health. Therefore the authors propose an additional measure of system effectiveness: the ratio of waiting time to EVT that balances the two goals of achieving equitable waiting times and maximizing the overall quality of transplants. They show that a small

fraction of type 0 blood kidneys should be cross-transplanted to blood type B candidates in order to optimize the effectiveness of the system.

Other papers consider deceased as well as living donors. Zenios (2002) assumes a flow of arrivals of two types of incompatible patient-donor pairs: in an A-B pair the patient is of blood group A and his/her associated donor of blood group B; in a B-A pairs the blood groups of the patient and the donor are exchanged. He also assumes that a kidney from a deceased donor is available immediately. Zenios asks how to dynamically manage the mix of direct (between two incompatible patient-donor pairs) and indirect exchanges (the living donor donates to DDWL and the associated recipient receives a kidney from a DD) to maximize the expected total discounted quality years of the candidates in the participating pairs. Direct exchanges are preferable because the candidate receives a living-donor organ that leads to better results than the inferior cadaveric organ an indirect exchange provides (Terasaki et al., 1995). However, the former involves waiting. To address this question, Zenios developed a double-ended queueing model for a simple exchange system with two types of donor-recipient pairs. The optimal policy takes the form of a two-sided regulator. A newly arrived incompatible pair enters the queue to wait for a direct exchange as long as the double-ended queueing process is between the two barriers. The author stressed that the exchange program can be successful only if the used policy strikes a balance between maximizing the social welfare of the participants, and simultaneously respects their autonomy and minimizes the risk of participant resentment.

3 Our model

We modelled the arrival of patients and deceased donors as two independent Poisson processes with parameters λ_p and λ_d, respectively. The used time unit was one month. Parameter $\lambda_p = 153/12$ corresponds to approximately 153 new registrants in the DDWL yearly and it has been estimated on the basis of clinical experience. For the flow of DDs we used two different values $\lambda_d = 108/24$ and $\lambda_d = 165/24$ that are consistent with 108 and 165 kidneys from deceased donors yearly. These values have been chosen on assumption that each deceased donor donates both kidneys and they capture the most pessimistic as well as the most optimistic scenario based on the National Transplant Organization (2019) data from the last 10 years, see also Table 1.

According to the data provided by the National Blood Service in Slovak Republic (2019), the most prevalent blood group in Slovak population is A with 42 %, then 0 with 32 %, blood groups B and AB account for 18% and 8%, respectively. For each arriving patient and DD we randomly generated his/her blood group according to this blood groups distribution. For each arriving patient we also randomly generated the level of his/her sensitization. H (high sensitization) patients are those with the PRA level above 80 %, the other patiens are considered to be L (patients with low sensitization). The chosen threshold is compatible with the thresholds used in Zenios et al. (1999)

and Segev et al. (2005a). The proportion of H patients in our starting sample was 8%, so we used this probability to model the patients flow too.

To be able to model living donation, for each patients arrival sample we also randomly generated a willing living donor for 20% or 40% of arriving patients. The donor's blood group was generated in the same manner as above, independently of his/her associated patient's blood group.

When a donation from a donor d to a patient p is considered, first we check the blood group compatibility of donor d with patient p. If they turn out not to be compatible, we do not proceed with transplantation. Otherwise we randomly generate a catch-all variable β with a uniform distribution in $[0, 1]$ interval that captures tissue type compatibility, possible cross-match and other immunological and health associated factors in the patient. If β turns out to be higher than 0.2 for a H patient and higher than 0.8 for a L patient then no transplantation is possible.

For allocation of DDs we use two different policies. The FCFT (first-come-first-transplanted) policy means that when a deceased donor d arrives, we process the patients on the DDWL in the order of their entry date to the DDWL. The first patient p for whom both compatibility tests give green light, is marked as transplanted and leaves the DDWL. The HL-policy gives priority to H patients for whom it is more difficult to find a suitable donor and it works as follows. When a deceased donor d arrives, we first process only the H patients on the DDWL in the order of their entry date to the DDWL and only if no such patient could use donor d, we repeat the procedure with L patients on the DDWL.

When we deem the transplantation to a patient p possible we record the time of the arrival of donor d as the time when p was transplanted.

We further consider 4 different allocation models.

Model 1, DD only. It assumes only DDs and it is described above.

Model 2, DD + LD. This model considers living donors in addition to DDs. If a patient p with his/her willing living donor $d(p)$ arrives then donation of $d(p)$ to p is attempted using the same two tests as with a DD. If the transplantation is not possible, donor $d(p)$ is lost and patient p is put into DDWL. By contrast, if this donation is not excluded, we assume that it is performed immediately, i.e., the waiting time of patient p is zero.

Model 3, KPD. Here we use the idea of kidney paired donation (KPD), first suggested by Rapaport (1986) and now a part of kidney transplantation programs in many countries (for the current situation in European countries see the overview in Biró et al. (2019), Ferrari et al. (2012) for Australia, Ashlagi et al. (2018) for USA or Kim (2014) for South Korea). Incompatible patient-donor pair $X = (p, d(p))$ is kept in the database and they wait for more incompatible patient-donor pairs. When another such pair $\widehat{X} = (\widehat{p}, d(\widehat{p}))$ arrives, kidney exchange is attempted. This means that we check whether the donor $d(p)$ of the pair X can donate to the patient $\widehat{p}$ of pair $\widehat{X}$ and simultaneously whether the donor $d(\widehat{p})$ of pair $\widehat{X}$ can donate to the patient p of pair X. If the

compatibility tests are positive in both cases, paired donation is performed at the time of the arrival of the second pair.

Model 4, C-KPD. This model extends Model 3 by considering possible cyclic exchanges of three incompatible patient-donor pairs in addition to paired exchanges. For three pairs $X, \widehat{X}, \overline{X}$ this means attempting either the cyclic exchange of kidneys $d(p)$ to $\widehat{p}$, $d(\widehat{p})$ to $\overline{p}$ and $d(\overline{p})$ to p, or the cyclic exchange of kidneys $d(p)$ to $\overline{p}$, $d(\overline{p})$ to $\widehat{p}$ and $d(\widehat{p})$ to p .

In Models 3 and 4, the patients of incompatible patient-donor pairs are kept in DDWL and while waiting for a possible kidney paired donation or cyclic kidney exchange they are considered as candidates for a cadaveric transplantation in accordance with Model 1 for each arriving DD. This assumption reflects the current legislative situation in the country of the authors that literally demands that 'an organ from a living person may not be used if an organ from a deceased donor is available'.

