Central European Journal of Operations Research
https://doi.org/10.1007/s10100-024-00914-3

Performance evaluation of national and international kidney exchange programmes with the ENCKEP simulator

Kristóf Druzsin[1,2] · **Péter Biró**[1,3] · **Xenia Klimentova**[4] · **Rita Fleiner**[2]

Accepted: 13 March 2024

Abstract
In this paper we present simulations for international kidney exchange programmes (KEPs). KEPs are organised in more than ten countries in Europe to facilitate the exchanges of immunologically incompatible donors. The matching runs are typically conducted in every three months for finding optimal exchanges using hierarchical optimisation with integer programming techniques. In recent years several European countries started to organise international exchanges using different collaboration policies. In this paper we conduct simulations for estimating the benefits of such collaborations with a simulator developed by the team of the ENCKEP COST Action. We conduct our simulations on generated datasets mimicking the practice of the three largest KEPs in Europe, the UK, Spanish and the Dutch programmes. Our main performance measure is the number of transplants compared to the number of registrations to the KEP pools over a 5-year period, however, as a novelty we also analyse how the optimisation criteria play a role in the lexicographic and weighted optimisation policies for these countries. Besides analysing the performances on a single instance, we also conduct large number of simulations to obtain robust findings on the performance of specific national programmes and on the possible benefits of international collaborations.

Keywords Kidney exchange · Integer programming · Hierarchical optimisation · Simulation

1 Introduction

Patients with the end-stage renal disease can be treated by dialysis, but their quality of life is poor and their life expectancy is short. The only long-term solution according to our knowledge is transplantation. One can get a kidney from a deceased donor, but the demand is very high and waiting lists are long even in the developed

In this paper we summarise the main findings of our conference papers (Druzsin et al. 2021) and Druzsin et al. (2022).

Extended author information available on the last page of the article

Published online: 10 May 2024

world (over 100,000 patients are on the US waiting list, with an average waiting time of 8–10 years). Therefore living donation became a common practice, also due to the longer graft survival rates. However, if someone has a willing, but immunologically incompatible donor then transplantation is not possible. To resolve this issue, kidney exchange programmes (KEPs) have been established in many countries to facilitate the exchanges of the donors. The largest ones in Europe are in the UK, the Netherlands, and Spain.

In a typical national KEP in Europe the matching runs are conducted in every three months. First the ABO and HLA-compatibility is tested in a so-called virtual compatibility testing phase for every donor and recipient based on their individual immunological data in order to identify the potential transplants (see more about virtual compatibility testing in [4]). These potential transplants are represented by the arcs of the so-called virtual compatibility graph, where the nodes represent the donor-recipient pairs. An exchange between several pairs is organised in a way that each recipient receives a kidney from a compatible donor from another pair. In the virtual compatibility graph such exchanges are represented by cycles. In most of the European programmes altruistic donation is also possible, where the first transplant is given by an altruistic donor and the altruistic chain terminates by the last donor giving his/her kidney to the waiting list for deceased organs.

In order to mitigate the risk that an exchange cycle would break and possibly leave a recipient without a transplant whose donor already donated, the transplant operations in the exchange cycles are conducted simultaneously. In the European practices the altruistic chains are also selected and performed as part of the quarterly matching rounds, and the operations are also simultaneous in the chains in most European KEPs. Due to the simultaneity of the exchanges the length of the exchange cycles and altruistic chains is limited. For example, only 2- and 3-way exchanges are allowed in the UK and Spain, whilst 4-way exchanges are also possible in the Netherlands. The goal of the KEPs is to find and implement optimal exchanges for the pool of registered patient-donor pairs in the regular matching runs. The European practices have been surveyed in Biró et al. (2019) and the optimisation aspects of the European KEPs were described in Biró et al. (2021), as the results of the COST Action European Network for Collaboration on Kidney Exchange Programmes (ENCKEP).

In the core of the KEPs, matching algorithms are used to compute optimal solutions under various optimisation criteria. The basic Integer Programming (IP) techniques were first described in Abraham et al. (2007) for finding the maximum number of transplants. In the main European applications hierarchical criteria are used, where the scores are only used as tiebreakers. For the most advanced IP optimisation for hierarchical criteria see ([9]). The fragmentation of the US kidney exchange programmes have been analysed in Agarwal et al. (2019), where the potential benefits of merging these programmes have been calculated. Our study is similar in spirit, just focused on the European context and using dynamic simulations for a time interval. For further literature on OR analyses of KEPs, we recommend a survey on kidney exchange simulators (Santos et al. 2017), and a recent summary on OR perspectives for KEPs (Ashlagi and Roth 2021).

After computing an optimal solution in a matching run for the virtual compatibility graph, the planned transplants in the selected exchange cycles and chains are tested in the laboratory. A positive crossmatch found in the lab, or other failure (such as the sickness of a patient) can cause the cancellation of the corresponding cycle or chain. If the timeline permits, re-optimisation can be conducted. Internal recourse is a special re-optimisation strategy, when an alternative solution is searched within each failed cycle or chain. A particular strategy used in the UK and Spanish KEPs is to prioritise 3-cycles with embedded two-cycles in the matching run solution in order to make their internal recourse strategy more successful. These probabilistic features are all implemented in the ENCKEP-simulator, where the user can set probabilities for arc or node failures in a generated instance of a KEP, internal recourse can be conducted, and the maximisation of the number of 3-cycles with embedded 2-cycles is a criterion that can be chosen in the optimisation policy.

To the best of our knowledge, there exist four international KEPs, and three of them are running in Europe. The first one involves hospitals in Czech Republic and Austria (Böhmig et al. 2017), and recently, this collaboration has been extended with Israeli hospitals. The second programme, called KEPSAT, engages national programs of Spain, Italy and Portugal (Valentín et al. 2019). The third one, STEP, is run by Scandiatransplant, the deceased organ sharing organisation for Denmark, Finland, Iceland, Norway, Sweden and Estonia. Finally, the only non-European programme is run between Australia and New Zealand, called ANZKX.

The recent Handbook (Klimentova 2021) of Working Group 3 and 4 of the ENCKEP COST Action described the practice of international KEPs in Europe, some modelling frameworks of international KEPs including results from Klimentova et al. (2021) and Mincu et al. (2021), and the new simulation and evaluation tools developed by these working groups. In this paper we illustrate the usage of the ENCKEP-simulator tool (Klimentova 2021) by a case study with generated data for the three largest KEPs currently operating in Europe, namely the national programmes of the UK, the Netherlands and Spain. We simulated 5 years of operation of each programme, with 3-month intervals between matching runs, the setting used in practice for most European KEPs. In order to obtain robust results, we conducted 10 simulations for large instances and 100 simulations for medium size instances for each collaboration policy considered. Therefore, we had $20 \times 10 = 200$ matching runs in our simulations for large instances and $20 \times 100 = 2000$ matching runs for medium size instances in total.

The remainder of the paper is organised as follows. In Sect. 2 we briefly introduce the ENCKEP-simulator that we used to conduct our simulations. Next, in Sect. 4, 5 and 6 we describe our results for national matching runs for the UK, Spain and the Netherlands, respectively. For each country, first we describe the basic characteristics of the national KEPs and the lexicographic and weighted optimisation criteria used. Afterwards we present the illustrative results of one simulation for a single, large size, generated instance, and then the findings of our robust simulations conducted for 10 large and 100 medium size instances for each country. In Sect. 7 we analyse the results of the international KEP simulations conducted for these three countries under three different collaboration policies.

