A BAYESIAN OPTIMIZATION APPROACH TO ESTIMATING EXPECTED MATCH TIME AND ORGAN QUALITY IN KIDNEY EXCHANGE

Anonymous authors
Paper under double-blind review

ABSTRACT

Kidney exchanges allow patients with end-stage renal disease to find a lifesaving living donor by way of an organized market. However, not all patients are equally easy to match, nor are all donor organs of equal quality—some patients are matched within weeks, while others may wait for years with no match offers at all. Knowledge of expected waiting time and organ quality affects medical and insurance decisions. This work presents a principled method to estimate the expected quality of the kidney that a specific patient who enters an exchange will receive, as well as how long it will take to find that match. Estimation is performed via a novel Bayesian-optimization-based approach that learns a model of a computationally complex underlying Monte Carlo simulator. With a limited number of expensive simulation trajectories, the model produces practically-applicable results. Such fast and accurate sampling could provide medical professionals near-instantaneous access to valuable insight regarding a patient's expected outcome in a kidney exchange system.

1 INTRODUCTION

Renal disease affects millions of people worldwide, with a societal burden comparable to diabetes (Neuen et al., 2013). A patient with end-stage renal failure requires one of two treatments to stay alive: frequent and costly filtration & replacement of their blood (dialysis), or the reception of an organ transplant from a donor with one or more healthy kidneys. The latter option is often preferable due to increased quality of life and other health outcomes (Santos et al., 2015). Donor kidneys are obtained from one of three sources: the deceased donor waiting list, where cadaveric kidneys are harvested from deceased donors with still-healthy kidneys; ad-hoc arrangements between a compatible living donor and a patient; and, recently, *kidney exchanges* – an organized market where patients swap willing donors with other patients (Roth et al., 2004; 2005a;b). Kidney exchanges, while still quite new, result in increased numbers and quality of transplants (Sönmez et al., 2017); furthermore, their design is a success story for fielded AI research (Abraham et al., 2007; Ashlagi & Roth, 2014; Anderson et al., 2015; Dickerson & Sandholm, 2015; Hajaj et al., 2015; Toulis & Parkes, 2015; Manlove & O'Malley, 2015). The act of getting a kidney transplant is time-sensitive, and affects healthcare and lifestyle decisions; furthermore, the expected quality of the kidney—if any—received by a patient affects the decision to accept or reject a particular match offer, and may be used to (de)prioritize patients in a matching mechanism (Bertsimas et al., 2013). Thus, decision-support systems that incorporate donor and patient features and quantify or predict the value of a current or future offered kidney are valuable to practitioners. The Kidney Donor Profile Index (KDPI) (Rao et al., 2009) and the Living Kidney Donor Profile Index (LKDPI) (Massie et al., 2016) are well-known and used to assess deceased- and living-donor kidneys, respectively. However, no method/system exists to find the expected quality of a donated kidney in a kidney exchange.

This paper presents a Bayesian-optimization-based system that takes as input features of a patient and their paired donor, and returns an estimate of (i) the expected quality of a match, and (ii) expected waiting time for a matched kidney offer. The use of modern tools from machine learning and combinatorial optimization is required due to the NP-hard and APX-hard nature of even the most basic problems in kidney exchange (Abraham et al., 2007; Biró & Cechlárová, 2007; Biró et al., 2009; Luo et al., 2016; Jia et al., 2017). Our method uses a realistic but expensive black-box Monte

Carlo simulator to produce estimates of match quality and time-to-match for a specific patient and donor; it samples new points in the space intelligently, balancing overall computational time with the accuracy of prediction for a new patient and donor. This prediction can be done in real or near-real time, a requirement for such a decision-support system. We give a proof-of-concept implementation on a reduced but realistic set of features in the kidney-exchange setting, and show that the method learns the necessary functions well.

2 PRELIMINARIES

Deceased- & Living-Donor Kidney Allocation. Our motivation is, in part, due to the widespread usage of the Kidney Donor Profile Index (KDPI) to quantify the value of deceased-donor kidneys, and the increasing use of the newer Living Kidney Donor Profile Index (LKDPI) to quantify the value of living-donor kidneys (Rao et al., 2009; Massie et al., 2016). Roughly speaking, both the KDPI and the LKDPI are metrics used to compute the expected lifetime (quality) of a kidney transplant. Both are based on multivariate Cox Regression models adapted from the statistics literature (Cox, 1992). The LKDPI was constructed such that LKDPI scores can be directly compared with KDPI scores, and evaluates transplants based on both donor features (e.g., estimated glomerular filtration rate, body mass index), as well as features indicating donor-patient compatibility (e.g., blood type (ABO) and human leukocyte antigen mismatches). We expand this metric of quality to fielded kidney exchange. Unlike a standard ad-hoc living-donor donation, in a donation through a kidney exchange, the features of the end donor are unknown, and are generated through a stochastic matching process. We aim to compute the expected LKDPI of the kidney received through kidney-paired donation, and the expected matching time that it would take to receive this kidney, in order to allow for comparison between the living donor, deceased donor, and kidney-paired donation options.