Let us mention here that although longer exchange cycles are possible, sometimes involving more than one country (Biró et al., 2019), we do not consider them in this paper. Also, in our case, always only the new arriving pair is compared with other incompatible pairs in the database, so the use of sophisticated graph matching algorithms or integer programming is not necessary, unlike in Santos et al. (2017).

4 Design and methods of the simulation study

We applied the previously described stochastic models in a simulation study based on the systematic design of simulation experiments described by Lorscheid et al. (2012) and Law (2014). Key steps of our simulation study follow the general scheme from Lorscheid et al. (2012, p. 30) using the technique called factorial design which assures a systematic analysis producing valid and objective results.

Our research goal consists in a systematic examination of the influence of four factors, according to the design of experiment (DOE) scheme proposed in Figure 1, namely (F1) the percentage of patients with a willing LD, (F2) the flow intensity of DDs given by λ_d, (F3) the policy for DDs kidney allocations and (F4) the type of chosen allocation model for willing LDs, on two response variables – (R1) the average median of waiting time (WT for short) on the DDWL for transplanted patients and (R2) the average probability of kidney transplantation (KT for short) in the given DDWL. Each combination of these four factors, called also a design point or treatment, provided the simulation settings for performing a corresponding simulation experiment. Therefore, in our case we have $v = (2\,\lambda_d \text{ levels}) \times (2 \text{ LDs levels}) \times (2 \text{ policy levels}) \times (4 \text{ models}) = 32$ design points with respect to the given factor levels. Such DOE is also known as $2 \times 2 \times 2 \times 4$ *factorial design* for the simulation study.

For each design point from the given 32 design points, we started with top-level factors of the corresponding simulation experiment, the (F1) per-

Fig. 1 The scheme of $2 \times 2 \times 2 \times 4$ factorial DOE for the simulation study

centage of patients with LDs and (F2) the flow intensity of DDs. As for LDs, we generated $n = 200$ samples of patients' arrival flows with $\lambda_p = 153/12$ as a fixed control variable together with two different percentages of patients with a willing LD. Regarding DDs, we independently generated the same number $n = 200$ samples of DD arrival flows with two different rates $\lambda_d = 108/24$ and $168/24$. We assumed that each DD provides both kidneys for transplantation, and we chose the values of parameter λ_d so as the obtained average monthly flows of organs correspond to the yearly minimum and maximum numbers of cadaveric transplantations that were performed in the last 10 years, as described in Section 3, see Table 1. After generating the samples, we continued with two sub-level factors (F3) and (F4) which means that we applied a corresponding allocation model for kidney transplantation with a policy given by levels in the design point.

For the fixed starting point of each simulation experiment representing time zero, we used the data of real patients that were registered in the national DDWL on May 18, 2019 in the following structure: the date of registration on the DDWL, blood group and the panel reactive antibody (PRA%) level. The number of these patients was 278. We simulated the evolution of the process in the timespan of 5 years.

The chosen number $n = 200$ of simulation runs was determined by a prior pilot simulation experiment. We started with a relatively small number of simulations ($n = 20$) which we subsequently increased iteratively to $n = 1000$. As the criterion for sufficient n recommended by Lorscheid et al. (2012), we employed the so-called coefficient of variation, the ratio of the standard deviation to the arithmetic mean which represents dimensionless, normalized measure of data variance. During increasing the number of runs n, we found that after 200 repetitions the coefficient of variation for all response variables stopped changing. In other words the variability of all response variables was stabilized for $n \geqq 200$ in one simulation experiment.

After performing $v = 32$ independent simulation experiments with settings given by particular design points, where every experiment contained $n = 200$ complex simulation runs (one run included one random sample of patients arrivals and one independent random sample of DD arrivals with subsequent steps of our model), we obtained $N = nv = 200 \times 32 = 6400$ complex simulation runs, which were summarized numerically by basic descriptive statistics measures.

The exploratory analysis was based on visualization of numerical summaries in the corresponding data histograms and plots of means (averages) with error bars given by standard deviations. The subsequent inferential statistical data analyses were connected to the addressed research questions. To test the statistical significance of main effects[1] given by all factors (F1)-(F4), we performed the standard four-way analysis of variance — ANOVA (Law,

[1] The main effect of a factor, shortly the effect of the factor, represents the average change in a response variable due to moving the factor from its "lower" to "higher" level while holding all other factors fixed (Law, 2014).

2014; Maxwell et al., 2017). Subsequently, for multiple comparisons of means leading to numerical values of main effects, Tukey's method was applied.

Generally, ANOVA requires three probability assumptions for responses in each designed point: sample independence, normality of the population distribution generating given samples and the same population variance (homogeneity of variance). Since the latter two assumptions usually may be violated in simulation modeling (Law, 2014), we also applied the nonparametric bootstrap with Efron percentil confidence intervals (Efron and Tibshirani, 1993; Chernick and LaBudde, 2014) as a cross-validation methodology.

Finally, in addition to evaluating the statistical significance of explored effects, we also calculated the practical significance of our ANOVA (bootstrap) results. As we compare effects within a single study, it is recommended to measure their effect size (ES for short), providing an indication of how large these effects are in reality (Lakens, 2013; Albers and Lakens, 2018). The two most common ES measures for ANOVA are η^2 and ω^2 or their nonparametric alternatives. However, since $\omega^2 = \eta^2 + \mathcal{O}(1/N)$, in case of a large number of data N it is sufficient to consider only η^2 with the following benchmarks: very small ($\eta^2 \leq 0.01$), small ($\eta^2 \in (0.01, 0.06]$), medium ($\eta^2 \in (0.06, 0.14]$) and large effect ($\eta^2 > 0.14$), suggested originally by Cohen (1988).

All statistical analyses with data manipulation and processing were realized in R (R Core Team, 2019) with libraries RcmdrMisc (Fox, 2020), sjstats (Lüdecke, 2020), reticulate (Ushey et al., 2019), dplyr (Wickham et al., 2019), in Scientific Python (Oliphant, 2007) with SciPy ecosystem (Jones et al., 2019) using libraries numpy (van der Walt et al., 2011), pandas (McKinney et al., 2010), matplotlib (Hunter, 2007) and cross-checked in SPSS software (IBM Corp., 2015).