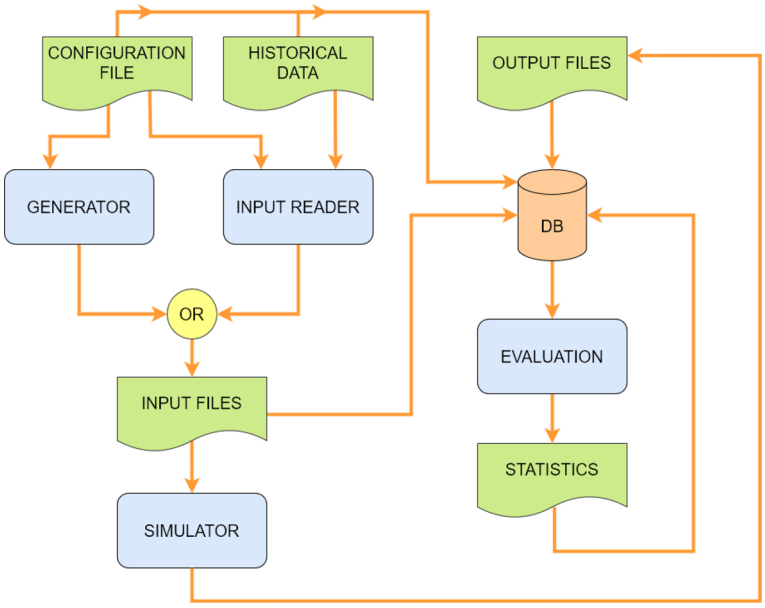


Fig. 1 Workflow of the ENCKEP simulator

2 ENCKEP simulator

The ENCKEP simulator (Klimentova 2021) is based on a standard technique of generating historical dataset for a period of time (e.g., five years) and conducting matching runs in regular time intervals (e.g., in every three months). Each pair in the generated dataset is provided with the timestamps of arrival, and potential departure (in case the recipient was not transplanted till then). During the simulation a pool is updated before each matching run by adding the arrived pairs and removing the departed. This framework of the simulator is similar to the ones described in Santos et al. (2017) and Klimentova et al. (2021), but has extended features to generate realistic pools and perform the matching runs for various optimisation criteria and international collaboration policies that are present in Europe (Biró et al. 2021).

The simulator has four main modules that provide the workflow of the simulations. The modules and the workflow are depicted in Fig. 1.

The Generator module uses an input file to generate the dataset according to the distributions provided. The input file is exemplified on Fig. 2 and includes the following parameters. The blood-types of the donors and recipients are sampled from a given distribution. A recipient may have multiple registered willing donors, and a distribution for the number of the donors can also be set. The PRA of the recipient shows the percentage of the HLA-incompatible donors in a donor population. This value is also sampled from a given distribution. The age distribution can be set separately for recipients, donors and altruistic donors. The parameters arrival time, duration of stay and probability of failure are defined independently for incompatible and compatible pairs and altruistic donors. By changing the two former parameters

```
dist_blood_type = { "O":0.46, "A":0.42, "B":0.09, "AB":0.03 }
dist_num_donors = { 1:0.9, 2:0.1, 3:0.0 }
dist_pra        = { "[0, 9]":0.64, "[10, 79]":0.27, "[80, 99]":0.09 }
dist_age        = { "patient"    : { "(18, 73)":1.0 },
                    "donor"      : { "(18, 73)":1.0 },
                    "altruistic" : { "(18, 73)":1.0 } }
arrival_time    = { "inc": { "avg":12.51 },
                    "com": { "avg":60.0 },
                    "alt": { "avg":80.0 } }
duration_stay   = { "inc": { "avg":365.0, "min":60, "max":2190 },
                    "com": { "avg":365.0, "min":60, "max":2190 },
                    "alt": { "avg":180.0, "min":60, "max":240  } }
prob_fail       = { "inc": { "max":0.5 },
                    "com": { "max":0.3 },
                    "alt": { "max":0.1 } }
```

Fig. 2 Possible settings for the generator module

one can increase or decrease the size of the generated pool. The probability of failure, when set above 0, means that a given transplant identified in the matching run for the virtual compatibility graph can fail to get implemented (classically the main cause of failure in the implementation of an exchange cycle was the positive crossmatch found in the laboratory test, but this issue has been reduced in many countries mainly due to the more accurate high resolution HLA-typing methods).

The Input Reader module is responsible for parsing historical datasets into a data format that the other modules can use. Moreover, in case some necessary attributes are missing from the input dataset, this module will fill them with generated values. The settings of distributions for the generation are provided in the format similar to the input file for the generator module (see Fig. 2). This way the module allows to run simulations not only with generated datasets, but also with datasets coming from real practice. Thus the user may compare the results of different simulations with the results implemented in practice.

The Simulator module conducts matching runs in a dynamic way for the provided dataset for a given time interval. In order to simulate different scenarios in the operation of a national or international KEP, the user can provide the settings in the input files. The simulation settings include the simulation time, frequency of matching runs during this time, constraints for cycle and chain lengths to be searched in optimisation runs, and the collaboration policy used in case of international KEPs.

The Optimisation module is a part of the Simulator and performs the optimisation in the matching runs. The settings for this module include the possibility of using multi-objective optimisation with the variety of built-in objective criteria. The user may specify the approach to be applied for those multiple objectives, lexicographic or weighted optimisation or their combination, and combine any number of optimisation criteria and their prioritisation in the selected approach.

In the case of lexicographic (or hierarchical) optimisation, the first optimisation criterion is considered in the first level. After that the maximum value obtained is set as a new constraint in the second level and the second optimisation criterion is maximised or minimised. This process continues until the unique optimal solution is obtained at any level or the last criterion in the list is reached. When combining

the lexicographic and weighted optimisation, the former always runs first, and the latter is used as a tiebreaker in case the lexicographic optimisation reached the last criterion in the hierarchy with multiple optimal solutions. When multiple criteria are provided also for the weighted optimisation, at first the weights are calculated separately for each criterion using the defined setting. Then the problem is optimised using the summed up weights of all criteria for each transplant or exchange cycle.

The objective coefficients for the selected criteria are calculated with respect to numerous flexible parameters settings that are available for each built-in criteria. This allows the user to apply any optimisation algorithm that is currently used in practice in European KEPs. We will provide specific examples for the optimisation criteria of the UK, Spain, and the Netherlands in Sects. 4, 5, and 6, respectively. We refer to Chapter 3 of the ENCKEP Handbook (Klimentova 2021) for more details on the optimisation process, description of criteria available and their parameters.