The Formal Kidney-Exchange Model. The most-used model represents a kidney exchange as a directed graph $G = (V, E)$, called a *compatibility graph*. Here, each patient and their paired donor who enter the pool are represented as a *single* vertex. Then, a directed edge is drawn from vertex v_i to vertex v_j if the patient at vertex v_j wants the donor kidney of vertex v_i. Weights $w_e \in \mathbb{R}$ represent the utility of an individual kidney transplant represented by an edge e, and are also used to (de)prioritize specific classes of patient (Dickerson et al., 2014; UNOS, 2015).

A *matching* M is a set of disjoint cycles and chains in a compatibility graph G; $M \in \mathcal{M}$, the set of all legal matchings.[1] No donor can give more than one of her kidneys, necessitating the disjointness of cycles and chains—although recent work explores multi-donor donation (Ergin et al., 2017; Farina et al., 2017). Given the set of all legal matchings $\mathcal{M}$, the *clearing problem* finds the matching M^* that maximizes utility function $u : \mathcal{M} \to \mathbb{R}$ (e.g., for maximum weighted matching, $u(M) = \sum_{c \in M} \sum_{e \in c} w_e$). Formally: $M^* \in \arg\max_{M \in \mathcal{M}} u(M)$. Ongoing research in the AI/Economics literature uses utility functions to enforce incentive properties via mechanism design (Ashlagi & Roth, 2014; Li et al., 2014; Hajaj et al., 2015; Blum et al., 2017; Mattei et al., 2017). Finding a maximum weight (capped-length) cycle and chain packing is NP-hard (Abraham et al., 2007; Biró et al., 2009), and is also hard to approximate (Biró & Cechlárová, 2007; Luo et al., 2016; Jia et al., 2017). In practice, integer program (IP) formulations are used to clear large exchanges (Abraham et al., 2007; Dickerson et al., 2013; Glorie et al., 2014; Anderson et al., 2015; Dickerson et al., 2016).

The kidney-exchange system is dynamic. In each iteration, new patients enter the pool, current patients may leave due to competition from other methods for receiving a kidney or death, and edges may appear or disappear based on the health characteristics of participants (e.g., pregnancy or sickness, leading to a change in compatibility with potential donors). Furthermore, the matching process is highly stochastic. In fielded kidney exchanges, matches are made without detailed knowledge of compatibility between a donor and patient. More-thorough physical *crossmatch tests* are done after an algorithmic match, but before the actual transplantation event, to ensure that a matched donor can donate to a paired patient. Even one failure of an edge in a cycle invalidates the *entire* cycle; similarly, given the incremental execution of chains, all potential transplants located after the first edge failure in a chain are invalidated.

[1] In fielded kidney exchanges, cycles are limited in size to, typically, 3; all surgeries in a cycle must be executed simultaneously, so longer cycles are nearly impossible to plan. Chains, however, can be much longer (or effectively endless) in practice.

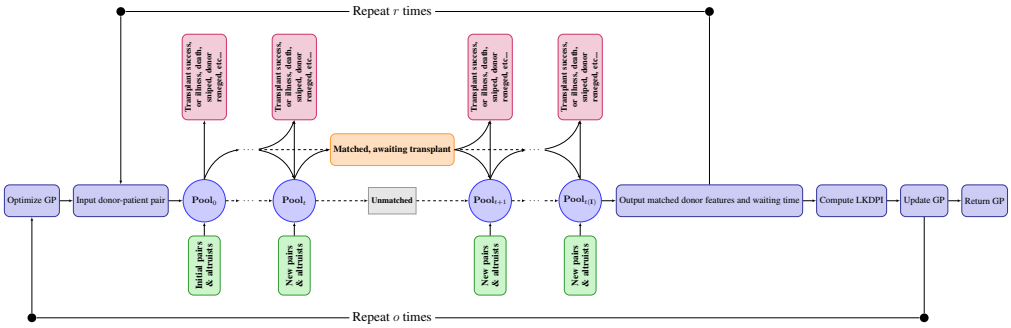

Figure 1: Estimating LKDPI and waiting time using Bayesian optimization

3 Estimating Match Time And Quality

We return to our initial motivation—creating a decision-support system that quickly and accurately estimates to a patient and paired donor the expected waiting time and offered organ quality. Our approach consists of two components: a realistic (but computationally expensive) simulator capable of sampling patient trajectories in the exchange, and a Bayesian-optimization-based active-sampling framework that learns the underlying model of the simulator for real-time prediction.