5 Results

Concerning the number of incoming patients (the flow intensity $\lambda_p = 153/12$) for Slovakia during 5 years (including the starting real sample) — the fixed control variable determining the number of willing LDs (top-level F1), we observed in generated samples for all 32 performed simulation experiments, one for each design point, that the minimum number of incoming patients varied from 942 to 1098 and the maximum number from 990 to 1143.

Regarding the independently generated number of DDs (top-level F2) with respect to all design points, the flow rate level $\lambda_d = 108/24$ led to the minimum number of DDs arrivals from interval [423, 469] and maximum number from [615, 663]. Consequently, these flows resulted in the corresponding minimum number of cadaveric transplantations from [418, 467] and maximum number from [615, 651]. As for $\lambda_d = 165/24$, these numbers were from [687, 729] and [915, 975] for minimum and maximum DDs arrivals with corresponding intervals [651, 717] and [796, 960] for minimum and maximum cadaveric transplantations.

The central tendency and variability of both response variables, (R1) median of WT on the DDWL and (R2) transplantation probability for all patients, H patients and L patients with respect to all possible levels of (F1)-(F4) are visually summarized by plots of means with corresponding standard deviations as error bars in Figure 2 and Figure 3. Each point with error bars in every horizontal pair of plots is the graphical representation of the mean and SD for one of 32 design-point simulation experiments. The detailed numerical values of these means and standard deviations for all considered designed points can be found in Tables 4 and 5 in Appendix.

As for the main effect of the percentage of LDs (F1), DDs intensity (F2), DDs allocation policy (F3) and allocation model (F4), the 4-way ANOVA proved that main effects of all four factors are statistically significant ($p < 0.001$), affecting the median of WT (R1) for all patients — for factor DD $F(1, 6368) = 22085.5$; for factor LD $F(1, 6368) = 1527.3$; for factor allocation policy $F(1, 6368) = 642.2$ and for factor model $F(3, 6368) = 2237.2$. It is worth to mention that the numerical ANOVA results for KT probability (R2) were very similar.

Although the effects of all factors are statistically significant, which simply means that these effects are larger than expected by random sampling fluctuations, from the viewpoint of practical significance, in case of all patients, our simulation results showed that only DDs arrival intensity (F2) and allocation model (F4) have a real, practically significant effect for median of WT (R1) and KT probability of patients (R2). The remaining two factors appear to be practically irrelevant.

In case of patients' subgroups – L patients and H patients, all the effects are statistically significant with the exception of the effect of DDs kidney allocation policy (F3) on the median WT in the group of L patients. Dealing with L patients, practically significant are again only effects of DDs arrival intensity (F2) and allocation model (F4) on both response variables. Regarding H patients, in case of both response variables, the most practically important is the effect of DDs allocation policy (F3) and less relevant but still practically important is the effect of the allocation model for LDs (F4).

The numerical values of the main effects on both response variables with the corresponding p-values, 95% confidence intervals and effect sizes η^2 for all groups of patients are shown in Table 3. Regarding interactions, plots of means also indicate possible interactions between factors as it was confirmed by ANOVA in several cases. But from the viewpoint of practical significance, the effect size of all interactions was very small or small ($\eta^2 < 0.015$) in comparison to main effects, which was the reason not to list them.

Here it is important to say that the main effects of all factors in this table are additive thanks to linearity of the ANOVA model. This means, for example in the case of waiting time for all patients (the first column of Table 3), that the main effect of F2 – DDs arrival intensity (shortening median

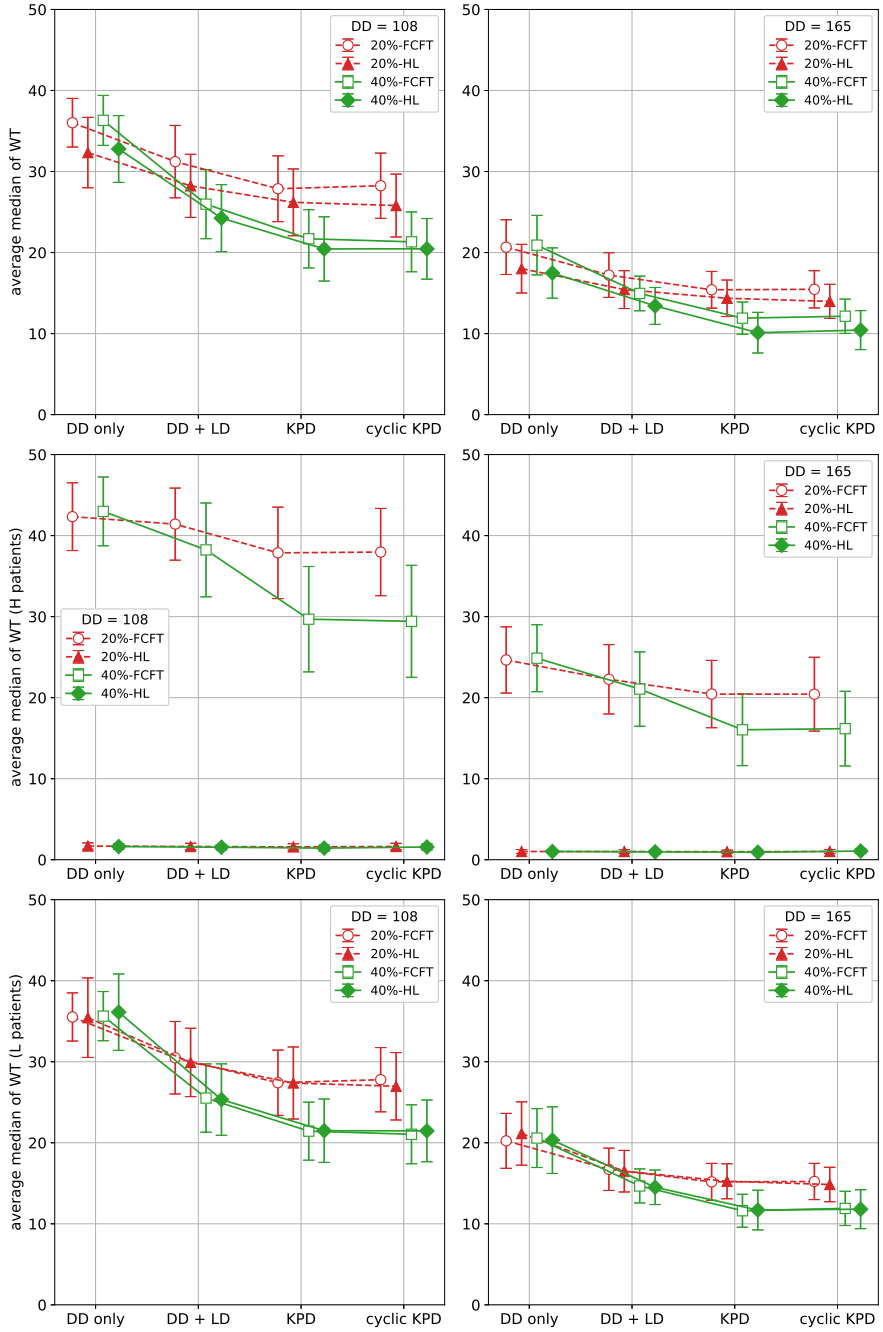

Fig. 2 Average median of transplantation waiting time for all, H and L patients.