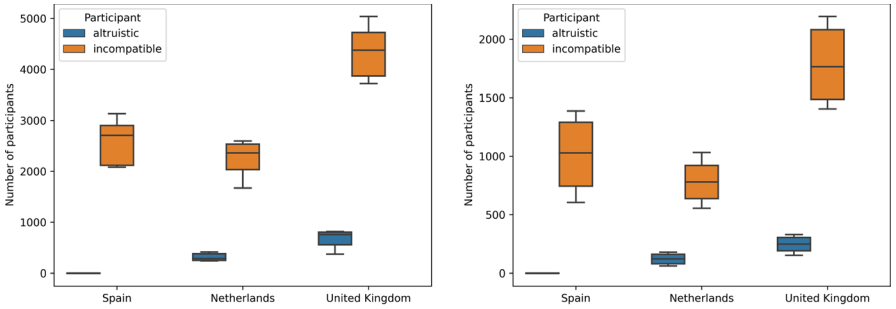
When the simulation is finished, the tool produces detailed output on the simulation, provided in four output files. These files contain information on matching runs, on the selected cycles, on pool of donors and recipients, and the implemented transplants subject to the simulation. The above mentioned four output files can be analysed independently with any tool, but the ENCKEP simulator also includes a built-in solution, the Evaluation module. That module provides the statistics of the simulation and stores them in the database. Note that when we count the number of transplants for an altruistic chain then we only consider the donations to the recipients in the KEP and not the last recipient in the deceased waiting list. So, for instance, if there are three transplants in an altruistic chain, the first one by an altruistic donor to a recipient in the kidney exchange pool and the last one from a registered paired donor to a recipient in the deceased waiting list, then we count two transplants, since two recipients of the KEP pool get transplanted.

3 Simulation setting and test instances

In the following sections, we present the results of the simulations for settings mimicking the national KEPs of the UK, Spain, and the Netherlands, and finally also for the hypothetical collaboration of these three countries.

First, for every country, we present the results for a single simulation. We set the parameters in the configuration files in such a way that the generated instances are realistic with respect to their sizes for each country. The number of recipient-donor pairs arriving to the KEP pools yearly are taken for the year 2015 from a European survey (Biró et al. 2019). The configuration files used to generate the realistic instance for the three countries are presented in the Appendix (see Figs. 15, 16, 17).

Second, in order to validate the robustness of the simulation, using the same configuration files as for single simulation (Figs. 15, 16, 17) we generated 10 large instances for each country. Furthermore in order to be able to make computational experiments with larger test bed we slightly decreased the frequency of arriving pairs in the pools (i.e. reduced the size of the pool of incompatible pairs in the simulation) and generated 100 medium size test instances. The other parameters in



(a) 10 large instances (b) 100 medium size instances

Fig. 3 Total number of incompatible pairs and altruistic donors in the pools for 5-years period for 10 large and 100 medium sizes test instances

configuration files for each country remain the same. The configuration files for generation of medium test instances are presented in Appendix, Figs. 18, 19, 20.

The overall sizes of the generated pools (that is, the number of recipient-donor pairs and altruistic donors arriving to the KEP pools in a 5-years period) are depicted in Fig. 3. Note that for Spain we did not generate altruistic donors, since in the current application the programme deals with the altruistic chains outside of the matching runs. When we simulate the international KEPs we merge these generated national pools.

When analysing the performance of a national or international policy, our focus in on the number of transplants conducted over the 5-years period. However, we also include additional details, such as the level of optimisation reached in a lexicographic policy, or the weight distribution for the weight-factors considered in the last level of the lexicographic optimisation, in the weighted optimisation, regarding the selected matchings. We are able to conduct the latter analysis as a novel approach in the literature, since the main optimisation criteria used in European KEP policies are implemented in the ENCKEP-simulator.

4 United Kingdom

The UK KEP is currently the largest programme in Europe that started operating in 2007. In the first few years a graph based algorithm was used to compute optimal 2-way and 3-way cycles according to the hierarchical optimisation criteria (Biró et al. 2009). Later the algorithm was replaced with an IP technique, described in Manlove and O'Malley (2021).

4.1 Optimisation criteria and simulation on a single instance

In the national KEP of the UK the matching runs are conducted in every 3 months, and the length upper bounds for exchange cycles and chains is set to 3 and 2,

respectively.[1] We slightly change this setting by allowing longer chains of the length at most 3, the same upper bound as for the length of cycles. We also allowed internal recourse in the simulation in order to search for embedded cycles in cycles with either arc or node failure, in the same way as they do in practice. As for the optimisation policy, we used the following set of criteria (see (Klimentova 2021) for further details).

Lexicographic:	Weighted:
1. Maximise the number of effective 2-cycles[2]	• Priority for waiting time in KEP (linear function with score 50)[3]
2. Maximise the number of transplanted KEP recipients in the solution	• Priority for highly sensitised recipients (linear function with score 50)
3. Maximise the number of cycles selected	• Minimise the donor-donor age differences (threshold function with score 3 and threshold 20 years)[4]
4. Maximise the number of back-arcs	

This optimisation policy is almost identical to the one used in real practice. However, since the Generator module of the ENCKEP simulator does not provide HLA-data, we have not used the maximisation of HLA-matching optimisation criterion.[5] Nevertheless, the software is prepared to apply this criterion, provided the HLA-data is available (e.g., in case of real historical datasets).

In Fig. 4 we presented the number of pairs in the pool together with the number of selected and completed transplants (i.e. those that proceeded after failures were taken into account) for each matching run.

One can observe a "warm up" period of simulation (the first six matching runs) when the size of the pool grows. It remains within the range of 220–250 pairs onwards, that corresponds to the size of the pools in the UK KEP. Similar behaviour can be observed also for the selected and performed transplants, though with the growth of numbers in the warm up period being less sharp.

[1] We count as a length of chain the number of transplants for the recipients in the KEP, i.e. the number of arcs in the chain.

[2] An effective 2-cycle is either a 2-cycle or a 3-cycle with embedded 2-cycle(s).

[3] In the weighted optimisation linear functions give scores proportional to a given value: waiting time of the recipient or the PRA-value of the recipient. For example, a recipient who has been present in the UK KEP pool for three matching runs would receive $3 \times 50 = 150$ points, and if she has PRA value 80% then she would receive another $50 \times 0.8 = 40$ points.

[4] The threshold function gives the score to each arc (transplant): if the difference between donors' age in pairs is less than 20, then the score is 3, and 0 otherwise.

[5] The HLA-data of a pool represent the HLA-antigens of the donors, and the HLA-antigens and HLA-antibodies of the recipients. A positive crossmatch in the HLA-compatibility testing occurs when the recipient has a HLA-antibody against the HLA-antigen of the donor. When the HLA-antigens are similar in between a donor and a recipient then the risk of having a positive crossmatch becomes smaller. The similarity of the HLA-antigens are traditionally measured by the HLA-matching value, that counts the similarity of the A,B and DR HLA-antigens resulting in an integer value in between 0 and 6, see (Takemoto et al. 2004) for further details.