Simulating the Kidney Exchange. Due to limitations in the availability of data, we consider a simplified kidney-exchange where edge weights in the compatibility graph are computed based on a patient's blood type, their potential donor's blood type, their paired but incompatible donor's blood type, and the respective CPRA (described next) of all involved parties (patient, paired donor, potential donor).[2] Although the feature set is simplified, we are able to draw from a reliable approximation of their joint distribution using data from the OPTN Standard Transplant Analysis and Research dataset[3], which contains the biological information of donors and patients—including pre- and post-transplantation event information— in conjunction with a modified graph structure generator from the medical literature due to Saidman et al. (2006). Here, the Calculated Panel Reactive Antibodies (CPRA) is a continuous-valued score in $[0, 1]$, roughly representing the fraction of donors, drawn from a general population, that would *not* be a match for a particular patient (e.g., a score of 1 signals extreme difficulty in matching). We let matches take place on a weekly basis, and cap the total match-time at 250 weeks. The pool is instantiated with an initial graph size of 250 patient-donor pairs and 10 altruists. Each time period, we draw from an arrival distribution of new vertices, and conduct crossmatch tests between each pair of vertices to determine the edges of the compatibility graph. Vertices expire with constant probability each iteration, and matches from the previous iteration are accepted with probability geometric in the number of edges. At the end of the iteration, we invoke an IP-based codebase (Dickerson & Sandholm, 2015) that solves the NP-hard optimization problem from §2 to generate new matches.[4] This model is standard in the realistic-kidney-exchange simulation literature (Saidman et al., 2006; Anderson et al., 2015; Dickerson & Sandholm, 2015).

A Bayesian-Optimization-Based Approach. One may naively estimate match time and transplant quality for a given patient by calling the simulator many times and aggregating the results. However, due to the aforementioned computational complexity of the simulator, this naive approach may not be executed quickly enough for real-time estimation. To remedy this issue, we propose an active-sampling approach based on the Bayesian Optimization (BO) framework that predicts the output of this naive approach. BO utilizes Gaussian processes (GPs) to maximize an unknown function—in our case, the expected output of a realistic, but noisy and expensive to run, simulator of a real-world process. In BO, an *acquisition function* is used to select the next point at which to sample that would, typically, lead to an accurate estimate of the maximum of that function. A kernel (covariance) function is used to interpolate between known values of the function, and determine the confidence

[2]For a discussion of all features that are included in a large, real-world kidney exchange due to the United Network for Organ Sharing (UNOS), we direct the reader to a white-paper from that group (UNOS, 2015).

[3]Request at `https://optn.transplant.hrsa.gov/data/request-data/`

[4]After the double-blind review period concludes, we will make our repository containing all code to replicate experiments in this paper publicly available.

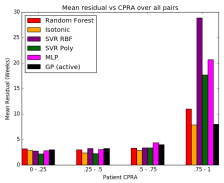

Figure 2: MAEs over all blood type pairs

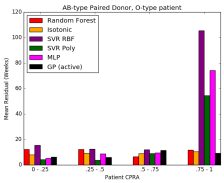

Figure 3: MAEs for the AB-O pair

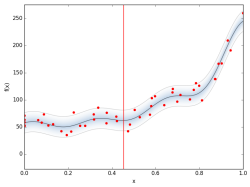

Figure 4: Learned GP for the AB-O pair

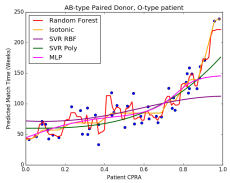

Figure 5: Comparison models for AB-O pair

at each point. While BO offers a method to maximize a function where getting output is time consuming, we learn the expected transplant quality/waiting time as a function of the features of the patient-donor pair using this method by taking as output the GP of the BO. We use this framework to choose informative input patient-donor pairs that we then simulate to obtain sample statistics on waiting time and transplant quality. We model the waiting-time and quality functions by a Gaussian automatic relevance determination kernel $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2}\sum_{i=1}^{D}\frac{(x_i - x_i')^2}{2\ell_i^2}\right)$. The lengthscale hyperparameters ℓ_i determine the relevance of each dimension in determining the covariance.