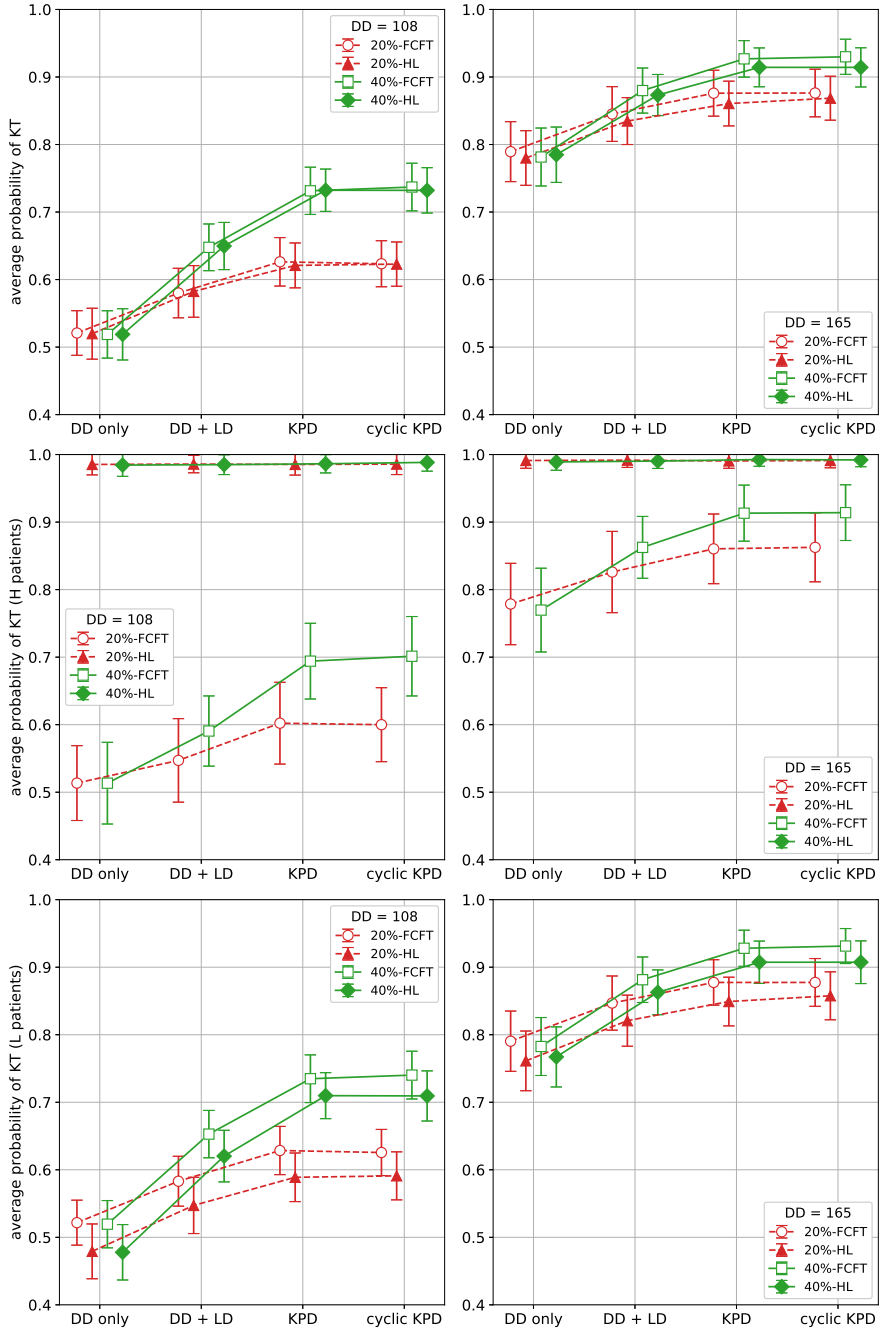

Fig. 3 Average probability of transplantation for all, H and L patients.

Table 3 Multiple comparisons between factor levels for median WT and KT probability (p value, statistical significance of main effect for factors — ***: $p < 0.001$, n.s.: not significant; effect size η^2, practical significance of main effect for factors — L: large, M: medium, S: small, VS: very small)

	Factor	Median WT (in months)			Median WT (L patients)			Median WT (H patients)		
		mean (95% CI)	p	η^2	mean (95% CI)	p	η^2	mean (95% CI)	p	η^2
DD	165 – 108	-12.33 (-12.49, -12.17)	***	L	-12.30 (-12.47, -12.13)	***	L	-8.67 (-8.82, -8.52)	***	L
LD	40 – 20	-3.24 (-3.41, -3.08)	***	S	-3.18 (-3.35, -3.01)	***	S	-1.84 (-2.00, -1.68)	***	VS
P	HL – FCFT	-2.10 (-2.27, -1.94)	***	S	-0.05 (-0.22, 0.12)	n.s.	VS	-27.81 (-28.00, -27.63)	***	L
M	DD+LD vs DD only	-5.48 (-5.78, -5.18)	***	L	-6.42 (-6.73, -6.10)	***	L	-1.50 (-1.73, -1.27)	***	VS
	KPD vs DD only	-8.31 (-8.62, -8.01)	***		-9.19 (-9.51, -8.88)	***		-3.91 (-4.13, -3.68)	***	
	C-KPD vs DD only	-8.33 (-8.64, -8.03)	***		-9.24 (-9.56, -8.93)	***		-3.86 (-4.10, -3.62)	***	
	KPD vs DD+LD	-2.84 (-3.14, -2.54)	***		-2.78 (-3.09, -2.46)	***		-2.40 (-2.65, -2.16)	***	
	C-KPD vs DD+LD	-2.86 (-3.16, -2.56)	***		-2.82 (-3.14, -2.51)	***		-2.36 (-2.62, -2.09)	***	
	C-KPD vs KPD	-0.02 (-0.32, 0.28)	n.s.		-0.05 (-0.36, 0.27)	n.s.		0.04 (-0.21, 0.30)	n.s.	