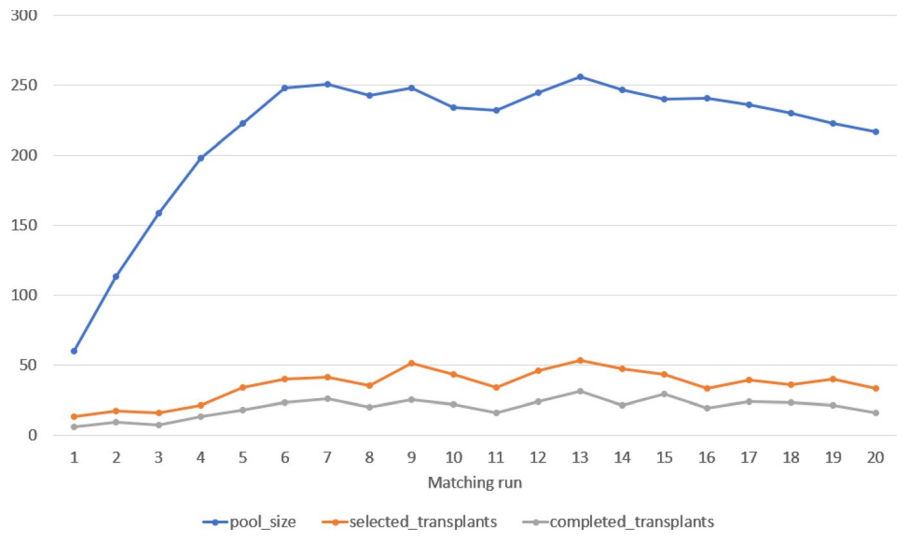


Fig. 4 The UK single instance simulation: dynamics of size of pool of pairs and number of transplants selected/completed per matching run

In order to evaluate the impact of each criterion in weighted optimisation, we depict in Fig. 5 for each matching run the fraction (in percentage) of the weight associated to each of the three criteria in the total weight of the optimal solution.

Figure 5 shows that in every matching run, the most influential weighted criterion was the prioritisation of waiting time in KEP. It is expected, as each matching run is worth 50 points in the waiting time component, that is the maximum score that a recipient can receive for the very high sensitivity (having PRA = 100%).

4.2 Robust simulations

We repeated the above described simulation for 10 large and 100 medium size instances, so we run the simulation on each instance for a five years periods, conducting the matching runs in every three months (thus having 200 and 2000 matching runs in total, respectively).

We describe the aggregated statistics for these simulations. The frequency distribution of final levels in the optimisation runs conducted is depicted in Fig. 6.

Level 0 means that there were no possible cycles found in that optimisation run (which never happened for these instances), and the lth level represents the case when the unique solution was found for the lth lexicographic criteria (from the 5 criteria in total). Level 5 means that the lexicographic optimisation reached the last level, the weighted optimisation.

As we can see in the figure, the weighted optimisation level was reached in the vast majority of the cases (in around 86% of the matching rounds for the large instances and in around 80% of the matching runs for the medium size instances), so

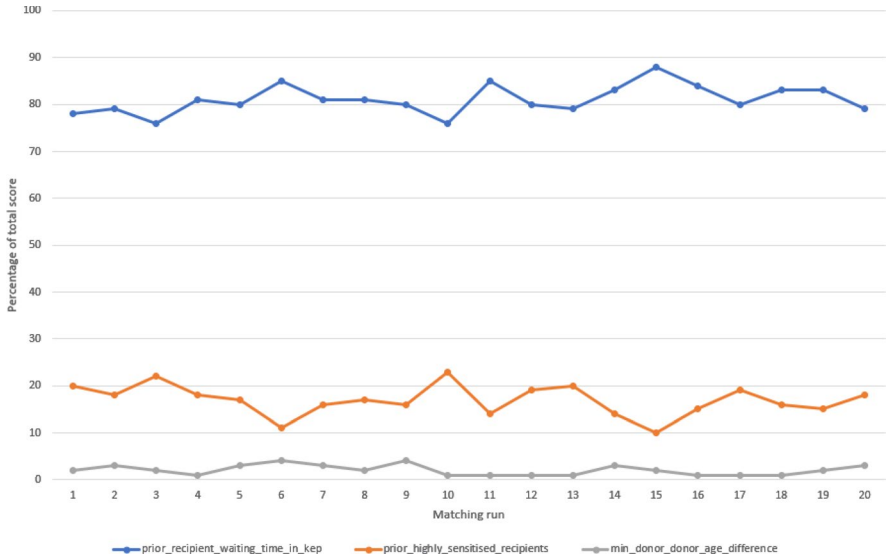
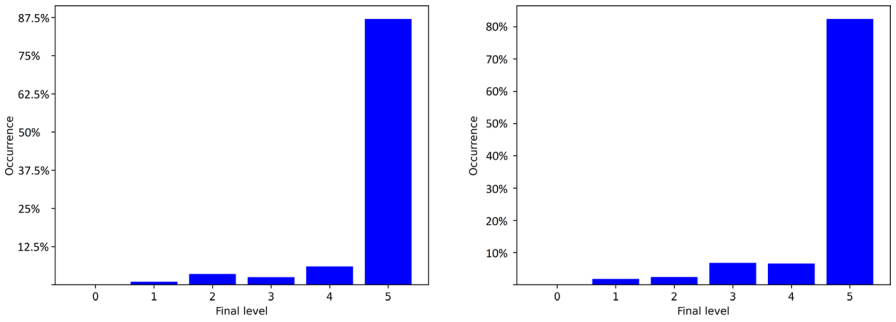


Fig. 5 The fraction (in percentage) of weight of each criterion in the total weight of the optimal solution in each matching run for the UK



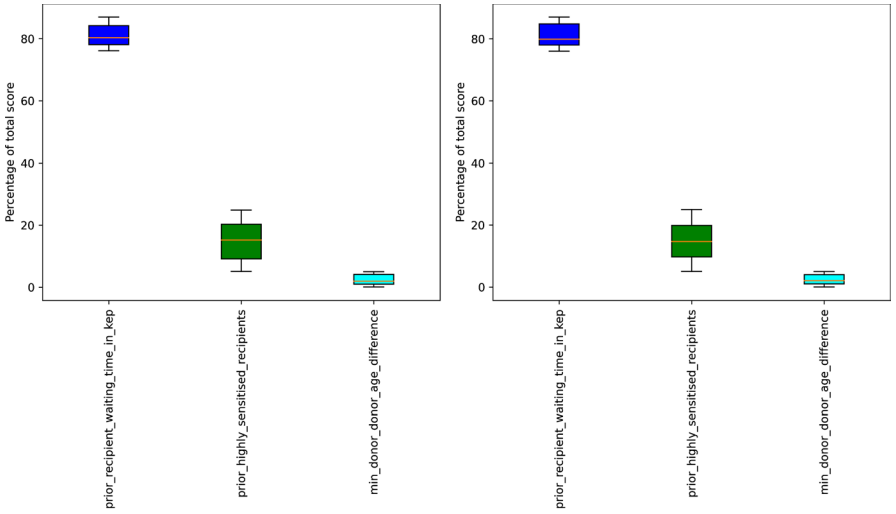
(a) 10 large size instances (b) 100 medium size instances

Fig. 6 Final level of optimisation runs for the UK

the criteria defined for this level very often had an important effect on the solution selected for implementation.

Since there are multiple criteria used for the weighted optimisation, checking which one had the largest impact can be useful to precisely adjust the scoring functions. In a similar way as for the single simulation in the previous section we depict in Fig. 7 the percentage of weight of each criterion in the total weight of the optimal solution in all matching runs for all instances of a given group (large and medium size instances).

We measured these scores only if the weighted optimisation level (5th) was reached, as in those cases the weighted criteria have an effect on the selection of the



(a) 10 large size instances (b) 100 medium size instances

Fig. 7 Weighted criteria impact for the UK

solution. As described in the previous section, the criterion for prioritising based on waiting time had the highest weight in the weighted optimisation. At the same time significantly lower weights were given for prioritising highly sensitised recipients. Minimising the donor-donor age differences seems to be the final discriminator in case of the first two weighted criteria resulted in approximately the same scores for some alternative solutions, as it is intended in the application.