We then select the point $\mathbf{x}$ that the GP is "least certain" about by letting the acquisition A be given by the posterior variance. Without hyperparameter optimization, the posterior variance of the GP depends only on the previous inputs that were chosen, and does not take into account the output information. We incorporate the output information in a principled way by approximating the *expected* posterior variance using the posterior distribution p of the hyperparameters ℓ_i given the observed data. That is, we let $A(\mathbf{x}) = \mathbb{E}_{(\ell_1,\ldots,\ell_D)\sim p}[\sigma_{\mathbf{x}}^2]$. We make such an approximation efficiently by using *Hamiltonian Monte Carlo*, a physics-inspired Markov Chain Monte Carlo technique (Duane et al., 1987). We implemented our method by modifying GPyOpt (GPyOpt, 2016), an open-source Bayesian-optimization platform.

Experiments. Due to the simplified feature set that we use in our simulator, the LKDPI turns out to be independent of the input features. Thus, we validate our approach on the waiting time. For all 16 blood-type-pair combinations for donor-patient pairs who enter the exchange, we performed Bayesian optimization over patient CPRA for $o = 50$ iterations. For each blood-type-pair combination, waiting time is estimated based on an aggregation of $r = 48$ trajectories in a realistic simulator. Figure 1 gives a graphical description of our setup. After constructing the GPs through BO, we test them by comparing the match time returned by the GP and the match time returned by the realistic simulator after s trajectories. To test the 16 generated GPs, the domain of CPRA $[0, 1]$ is partitioned uniformly into 4 zones $\{[0, .25), [.25, .5), [.5, .75), [.75, 1)\}$. In each zone, 5 random trials are done, with $s = 128$ trajectories each. We test the system with $s > r$ in order to measure the ability of the model to make good estimations of the expected value even with more noisy input. As no baseline exists to validate our method, we show that our approach can produce clinically promising estimates relative to common regression models in the Scikit-learn framework (Pedregosa et al., 2011). [5] Figure 2 shows that our approach makes clinically promising estimates over all blood types and CPRAs. Other methods, while performing well on easy-to-match patients with low CPRA, make comparatively poor estimates for the harder-to-match patients. This difference is more apparent for the hardest-to-match (AB-donor, O-patient) pair. Our approach clearly outperforms other methods for this pair (Figure 3), and fits the collected data more robustly (Figures 4 and 5).

While our proof-of-concept demonstrates the promise of this system in a reduced environment, before making a policy recommendation or deploying a support tool in practice, we note the following: (i) The number of trajectories r should be (much) greater, for greatly reduced stochasticity, and thus far smaller mean residuals from the GP to the realistic simulator. (ii) The number of features considered should be much higher, and informed by experts in the field. Yet, given these proof-of-concept experimental results, we feel confident that a decision-support system can be deployed for use by practitioners, in order to give patients and donors this information on demand.

[5] We compare against random forests, isotonic regression, kernelized support vector regression, and multilayer perceptrons. All models were tuned with 200 iterations of randomized hyperparameter optimization. None of these models are capable of learning the underlying noise model of the function, which is of clinical importance, as it allows stakeholders to know the margin of error on the expected-waiting-time estimate.

REFERENCES

David Abraham, Avrim Blum, and Tuomas Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pp. 295–304, 2007.

Ross Anderson, Itai Ashlagi, David Gamarnik, and Alvin E Roth. Finding long chains in kidney exchange using the traveling salesman problem. *Proc. National Academy of Sciences*, 2015.

Itai Ashlagi and Alvin E Roth. Free riding and participation in large scale, multi-hospital kidney exchange. *Theoretical Economics*, 9(3):817–863, 2014.

Dimitris Bertsimas, Vivek F Farias, and Nikolaos Trichakis. Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Operations Research*, 61(1):73–87, 2013.

Péter Biró and Katarína Cechlárová. Inapproximability of the kidney exchange problem. *Information Processing Letters*, 101(5), March 2007.

Péter Biró, David F Manlove, and Romeo Rizzi. Maximum weight cycle packing in directed graphs, with application to kidney exchange programs. *Discrete Mathematics, Algorithms and Applications*, 1(04):499–517, 2009.

Avrim Blum, Ioannis Caragiannis, Nika Haghtalab, Ariel D. Procaccia, Eviatar B. Procaccia, and Rohit Vaish. Opting into optimal matchings. In *Proceedings of the Symposium on Discrete Algorithms (SODA)*, 2017.

David R Cox. Regression models and life-tables. In *Breakthroughs in Statistics*, pp. 527–541. Springer, 1992.

John P. Dickerson and Tuomas Sandholm. FutureMatch: Combining human value judgments and machine learning to match in dynamic environments. In *AAAI Conference on Artificial Intelligence (AAAI)*, pp. 622–628, 2015.