	Factor	KT probability			KT Probability (L patients)			KT probability (H patients)		
		mean (95% CI)	p	η^2	mean (95% CI)	p	η^2	mean (95% CI)	p	η^2
DD	165 – 108	0.236 (0.234, 0.237)	***	L	0.245 (0.243, 0.247)	***	L	0.129 (0.127, 0.131)	***	M
LD	40 – 20	0.053 (0.051, 0.055)	***	S	0.055 (0.054, 0.057)	***	S	0.023 (0.021, 0.025)	***	VS
P	HL – FCFT	-0.005 (-0.007, -0.003)	***	VS	-0.029 (-0.031, -0.027)	***	VS	0.267 (0.265, 0.268)	***	L
M	DD+LD vs DD only	0.085 (0.082, 0.088)	***	L	0.089 (0.086, 0.093)	***	L	0.032 (0.029, 0.035)	***	S
	KPD vs DD only	0.134 (0.131, 0.138)	***		0.141 (0.137, 0.144)	***		0.063 (0.060, 0.065)	***	
	C-KPD vs DD only	0.136 (0.133, 0.139)	***		0.143 (0.139, 0.146)	***		0.064 (0.061, 0.067)	***	
	KPD vs DD+LD	0.049 (0.046, 0.053)	***		0.051 (0.048, 0.054)	***		0.031 (0.028, 0.033)	***	
	C-KPD vs DD+LD	0.051 (0.048, 0.055)	***		0.053 (0.050, 0.056)	***		0.032 (0.029, 0.035)	***	
	C-KPD vs KPD	0.002 (-0.001, 0.005)	n.s.		0.002 (-0.001, 0.005)	n.s.		0.001 (-0.001, 0.004)	n.s.	

WT by 12.3 months) adds to the main effect of F4 – allocation model (shortening median WT by 5.5 months in the case of at least 20% patients with LD without KPD or by 8.3 months with KPD or C-KPD) giving overall shortening by 18.8 – 20.6 months. If we consider the main effects of factors with practically small size, namely LDs percentage (-3.2 months) and DDs allocation policy (-2.1 months), this would shorten the WT by another 5.3 months leading to total median WT shortening by 25.9 months.

On the other hand, as the middle pairs of plots of means in Figure 2 and Figure 3 evidently indicate, in the case of H patients the third ANOVA assumption, homogeneity of variance, is almost certainly violated. Therefore in Table 3 the results for H patients came from the applied non-parametric bootstrap with the non-parametric version of effect size η^2.

Histograms of ANOVA residuals' distributions with Q-Q plots and Shapiro-Wilk tests also showed that (R1) and (R2) response samples came from symmetrical, but not necessarily normal distributions. However, from the practical point of view, if we compare ANOVA and bootstrap results, we find that all corresponding results in Table 3 are pretty close, even in the case of H patients. Therefore, it is clear that our simulation study represents an example demonstrating the well-known robustness of ANOVA testing with respect to certain violations of normality or variance homogeneity (Norman, 2010).

Finally, we also present summarizing numerical results of main effects for (F1), (F2), (F3) on response variables with respect to fixed levels of model (F4) calculated by three-way ANOVA, see Table 6 in Appendix.

6 Discussion and conclusions

Besides of medical and logistical issues associated with kidney transplantations that have to be solved, there are also many important ethical, religious and legal questions which have to be answered. Should the first aim of the transplantation policy be to help as many patients as possible, or should some patients be given preference, the patients who wait longest, those that have longer life expectancy or those whose condition is worst? Will it not be the case that prioritizing some will harm the others or the system as a whole? Is it right to ask a healthy person to undergo the risk of removing an organ when perhaps in a near or more distant future a deceased donor might be available?

Mathematical modelling can help to weigh various pros and cons in facilitating decision making connected with answers to these questions. It also has the power to visualize and quantify possible outcomes and the significance of various scenarios before they have to be actually observed in reality.

In this paper we have presented the what-if analysis in the form of a complex simulation study based on actual Slovak medical data using stochastic modeling and the systematic design of simulation experiments to see what happens in the waiting list for transplantation if various parameters and policies are changed.

After careful statistical analysis of our simulation study results, performed and confirmed by two different statistical methodologies (ANOVA and bootstrap), we can draw some important conclusions or recommendations for the medical community. These recommendations are based on the estimated average influence of the considered parameters and policies (represented by four factors) on two response variables — the waiting time in the DDWL in Slovakia and the corresponding kidney transplantation probability.

Our results demonstrate that the largest and most counter-intuitive effect on waiting time and transplantation probability was obtained by the change in the currently implemented FCFT allocation policy to a policy that prioritizes the most vulnerable group of patients — highly sensitized patients (8% of all patients population). This change has led to shortening the waiting time of H patients by enormous 28 months in average to the final waiting time shorter than two months. Simultaneously it increases the transplantation probability by average percentage points 27% to final 99%.

At first glance, prioritizing H patients over remaining 92% L patients may appear as highly unethical. But simulations clearly show that the waiting time and transplantation probability of L patients stay practically the same as the original ones under FCFT policy. Since such a change needs no medical or logistic decisions, this important observation may serve as a starting point for a discussion about such a modification of the current Slovak legislation.

The number of deceased donors (DDs) has the second greatest impact. Increasing the number of DDs from the observed Slovak historical minimum (108 kidneys per year) to the Slovak historical maximum (165 kidneys per year) leads to the decrease of waiting time by 12 months in average and increase of transplantation probability by 0.24 for all patients.

Unlike the legislative and in some sense ethical change in the applied policy, increasing the number of kidneys from deceased donors depends mainly on improved education and devotion of doctors working at intensive care units so that they can correctly identify, examine and treat potential donors. Of course, even though Slovakia has an opt-out system for organ donation, education of public towards a positive attitude towards cadaveric donation is also essential.

It is worth to mention that although approximately 168 DDs arriving yearly exceed the yearly average numbers of new patients, the transplantation probability is still only about 0.78. This can be significantly changed by the presence of willing living donors.