5 Spain

The Spanish national kidney exchange programme was developed in 2009 by Organización Nacional de Trasplantes (ONT). Currently it is the second largest KEP in Europe after the UK programme.

5.1 Optimisation criteria and simulation on one instance

The KEP operating in Spain sets the limit for maximum length of exchange cycles to 3. Recall that we do not consider altruistic donors in the simulations for Spain. The simulation time and the frequency of matching runs was set in the same way as for the UK (5 years period, matching runs in every 3 months, respectively). Again, we allowed the internal recourse in the simulations, as indeed there is a re-optimisation round conducted by the programme after the possible cancellations due to failure by positive crossmatches identified in the laboratory, or other reasons. The optimisation policy consists of the following set of criteria (see (Klimentova 2021) for details).

Lexicographic:	Weighted:
1. Maximise the number of transplanted KEP recipients in the solution	• Minimise the age-differences between donors and recipients (threshold function with score 15 and threshold 10 years)
2. Maximise the number of cycles selected	• Priority for recipients with low matching probability (linear function with score 30)
3. Maximise the number of back-arcs	• Priority for waiting time in KEP (threshold function with score 30 and threshold 1 year)
4. Priority for highly sensitised recipients (reciprocal function with scor	• Priority for same blood-group transplants (30 points) e 5)

This set of criteria is similar to the one used in practice. The difference is that three criteria for the weighted optimisation were not considered. Priority for paediatric recipients was not implemented as the generated pool contains exclusively adult patients. The two criteria that give priorities based on the time spent on dialysis and for the donor being in the same region as the recipient were left out from the analysis, because no data is available for these criteria in the generated datasets.

As the dynamics of size of pool of pairs and the number of transplants for each matching run are similar to those presented for the UK in the previous section (see Fig. 4). We will skip similar graphics for Spain and the Netherlands.

Figure 8 depicts the fraction (in percentage) of the weight for each weighted optimisation criteria in the total weight of the optimal solution in each matching run.

As we can see, priority for same blood-group accounted for the largest total score in all matching runs, followed by the total score on matching probability, whilst waiting time and donor-recipient age differences contributed almost equally to the total score.

5.2 Robust simulations

Next we conducted simulations for 10 large and 100 medium size instances. The statistics on the number of levels reached in the lexicographical optimisation are depicted in Fig. 9.

The patterns are similar to the UK case, the final weighted optimisation level was reached in the vast majority of the matching runs.

Figure 10 depicts the percentage of weight of each criterion in the total weight of the optimal solution in all matching runs for all instances of a given group for the cases when the weighted optimisation was reached.

The priority for the same blood-group transplants had the highest weight in weighted optimisation runs, while the values for the other three criteria were significantly lower. In contrast with the UK, the weights given based on the waiting time of recipients were the lowest. This can be explained by the fact that this score is only awarded for the waiting time if the recipient stayed in the pool for 12 months, which is an unlikely scenario here since the average duration of stay is set to 12 months, and many recipients also get transplanted before they would leave the pool (i.e., before their projected departure time).

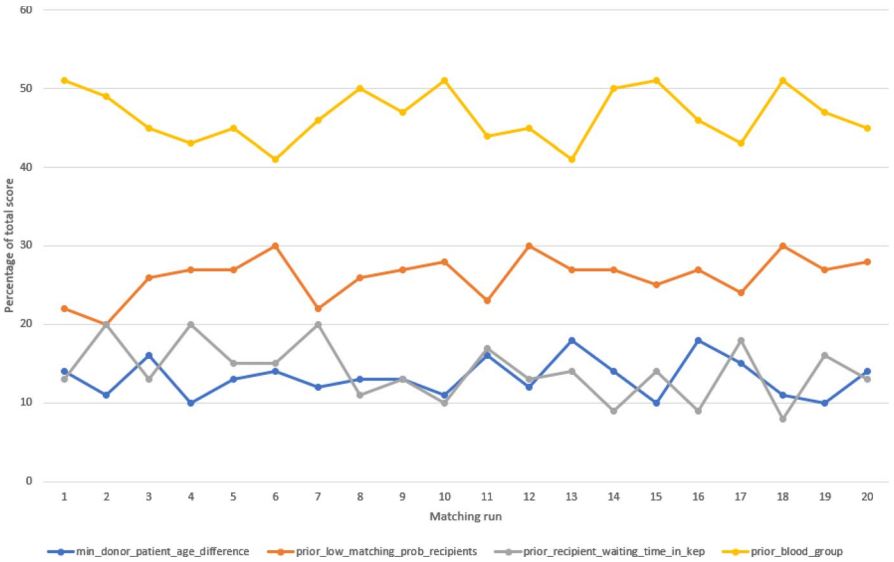
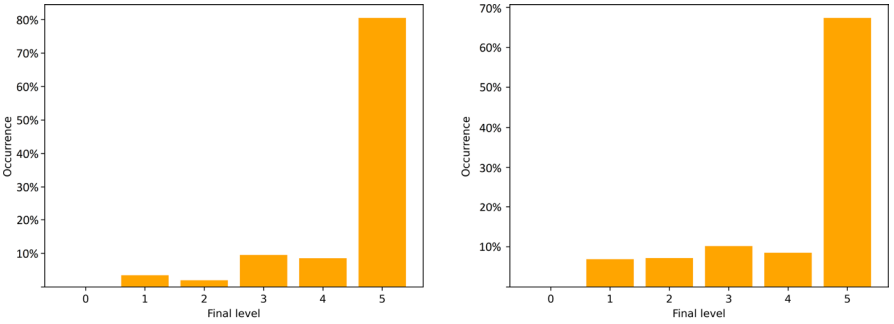


Fig. 8 The fraction (in percentage) of weight of each criterion in the total weight of the optimal solution in each matching run for Spain

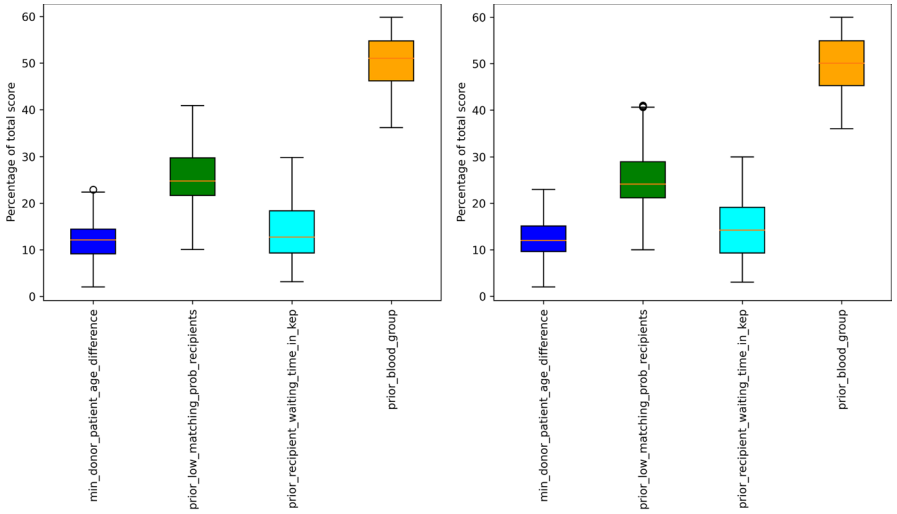


(a) 10 large instances (b) 100 medium size instances

Fig. 9 Final level of optimisation runs for Spain

6 The Netherlands

The Dutch KEP was the first in Europe, established in 2004. It is coordinated by the Dutch Transplant Foundation with 8 transplant centres and a single central laboratory in Leiden for histocompatibility testing. Whenever an optimal solution is selected, a quick (2 day) laboratory testing can be conducted with the blood-samples stored in the central lab, and if some positive crossmatch or other issue would occur then a re-optimisation can take place. This is the main reason why 4-way exchanges are also considered in the solution, as opposed to the UK and Spain, where only one limited