John P. Dickerson, Ariel D. Procaccia, and Tuomas Sandholm. Failure-aware kidney exchange. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pp. 323–340, 2013.

John P. Dickerson, Ariel D. Procaccia, and Tuomas Sandholm. Price of fairness in kidney exchange. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 1013–1020, 2014.

John P. Dickerson, David Manlove, Benjamin Plaut, Tuomas Sandholm, and James Trimble. Position-indexed formulations for kidney exchange. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2016.

Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216–222, 1987.

Haluk Ergin, Tayfun Sönmez, and M. Utku Ünver. Multi-donor organ exchange, 2017. Working paper.

Gabriele Farina, John P. Dickerson, and Tuomas Sandholm. Operation frames and clubs in kidney exchange. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

Kristiaan M. Glorie, J. Joris van de Klundert, and Albert P. M. Wagelmans. Kidney exchange with long chains: An efficient pricing algorithm for clearing barter exchanges with branch-and-price. *Manufacturing & Service Operations Management (MSOM)*, 16(4), 2014.

GPyOpt. GPyOpt: A Bayesian optimization framework in Python, 2016.

Chen Hajaj, John P. Dickerson, Avinatan Hassidim, Tuomas Sandholm, and David Sarne. Strategy-proof and efficient kidney exchange using a credit mechanism. In *AAAI Conference on Artificial Intelligence (AAAI)*, pp. 921–928, 2015.

Zhipeng Jia, Pingzhong Tang, Ruosong Wang, and Hanrui Zhang. Efficient near-optimal algorithms for barter exchange. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 362–370, 2017.

Jian Li, Yicheng Liu, Lingxiao Huang, and Pingzhong Tang. Egalitarian pairwise kidney exchange: Fast algorithms via linear programming and parametric flow. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 445–452, 2014.

Suiqian Luo, Pingzhong Tang, Chenggang Wu, and Jianyang Zeng. Approximation of barter exchanges with cycle length constraints. *CoRR*, abs/1605.08863, 2016. URL `http://arxiv.org/abs/1605.08863`.

David Manlove and Gregg O'Malley. Paired and altruistic kidney donation in the UK: Algorithms and experimentation. *ACM Journal of Experimental Algorithmics*, 19(1), 2015.

Allan B Massie, Joseph Leanza, LM Fahmy, EKH Chow, Niraj M Desai, X Luo, EA King, MG Bowring, and DL Segev. A risk index for living donor kidney transplantation. *American Journal of Transplantation*, 16(7):2077–2084, 2016.

Nicholas Mattei, Abdallah Saffidine, and Toby Walsh. Mechanisms for online organ matching. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

Brendon L Neuen, Georgina E Taylor, Alessandro R Demaio, and Vlado Perkovic. Global kidney disease. *The Lancet*, 382(9900):1243, 2013.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Panduranga S Rao, Douglas E Schaubel, Mary K Guidinger, Kenneth A Andreoni, Robert A Wolfe, Robert M Merion, Friedrich K Port, and Randall S Sung. A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. *Transplantation*, 88(2):231–236, 2009.

Alvin Roth, Tayfun Sönmez, and Utku Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119(2):457–488, 2004.

Alvin Roth, Tayfun Sönmez, and Utku Ünver. A kidney exchange clearinghouse in New England. *American Economic Review*, 95(2):376–380, 2005a.

Alvin Roth, Tayfun Sönmez, and Utku Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125(2):151–188, 2005b.

Susan L. Saidman, Alvin Roth, Tayfun Sönmez, Utku Ünver, and Frank Delmonico. Increasing the opportunity of live kidney donation by matching for two and three way exchanges. *Transplantation*, 81(5):773–782, 2006.

Alfonso H Santos, Michael J Casey, Xuerong Wen, Ivan Zendejas, Shehzad Rehman, Karl L Womer, and Kenneth A Andreoni. Survival with dialysis versus kidney transplantation in adult hemolytic uremic syndrome patients: A fifteen-year study of the waiting list. *Transplantation*, 99(12):2608–2616, 2015.

Tayfun Sönmez, Utku Unver, and M Bumin Yenmez. Incentivized kidney exchange, 2017. Tech. report, Boston College Dept. of Economics.

Panos Toulis and David C. Parkes. Design and analysis of multi-hospital kidney exchange mechanisms using random graphs. *Games and Economic Behavior*, 91(0):360–382, 2015.

UNOS. Revising kidney paired donation pilot program priority points, 2015. OPTN/UNOS Public Comment Proposal.