In particular, the presence of LDs additively decreases waiting time for patients at least by another 5 months without KPD program and 8 months with KPD program, leading to total 17 – 20 months decrease in waiting time which represents shortening the total waiting time by almost one half.

Similarly, the presence of given factors increases the probability at least by 0.32 without KPD and 0.37 with KPD to the final average of 0.88. It can also be seen that considering C-KPD, which is logistically more challenging, leads to no difference in comparison with simple KPD. One explanation of this lack of difference between KPD and C-KPD may be small size of patients population in Slovakia.

Our simulations also showed another important result. The significant increase of the proportion of patients with a willing living donor from 20% to 40% which might be an unrealistic demand, brings practically only a small effect on both response variables (-3 months for waiting time and +0.05 for transplantation probability). However, without a reasonable number of LDs, in our simulations it is at least 20%, there is no room for the improvement. What may be very frustrating for the patient and those who love him/her so much that they are offering a part of their body to help to lessen his/her suffering, is the fact that even the blood groups incompatibility, not yet considering the positive cross-match possibility, may prevent almost one half of intended donations. Therefore, we think that introducing a program of kidney paired donations should be considered despite of relatively small 3 month reduction of the WT and 5% increase of transplantation probability. Again, education, encouragement and trust of patients in the healthcare system are the key factors.

Finally, it is important to emphasize that our stochastic modelling and the obtained simulation results have naturally some limitations. For example 99% transplantation probability of H patients under HL-policy seems to be unrealistic given the real clinical experience. Still, one can expect that HL-policy can substantially increase the transplantation probability of hard-to-match patients.

We believe that the main sources of model limitations stem from the fact that some real-life medical data is not collected or recorded, or because the size of the modelled population is too small to enable modelling some features more precisely. For example, we do not know in detail the time from entering the DDWL till the patient's condition deteriorates to such a state when no transplantation is possible or when he/she dies. We do not have sufficient statistical data about the survival time on the DDWL or after transplantation that would enable us to include some considerations of life expectancy of waiting and transplanted patients into the design of some other allocation policies. Also, because of the very small patients' population it is almost impossible to realistically model the distribution of HLA antigens or age of patients.

Acknowledgements The authors would like to thank Daniel Kuba and Magdaléna Krátka, National Transplant Organization, for providing the details of the composition of the Slovak DDWL. Special thanks go to Jozef Hanč, Institute of Physics, P.J. Šafárik University, for his essential help in simulation data preprocessing and cross-validation for the subsequent statistical analysis and efficient bootstrap computations, using Scientific Python.

Conflict of interest

The authors declare that they have no conflict of interest.

References

J. Abellán, C. Armero, D. Conesa, J. Pérez-Panadés, M. Martinez-Beneito, O. Zurriaga, M. J. Garcia-Blasco, and H. Vanaclocha. Analysis of the renal transplant waiting list in the país Valenciá (Spain). *Stat.Med.*, 25:345–358, 2006.

C. Albers and D. Lakens. When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, 74:187–195, 2018.

I. Ashlagi, A. Bingaman, M. Burq, V. Manshadi, D. Gamarnik, C. Murphey, A. E. Roth, M. L. Melcher, and M. A. Rees. Effect of match-run frequencies on the number of transplants and waiting times in kidney exchange. *Am J Transplant*, 18:1177–1186, 2018.

P. Biró, B. Haase-Kromwijk, T. Anderson, and et.al. Building kidney exchange programmes in Europe — an overview of exchange practice and activities. *Transplantation*, 103, 2019.

M. R. Chernick and R. A. LaBudde. *An introduction to bootstrap methods with applications to R.* Wiley, Hoboken, 2014.

J. Cohen. *Statistical Power Analysis for the Behavioral Sciences.* Routledge, Hillsdale, N.J, 2nd edition, 1988.

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap.* Chapman and Hall/CRC, New York, 1993.

P. Ferrari, S. Fidler, C. Woodroffe, G. Tassone, and L. D'Orsogna. Comparison of time on the deceased donor kidney waitlist versus time on the kidney paired donation registry in the Australian program. *Transplant International*, 25:1026–1031, 2012.

J. Fox. *RcmdrMisc: R Commander Miscellaneous Functions*, 2020. URL `https://CRAN.R-project.org/package=RcmdrMisc`. R package version 2.7-0.

P. F. Halloran. Immunosuppressive drugs for kidney transplantation. *New England Journal of Medicine*, 351(26):2715–2729, 2004.

J. D. Hunter. Matplotlib: A 2d Graphics Environment. *Computing in science & engineering*, 9(3):90–95, 2007.

IBM Corp. IBM SPSS Statistics for Windows, version 23.0. Armonk, NY: IBM Corp., 2015.

E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2019. URL `http://www.scipy.org/`.

J. H. Kim. Features and ethical considerations associated with living kidney and liver transplantations in South Korea. *Transplantation Proceedings*, 46: 3647–3652, 2014.

D. Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.*, 4, 2013.

A. Law. *Simulation Modeling and Analysis.* McGraw-Hill Education, 5th ed. edition, 2014.

I. Lorscheid, B.-O. Heine, and M. Meyer. Opening the 'black box' of simulations: increased transparency and effective communication through the

systematic design of experiments. *Comput Math Organ Theory*, 18(1):22–62, 2012. ISSN 1572-9346.

D. Lüdecke. *sjstats: Statistical Functions for Regression Models (Version 0.17.8)*, 2020. URL `https://CRAN.R-project.org/package=sjstats`.

S. E. Maxwell, H. D. Delaney, and K. Kelley. *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Routledge, 3rd edition, 2017.

W. McKinney et al. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

National Blood Service in Slovak Republic. *(Národná transfúzna služba)*, 2019. Accessed: 15 June 2019, URL `http://www.nto.sk/`.

National Health Information Center. *(Národné centrum zdravotníckych informácií Slovenskej republiky), Health Statistics Yearbooks*, 2019.

National Transplant Organization. *(Národná transplantačná organizácia Slovenskej republiky)*, 2019. Accessed: 15 June 2019, URL `http://www.nto.sk/`.

G. Norman. Likert scales, levels of measurement and the "laws" of statistics. *Adv in Health Sci Educ*, 15(5):625–632, 2010.

T. E. Oliphant. Python for Scientific Computing. *Comput Sci Eng*, 9(3):10–20, 2007.

Y. Perlman, A. Elalouf, and U. Yechiali. Dynamic allocation of stochastically-arriving flexible resources to random streams of objects with application to kidney cross-transplantation. *European Journal of Operational Research*, 265:169–177, 2018.