(a) 10 large instances (b) 100 medium size instances

Fig. 10 Weighted criteria impact for the datasets for Spain

re-optimisation round is allowed due to the multiple HLA-labs used in these countries. For further details, see (de Klerk et al. 2008) and de et al. (2011).

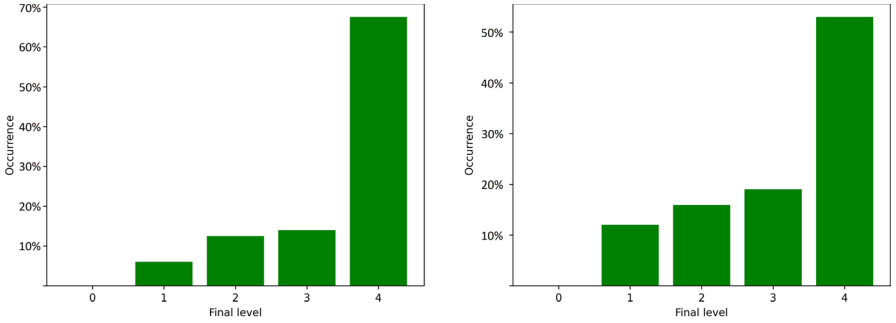
6.1 Optimisation criteria

In our simulations for the Netherlands, we used 4 as upper length limit for both exchange cycles and chains, and internal recourse was enabled here as well. The optimisation policy we used in the simulation was the following (see (Klimentova 2021) for further details).

Lexicographic:

1. Maximise the number of transplanted KEP recipients in the solution
2. Priority for same blood-group transplants
3. Priority for recipients with low matching probability (using reciprocal function with coefficient 5)
4. Minimise the lengths of cycles selected

The policy is similar to the one used in practice, the difference is that we left out the 5th and 6th criteria (in hierarchy): the maximisation of the number of transplant centres in long cycles and priority for time on dialysis, respectively, since these criteria are not yet supported in the ENCKEP simulator.



(a) 10 large instances (b) 100 medium size instances

Fig. 11 Final level of optimisation runs for the Netherlands

6.2 Robust simulations

As mentioned above, in the Netherlands the upper length limit of cycles and chains is set to 4. Since only lexicographic optimisation is applied for the used criteria, the results do not include figures on weighted optimisation that we have seen for the other two applications. However, we can see in Fig. 11 that the optimisation stopped on the last level in our model in most of the optimisation runs conducted. This means, that the last two criteria that were omitted could have made an impact on the solution selected.

7 International collaboration

As mentioned in introduction, we are aware of four international KEPs running worldwide:

- Czech-Austrian-Israeli collaboration,
- KEPSAT involving Italy, Portugal, Spain,
- STEP, run by Scandiatransplant,
- NZKX by Australia and New Zealand.

One of the main aspects that can differ in between international KEPs is collaboration polices. In KEPSAT each of the three countries conducts a national matching run first, and only after that they seek for international cycles in the remaining pools. The other three international KEPs merge their pools and conduct one single matching run.

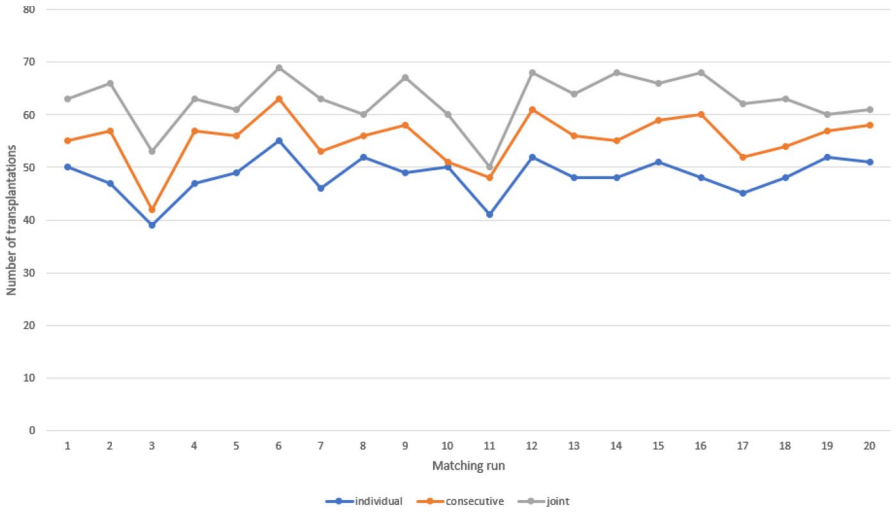


Fig. 12 Number of transplants identified in each matching run with different collaboration policies

Table 1 Total number of transplants by pools and collaboration policies

	Individual	Consecutive	Joint
UK	409	484	537
Spain	281	331	409
The Netherlands	278	293	309
Total	968	1108	1255

7.1 Optimisation criteria and simulation on a single instance

The variants of collaboration policies mentioned above can be simulated using the software:

- Individual policy: Each participating pool performs matching runs separately. By use of this policy one can access the benefits of participants from participation in the international collaboration.
- Consecutive policy: In each matching run, first the optimisation run is performed for every pool separately. Then the pairs that are still remaining in the pools are merged into one joint pool, and another optimisation run is performed for that pool.
- Joint policy: All the participating pools are merged into one pool, and this pool is used in the matching runs.

In order to compare the policies, we applied them for the same generated pool of pairs. For the merged pools, we used the optimisation policy of the UK. The results are depicted on Fig. 12.

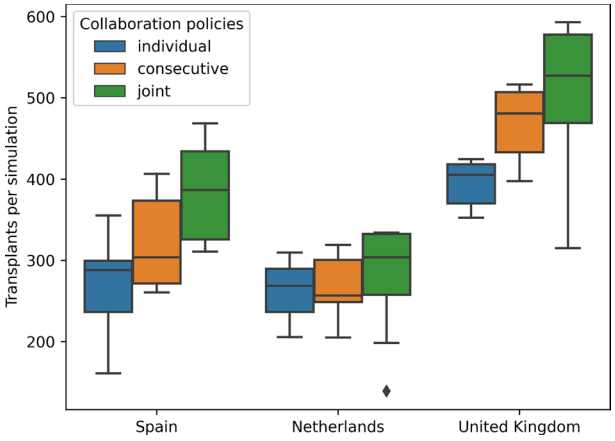


Fig. 13 Total number of transplants performed by each country for each collaboration policy, 10 large instances

We can see that for this realistic size instance, the joint policy produced the most transplants in each matching run, followed by the consecutive policy.

As reported in Table 1 the total number of transplants for the individual, consecutive and joint collaboration policies were 968, 1108, and 1255, respectively. According to the simulation result, the total number of transplants can be increased by initiating a collaboration between the countries, where the joint policy seems to be the best approach. At the same time, the gains observed for the Netherlands are relatively small, since in their national KEP (individual policy) they allow 4-long cycles and chains, whilst in the joint policy only 3-long cycles and chains are allowed. This reveals the potential need to apply different constraints or more advanced policies in order to incentivise a country to enter an international collaboration, e.g. the approach proposed in Klimentova et al. (2021).

7.2 Robust simulations

We repeated the previously described international KEP simulations for each of three collaboration policies for 10 large and 100 medium size instances. The total number of transplants performed by each country under each collaboration policy for large and medium size test instances is depicted in Figs. 13 and 14, respectively.

The presented robust results based on large scale performance analysis led to essentially the same conclusion. The benefit of the collaboration was the largest for Spain, and the smallest for the Netherlands, explained by the pool sizes and the different optimisation policy used for the merged pool.

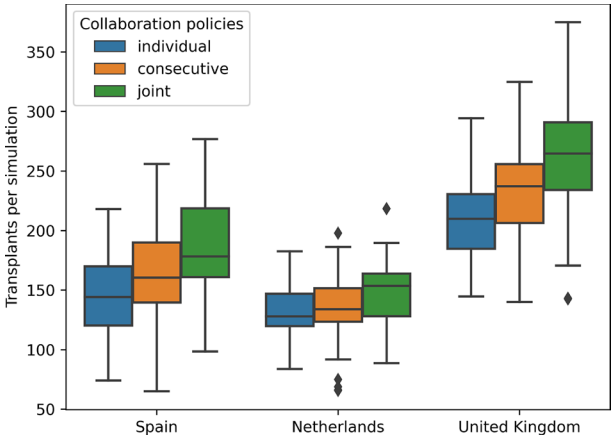


Fig. 14 Total number of transplants performed by each country for each collaboration policy, 100 medium size instances

8 Conclusion

In this paper we demonstrated the ENCKEP-simulator by performing analysis of kidney exchange programmes for three European countries (UK, Spain, the Netherlands) using realistically generated datasets and optimisation criteria that approximate the current practices. We described the findings for a single instance and also for 10 large and 100 medium size instances for each country by taking the averages. Furthermore, we studied the expected benefits of international KEPs under different collaboration policies again for a single and robust simulations as well.

In a follow-up paper Matyasi and Biró studied various re-optimisation policies for international KEPs with the ENCKEP-simulator (Bir´o et al. 2009). As future work, we would like to conduct further analyses on real historical datasets, and investigate additional questions, such as measuring the expected quality of the transplants using quality indices that can predict the expected graft survival times, and allowing ABO-incompatible transplants in the exchanges.

Appendix

Additional figures added after revision:

```
{
    "name": "general",
    "dist_blood_type": {"O": 0.46, "A": 0.42, "B": 0.09, "AB": 0.03},
    "dist_num_donors": {"1": 0.9, "2": 0.1, "3": 0.0},
    "dist_pra": {"[0, 9]": 0.64, "[10, 79]": 0.27, "[80, 99]": 0.09},
    "dist_age": {"patient": {"(18, 73)": 1.0}, "altruistic": {"(18, 73)": 1.0}},
    "arrival_time": {"inc": {"avg": 2.31}, "com": {"avg": -1}, "alt": {"avg": 68.5}},
    "duration_stay": {"inc": {"avg": 365.0, "min": 60, "max": 2190}, "com": {"avg": 365.0, "min": 60, "max": 2190}, "alt": {"avg": 180.0, "min": 60, "max": 240}},
    "prob_fail": {"inc": {"max": 0.5}, "com": {"max": 0.3}, "alt": {"max": 0.1}},
    "prob_pass_hla": {"*-*": 1.0}
}
```


Fig. 15 The configuration file for the UK, single simulation and large size test instances


```
{
    "name": "general",
    "dist_blood_type": {"O": 0.46, "A": 0.42, "B": 0.09, "AB": 0.03},
    "dist_num_donors": {"1": 0.9, "2": 0.1, "3": 0.0},
    "dist_pra": {"[0, 9]": 0.64, "[10, 79]": 0.27, "[80, 99]": 0.09},
    "dist_age": {"patient": {"(18, 73)": 1.0}, "altruistic": {"(18, 73)": 1.0}},
    "arrival_time": {"inc": {"avg": 8.51}, "com": {"avg": -1}, "alt": {"avg": -1}},
    "duration_stay": {"inc": {"avg": 365.0, "min": 60, "max": 2190}, "com": {"avg": 365.0, "min": 60, "max": 2190}, "alt": {"avg": 180.0, "min": 60, "max": 240}},
    "prob_fail": {"inc": {"max": 0.5}, "com": {"max": 0.3}, "alt": {"max": 0.1}},
    "prob_pass_hla": {"*-*": 1.0}
}
```


Fig. 16 The configuration file for Spain, single simulation and large size test instances


```
{
    "name": "general",
    "dist_blood_type": {"O": 0.46, "A": 0.42, "B": 0.09, "AB": 0.03},
    "dist_num_donors": {"1": 0.9, "2": 0.1, "3": 0.0},
    "dist_pra": {"[0, 9]": 0.64, "[10, 79]": 0.27, "[80, 99]": 0.09},
    "dist_age": {"patient": {"(18, 73)": 1.0}, "altruistic": {"(18, 73)": 1.0}},
    "arrival_time": {"inc": {"avg": 10.51}, "com": {"avg": 60.0}, "alt": {"avg": 80.0}},
    "duration_stay": {"inc": {"avg": 365.0, "min": 60, "max": 2190}, "com": {"avg": 365.0, "min": 60, "max": 2190}, "alt": {"avg": 180.0, "min": 60, "max": 240}},
    "prob_fail": {"inc": {"max": 0.5}, "com": {"max": 0.3}, "alt": {"max": 0.1}},
    "prob_pass_hla": {"*-*": 1.0}
}
```


Fig. 17 The configuration file for the Netherlands, single simulation and large size test instances


```
{
    "name": "general",
    "dist_blood_type": {"O": 0.46, "A": 0.42, "B": 0.09, "AB": 0.03},
    "dist_num_donors": {"1": 0.9, "2": 0.1, "3": 0.0},
    "dist_pra": {"[0, 9]": 0.64, "[10, 79]": 0.27, "[80, 99]": 0.09},
    "dist_age": {"patient": {"(18, 73)": 1.0}, "altruistic": {"(18, 73)": 1.0}},
    "arrival_time": {"inc": {"avg": 9.31}, "com": {"avg": -1}, "alt": {"avg": 77.5}},
    "duration_stay": {"inc": {"avg": 365.0, "min": 60, "max": 2190}, "com": {"avg": 365.0, "min": 60, "max": 2190}, "alt": {"avg": 180.0, "min": 60, "max": 240}},
    "prob_fail": {"inc": {"max": 0.5}, "com": {"max": 0.3}, "alt": {"max": 0.1}},
    "prob_pass_hla": {"*-*": 1.0}
}
```