R Core Team. R: A language and environment for statistical computing, 2019. URL `http://www.r-project.org/`.

F. Rapaport. The case for a living emotionally related international kidney donor exchange registry. *Transplant Proceedings*, 18(3):5–9, 1986.

N. Santos, P. Tubertini, A. Viana, and J. Pedroso. Kidney exchange simulation and optimization. *Journal of the Operational Research Society*, 68:1521–1532, 2017.

D. L. Segev, S. E. Gentry, J. K. Melancon, and R. A. Montgomery. Characterization of waiting times in a simulation of kidney-paired donation. *American Journal of Transplantation*, 5:2248–2455, 2005a.

D. L. Segev, S. E. Gentry, D. S. Warren, B. Reeb, and R. A. Montgomery. Kidney paired donation and optimizing the use of live donor organs. *JAMA*, 293:1883–1890, 2005b.

D. A. Stanford, J. M. Lee, N. Chandok, and V. C. McAlister. A queueing model to address wait time inconsistency in solid-organ transplantation. *Oper. Res. Health Care*, 3:40–45, 2014.

P. I. Terasaki, J. M. Cecka, D. W. Gjertson, and S. Takemoto. High survival rates of kidney transplants from spousal and living unrelated donors. *New England Journal of Medicine*, 333:333–336, 1995.

K. Ushey, J. Allaire, and Y. Tang. *reticulate: Interface to 'Python'*, 2019. URL `https://CRAN.R-project.org/package=reticulate`. R package version

1.14.
S. van der Walt, S. C. Colbert, and G. Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput Sci Eng*, 13(2): 22–30, 2011.
H. Wickham, R. François, L. Henry, and K. Müller. *dplyr: A Grammar of Data Manipulation*, 2019. URL `https://CRAN.R-project.org/package=dplyr`. R package version 0.8.3.
S. A. Zenios. Modeling the transplant waiting list: A queueing model with reneging. *Queueing Systems*, 31:239–251, 1999.
S. A. Zenios. Optimal control of a paired-kidney exchange program. *Management Science*, 48:328–342, 2002.
S. A. Zenios, L. M. Wein, and G. M. Chertow. Evidence-based organ allocation. *The American Journal of Medicine*, 107:52–61, 1999.

Appendix

Here we present supplementary tables of results collected in our simulation study. Tables 4 and 5 are graphically visualized in plots (Fig. 2 and Fig. 3). Each cell in Tab. 4 and 5 for any particular group of patients represents the numerical summary by the mean and SD for one of 32 design-point simulation experiments.

Table 4 Median waiting time of transplanted patients on the DDWL (in months) during 5 years.

DD	LD - Policy	Patients	DD only Mean (SD)	DD + LD Mean (SD)	KPD Mean (SD)	cyclic KPD Mean (SD)
108 yearly	20% FCFT	All	36.0 (3.0)	31.2 (4.5)	27.9 (4.1)	28.2 (4.0)
		L	35.5 (3.0)	30.5 (4.5)	27.4 (4.0)	27.8 (4.0)
		H	42.3 (4.2)	41.4 (4.4)	37.9 (5.6)	38.0 (5.4)
	20% HL	All	32.3 (4.3)	28.2 (3.9)	26.2 (4.1)	25.8 (3.9)
		L	35.4 (4.9)	29.9 (4.2)	27.4 (4.4)	27.0 (4.2)
		H	1.7 (0.4)	1.6 (0.4)	1.6 (0.4)	1.6 (0.4)
	40% FCFT	All	36.3 (3.1)	26.0 (4.3)	21.7 (3.6)	21.3 (3.7)
		L	35.6 (3.0)	25.5 (4.2)	21.4 (3.6)	21.1 (3.6)
		H	43.0 (4.2)	38.2 (5.8)	29.7 (6.5)	29.4 (6.9)
	40% HL	All	32.8 (4.1)	24.2 (4.1)	20.4 (4.0)	20.5 (3.7)
		L	36.1 (4.7)	25.3 (4.4)	21.5 (3.9)	21.5 (3.8)
		H	1.6 (0.3)	1.5 (0.3)	1.4 (0.4)	1.6 (0.4)
165 yearly	20% FCFT	All	20.7 (3.4)	17.2 (2.8)	15.4 (2.3)	15.5 (2.3)
		L	20.2 (3.4)	16.7 (2.6)	15.2 (2.3)	15.2 (2.2)
		H	24.7 (4.1)	22.3 (4.3)	20.4 (4.2)	20.4 (4.6)
	20% HL	All	18.0 (3.0)	15.4 (2.3)	14.4 (2.3)	14.0 (2.1)
		L	21.1 (3.9)	16.5 (2.6)	15.3 (2.2)	14.9 (2.1)
		H	1.0 (0.2)	1.0 (0.2)	1.0 (0.2)	1.0 (0.2)
	40% FCFT	All	20.9 (3.7)	15.0 (2.1)	11.9 (2.0)	12.1 (2.1)
		L	20.6 (3.6)	14.7 (2.1)	11.6 (2.0)	11.9 (2.1)
		H	24.9 (4.1)	21.1 (4.6)	16.0 (4.4)	16.2 (4.6)
	40% HL	All	17.5 (3.1)	13.4 (2.3)	10.1 (2.5)	10.4 (2.4)
		L	20.3 (4.1)	14.5 (2.1)	11.7 (2.5)	11.8 (2.4)
		H	1.0 (0.2)	1.0 (0.2)	0.9 (0.2)	1.1 (0.3)

Table 5 Probability of kidney transplantation on DDWLs during 5 years.

DD	LD - Policy	Patients	DD only Mean (SD)	DD + LD Mean (SD)	KPD Mean (SD)	cyclic KPD Mean (SD)
108 yearly	20% FCFT	All	0.52 (0.03)	0.58 (0.04)	0.63 (0.04)	0.62 (0.03)
		L	0.52 (0.03)	0.58 (0.04)	0.63 (0.04)	0.63 (0.03)
		H	0.51 (0.06)	0.55 (0.06)	0.60 (0.06)	0.60 (0.05)
	20% HL	All	0.52 (0.04)	0.58 (0.04)	0.62 (0.03)	0.62 (0.03)
		L	0.48 (0.04)	0.55 (0.04)	0.59 (0.04)	0.59 (0.04)
		H	0.99 (0.02)	0.99 (0.01)	0.99 (0.02)	0.99 (0.02)
	40% FCFT	All	0.52 (0.04)	0.65 (0.03)	0.73 (0.03)	0.74 (0.04)
		L	0.52 (0.04)	0.65 (0.04)	0.73 (0.04)	0.74 (0.04)
		H	0.51 (0.06)	0.59 (0.05)	0.69 (0.06)	0.70 (0.06)
	40% HL	All	0.52 (0.04)	0.65 (0.03)	0.73 (0.03)	0.73 (0.03)
		L	0.48 (0.04)	0.62 (0.04)	0.71 (0.03)	0.71 (0.04)
		H	0.98 (0.02)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)
165 yearly	20% FCFT	All	0.79 (0.04)	0.85 (0.04)	0.88 (0.03)	0.88 (0.04)
		L	0.79 (0.04)	0.85 (0.04)	0.88 (0.03)	0.88 (0.04)
		H	0.78 (0.06)	0.83 (0.06)	0.86 (0.05)	0.86 (0.05)
	20% HL	All	0.78 (0.04)	0.83 (0.03)	0.86 (0.03)	0.87 (0.03)
		L	0.76 (0.04)	0.82 (0.04)	0.85 (0.04)	0.86 (0.04)
		H	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)
	40% FCFT	All	0.78 (0.04)	0.88 (0.03)	0.93 (0.03)	0.93 (0.03)
		L	0.78 (0.04)	0.88 (0.03)	0.93 (0.03)	0.93 (0.03)
		H	0.77 (0.06)	0.86 (0.05)	0.91 (0.04)	0.91 (0.04)
	40% HL	All	0.78 (0.04)	0.87 (0.03)	0.91 (0.03)	0.91 (0.03)
		L	0.77 (0.04)	0.86 (0.03)	0.91 (0.03)	0.91 (0.03)
		H	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)

Table 6 Multiple comparisons between factor levels for median WT and KT probability with respect to a fixed level model computed from three-way ANOVA and bootstrap (p value, statistical significance of main effect for factors — ***: $p < 0.001$, n.s.: not significant)

Model		Factor	Median WT (in months)		Median WT (L patients)		Median WT (H patients)	
			mean (95% CI)	p	mean (95% CI)	p	mean (95% CI)	p
DD only	DD	165 – 108	-15.09 (-15.44, -14.75)	***	-15.10 (-15.49, -14.72)	***	-9.26 (-9.53, -8.98)	***
	LD	40 – 20	0.11 (-0.23, 0.45)	n.s.	0.08 (-0.30, 0.46)	n.s.	0.20 (-0.08, 0.48)	n.s.
	P	HL – FCFT	-3.33 (-3.67, -2.98)	***	0.26 (-0.12, 0.64)	n.s.	-32.37 (-32.65, -32.09)	***
DD + LD	DD	165 – 108	-12.16 (-12.50, -11.83)	***	-12.21 (-12.56, -11.87)	***	-9.37 (-9.70, -9.02)	***
	LD	40 – 20	-3.38 (-3.71, -3.04)	***	-3.40 (-3.74, -3.06)	***	-1.14 (-1.45, -0.82)	***
	P	HL – FCFT	-2.01 (-2.35, -1.68)	***	-0.29 (-0.63, 0.05)	n.s.	-29.45 (-29.83, -29.10)	***
KPD	DD	165 – 108	-11.11 (-11.43, -10.79)	***	-10.99 (-11.31, -10.67)	***	-8.05 (-8.43, -7.65)	***
	LD	40 – 20	-4.92 (-5.23, -4.60)	***	-4.74 (-5.06, -4.43)	***	-3.20 (-3.55, -2.82)	***
	P	HL – FCFT	-1.44 (-1.76, -1.13)	***	0.05 (-0.27, 0.36)	n.s.	-24.76 (-25.11, -24.41)	***
C - KPD	DD	165 – 108	-10.96 (-11.26, -10.65)	***	-10.87 (-11.19, -10.56)	***	-7.98 (-8.33, -7.61)	***
	LD	40 – 20	-4.78 (-5.09, -4.48)	***	-4.65 (-4.96, -4.34)	***	-3.21 (-3.62, -2.81)	***
	P	HL – FCFT	-1.63 (-1.93, -1.32)	***	-0.22 (-0.53, 0.10)	n.s.	-24.67 (-25.02, -24.28)	***

Model		Factor	KT probability		KT probability (L patients)		KT probability (H patients)	
			mean (95% CI)	p	mean (95% CI)	p	mean (95% CI)	p
DD only	DD	165 – 108	0.264 (0.261, 0.268)	***	0.276 (0.272, 0.280)	***	0.133 (0.129, 0.137)	***
	LD	40 – 20	-0.002 (-0.005, 0.002)	n.s.	-0.001 (-0.005, 0.003)	n.s.	-0.003 (-0.007, 0.001)	n.s.
	P	HL – FCFT	-0.002 (-0.006, 0.002)	n.s.	-0.032 (-0.036, -0.028)	***	0.344 (0.340, 0.348)	***
DD + LD	DD	165 – 108	0.243 (0.240, 0.247)	***	0.252 (0.248, 0.256)	***	0.141 (0.137, 0.144)	***
	LD	40 – 20	0.052 (0.049, 0.056)	***	0.055 (0.051, 0.059)	***	0.020 (0.016, 0.023)	***
	P	HL – FCFT	-0.003 (-0.007, 0.000)	n.s.	-0.028 (-0.032, -0.025)	***	0.282 (0.278, 0.286)	***
KPD	DD	165 – 108	0.217 (0.214, 0.220)	***	0.225 (0.222, 0.228)	***	0.122 (0.118, 0.126)	***
	LD	40 – 20	0.080 (0.077, 0.083)	***	0.084 (0.081, 0.087)	***	0.037 (0.033, 0.041)	***
	P	HL – FCFT	-0.008 (-0.011, -0.005)	***	-0.028 (-0.032, -0.025)	***	0.221 (0.217, 0.225)	***
C - KPD	DD	165 – 108	0.218 (0.215, 0.222)	***	0.227 (0.224, 0.230)	***	0.121 (0.118, 0.125)	***
	LD	40 – 20	0.081 (0.077, 0.084)	***	0.084 (0.081, 0.087)	***	0.039 (0.035, 0.043)	***
	P	HL – FCFT	-0.007 (-0.010, -0.004)	***	-0.027 (-0.031, -0.024)	***	0.220 (0.216, 0.223)	***