Fig. 18 The configuration file for the UK, medium size test instances


```
{
    "name": "general",
    "dist_blood_type": {"O": 0.46, "A": 0.42, "B": 0.09, "AB": 0.03},
    "dist_num_donors": {"1": 0.9, "2": 0.1, "3": 0.0},
    "dist_pra": {"[0, 9]": 0.64, "[10, 79]": 0.27, "[80, 99]": 0.09},
    "dist_age": {"patient": {"(18, 73)": 1.0}, "altruistic": {"(18, 73)": 1.0}},
    "arrival_time": {"inc": {"avg": 15.51}, "com": {"avg": -1}, "alt": {"avg": -1}},
    "duration_stay": {"inc": {"avg": 365.0, "min": 60, "max": 2190}, "com": {"avg": 365.0, "min": 60, "max": 2190}, "alt": {"avg": 180.0, "min": 60, "max": 240}},
    "prob_fail": {"inc": {"max": 0.5}, "com": {"max": 0.3}, "alt": {"max": 0.1}},
    "prob_pass_hla": {"*-*": 1.0}
}
```


Fig. 19 The configuration file for Spain, medium size test instances


```
{
    "name": "general",
    "dist_blood_type": {"O": 0.46, "A": 0.42, "B": 0.09, "AB": 0.03},
    "dist_num_donors": {"1": 0.9, "2": 0.1, "3": 0.0},
    "dist_pra": {"[0, 9]": 0.64, "[10, 79]": 0.27, "[80, 99]": 0.09},
    "dist_age": {"patient": {"(18, 73)": 1.0}, "altruistic": {"(18, 73)": 1.0}},
    "arrival_time": {"inc": {"avg": 17.51}, "com": {"avg": -1}, "alt": {"avg": 89.0}},
    "duration_stay": {"inc": {"avg": 365.0, "min": 60, "max": 2190}, "com": {"avg": 365.0, "min": 60, "max": 2190}, "alt": {"avg": 180.0, "min": 60, "max": 240}},
    "prob_fail": {"inc": {"max": 0.5}, "com": {"max": 0.3}, "alt": {"max": 0.1}},
    "prob_pass_hla": {"*-*": 1.0}
}
```


Fig. 20 The configuration file for the Netherlands, medium size test instances

Acknowledgements We acknowledge the financial support by the Hungarian Academy of Sciences, Momentum Grant No. LP2021-2, and by the Hungarian Scientific Research Fund, OTKA, Grant No. K143858.

Funding Open access funding provided by HUN-REN Centre for Economic and Regional Studies.

Declarations

Conflict of interest The authors declared that they have no Conflict of interest.

References

Abraham D, Blum A, Sandholm T (2007) Clearing algorithms for barter exchange markets: enabling nationwide kidney exchanges. Proc EC 295–304:2007

Agarwal N, Ashlagi I, Azevedo E, Featherstone CR, Karaduman Ö (2019) Market failure in kidney exchange. Am Econ Rev 109:4026–4070

Ashlagi I, Roth AE (2021) Kidney exchange: an operations perspective. Manag Sci 67:5455–5478

Bhaskaran MC, Heidt S, Muthukumar T (2022) Principles of virtual crossmatch testing for kidney transplantation. Kidney Int Rep 7:1179–1188

Biró P, Haase B, van de Klundert J et al (2019) Building kidney exchange programmes in Europe: an overview of exchange practice and activities. Transplantation 103:1514–1522

Biró P, van de Klundert J, Manlove DF et al (2021) Modelling and optimisation in European kidney exchange programmes. Eur J Oper Res 291:447–456

Biró P, Manlove DF, Rizzi R (2009) Maximum weight cycle packing in directed graphs, with application to kidney exchange programs. Dis Math Algorithms Appl 1:499–517

Böhmig GA, Fronek J, Slavcev A, Fischer GF, Berlakovich G, Viklicky O (2017) Czech-Austrian kidney paired donation: first European cross-border living donor kidney exchange. Transpl Int 30:638–639

Delorme M, Garcıa S, Gondzio J, Kalcsics J, Manlove DF, Petterson W (2023) New algorithms for hierarchical optimisation in kidney exchange programmes. Oper Res. https://doi.org/10.1287/opre.2022.2374

Druzsin K, Biró P, Fleiner R, Klimentova X (2021) Simulations for measuring efficiency of international kidney exchange programmes. In: 16th international symposium on operational research in Slovenia

Druzsin K, Biró P, Fleiner R, Klimentova X (2022) Large scale performance analysis of international kidney exchange programmes by the ENCKEP simulator. In: VOCAL optimization conference: advaced algorithms

de Klerk M, Witvliet MD, Haase-Kromwijk BJ et al (2008) A flexible national living donor kidney exchange program taking advantage of a central histocompatibility laboratory: the Dutch model. Clin Transpl 69:73

de Klerk M, Kal-van Gestel JA, Haase-Kromwijk BJ, et al (2011) Living donor kidney exchange program. Eight years of outcomes of the dutch living donor kidney exchange program. Clin Transpl, p 287

Klimentova X, Viana A, Pedroso JP, Santos N (2021) Fairness models for multi-agent kidney exchange programmes. Omega 102:102333

Klimentova X, et al (2021) International kidney exchange programmes in Europe: practice, solution models, simulation and evaluation tools. Handbook of working group 3 and 4 of the ENCKEP cost action

Manlove DF, O'Malley G (2021) Paired and altruistic kidney donation in the UK: algorithms and experimentation. J Exp Algorithmics 19:1–21

Matyasi L, Biro P (2023) Testing re-optimisation strategies in international kidney exchange programmes by the ENCKEP simulator. Central Eur J Oper Res. https://doi.org/10.1007/s10100-023-00880-2

Mincu RS, Biró P, Gyetvai M, Popa A, Verma U (2021) IP solutions for international kidney exchange programmes. Central Eur J Oper Res 29(2):403–423

Santos N, Tubertini P, Viana A, Pedroso JP (2017) Kidney exchange simulation and optimization. J Oper Res Soc 68:1521–1532

Takemoto S, Port FK, Claas FH, Duquesnoy RJ (2004) HLA matching for kidney transplantation. Hum Immunol 65:1489–1505

Valentín MO, Garcia M, Costa AN, Bolotinha C, Guirado L, Vistoli F, Breda A, Fiaschetti P, Dominguez-Gil B (2019) International cooperation for kidney exchange success. Transplantation 103(6):180–181

Authors and Affiliations

Kristóf Druzsin[1,2] · Péter Biró[1,3] · Xenia Klimentova[4] · Rita Fleiner[2]

✉ Kristóf Druzsin
druzsin.kristof@krtk.hu

Péter Biró
biro.peter@krtk.hu

Xenia Klimentova
xenia.klimentova@inesctec.pt

Rita Fleiner
fleiner.rita@nik.uni-obuda.hu

[1] Institute of Economics, KRTK, Tóth Kálmán utca 4., Budapest 1097, Hungary

[2] John von Neumann Faculty of Informatics, Óbuda University, Bécsi út 96/B, Budapest 1034, Hungary

[3] Department of Operations and Decisions, Corvinus University of Budapest, Fővám tér 13., Budapest 1093, Hungary

[4] INESC TEC, Campus da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal