Ethics-by-design: efficient, fair and inclusive resource allocation using machine learning

Theodore P. Papalexopoulos 1, Dimitris Bertsimas 1,*, I. Glenn Cohen 2, Rebecca R. Goff 3, Darren E. Stewart 3 and Nikolaos Trichakis 1

1 Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
2 Harvard Law School, Harvard University, Cambridge, MA 02138, USA
3 Research Department, United Network for Organ Sharing, Richmond, VA 23219, USA

*Corresponding author. E-mail: dbertsim@mit.edu

ABSTRACT

The distribution of crucial medical goods and services in conditions of scarcity is among the most important, albeit contested, areas of public policy development. Policymakers must strike a balance between multiple efficiency and fairness objectives, while reconciling disparate value judgments from a diverse set of stakeholders. We present a general framework for combining ethical theory, data modeling, and stakeholder input in this process and illustrate through a case study on designing organ transplant allocation policies. We develop a novel analytical tool, based on machine learning and optimization, designed to facilitate efficient and wide-ranging exploration of policy outcomes across multiple objectives. Such a tool enables all stakeholders, regardless of their technical expertise, to more effectively engage in the policymaking process by developing evidence-based value judgments based on relevant tradeoffs.

KEYWORDS: Analytics, ethics by design, machine learning, organ allocation, organ transplantation, resource allocation

I. INTRODUCTION

As has been made clear by current controversies regarding criteria for allocating COVID-19 vaccines in U.S. states and across countries, the distribution of crucial medical goods and services in conditions of scarcity is among the most important,
albeit contested, areas for public policy development.\textsuperscript{1–3} COVID-19 vaccine allocation policy questions—for example, how much to prioritize age and measures of vulnerability and whether to consider ‘indirect benefits’—although in some sense specific to the pandemic, are also representative of a more general challenge in devising allocation policies: how to combine ethical theory, data modeling, and stakeholder input into a single process for generating a sound allocation system.

The importance of deriving and implementing good models for doing this can be seen in what is widely considered as a major failure in this area: the Oregon Medicaid prioritization process of the late 1980s and early 1990s. Oregon’s goal was to expand Medicaid to all persons below the poverty level while keeping costs manageable by restricting which services were reimbursable. What services to cover was determined using a combination of methods including (i) actuarial estimates of the cost of providing Medicaid coverage, (ii) a state-wide telephone survey formulated to measure ‘Quality of Well Being’ and calculate the effectiveness of ∼1600 treatments for particular conditions, and (iii) input from physician specialist panels regarding benefit duration and average age of onset of relevant conditions.\textsuperscript{4} The initial process’s focus on cost–benefit analysis generated a list of covered and non-covered procedures that was discomfiting to some, for example by covering tooth-capping but not emergent appendicitis. Ensuing controversy prompted a move to ‘correct’ the process by, among other things, creating larger overarching categories and ‘moving by hand’ into covered categories procedures or treatments that seemed to the commissioner to be common-sense priorities.\textsuperscript{5}

The Oregon experiment mixed evidence, ethics, and stakeholder input in a witch’s brew that was perceived as ad hoc and political. The process nevertheless had (at least) some of the right ingredients. A purely top-down ethics approach, absent testing of its ramifications through data modeling, is problematic, but so is prioritizing what can be measured rather than measuring what should be prioritized; and both ethics- and data-driven approaches will fail without stakeholder pressure testing and buy in. Can all of this be combined into a better model?

We argue yes, and describe a general framework that we view as a major step forward for ethically informed policymaking. The key idea is to incorporate ethics into a data-driven policy design process from the outset. At the core of our approach is a novel analytical tool, based on machine learning and optimization, that enables stakeholders to assess tradeoffs between different policy objectives. To achieve that aim, users specify their desired system-level outcomes, encompassing diverse ethical and utility considerations, and the tool identifies a conforming policy in near-real time. By exploring the impact of changing user inputs, stakeholders, from ethicists to community leaders, can develop evidence-based value judgments on relevant tradeoffs.

\textsuperscript{1} National Academies of Sciences Engineering and Medicine, \textit{Framework for Equitable Allocation of COVID-19 Vaccine} (Helene Gayle, et al. eds., The National Academies Press. 2020).
\textsuperscript{2} Muriel Jean-Jacques & Howard Bauchner, \textit{Vaccine Distribution—Equity Left Behind?}, 325 JAMA 829 (2021).
\textsuperscript{3} Tim Jin, \textit{Op-Ed: Why Is California’s Age-based COVID-19 Vaccine Policy Overlooking Disabled People Like Me?}, Los Angeles Times, https://www.latimes.com/opinion/story/2021-01-29/covid-vaccine-disabled-people-priority (Accessed April 10, 2022).
\textsuperscript{4} Arti K. Rai, \textit{Rationing Through Choice: A New Approach to Cost-Effectiveness Analysis in Health Care}, 72 Indiana Law J 1015–1097 (1997).
\textsuperscript{5} Philip A. Perry & Timothy Hotze, \textit{Oregon’s Experiment with Prioritizing Public Health Care Services}, 13 AMA J ETHICS, 241–247 (2011).
regardless of their technical expertise, and more effectively engage in the policymaking process.

To present our framework, we use as a case study the Organ Procurement & Transplant Network’s (OPTN) policymaking process for migrating from the current classification-based policy to a continuous distribution (CD) model for organ allocation.\textsuperscript{6,7} We begin by providing background information on policy development for organ allocation in the United States.

II. REDESIGNING ORGAN ALLOCATION IN THE U.S.

II.A Background

Since 1986, the OPTN has been operated under a federal contract by the United Network for Organ Sharing (UNOS). A core obligation of UNOS is to facilitate development of policies that determine how organs from deceased donors in the United States are allocated to medically suitable candidates on the national waiting list. The unfortunate reality is that demand for transplants far outstrips available supply, and transplant candidates generally wait a significant amount of time before receiving an organ. But not everyone can afford to wait. Some end-stage organ failure patients are so critically ill that without a transplant they will die in a matter of days.

The guiding principles for developing organ allocation policy are set by the OPTN Final Rule.\textsuperscript{8} Among other requirements, the Final Rule states that organ allocation policies must be based on ‘sound medical judgement,’ seek to achieve the ‘best use’ of donated organs, and promote ‘efficient management’ of organ placement. Allocation policies should aim to equitably distribute organs to those most in need over as broad a geographic area as is feasible.

In translating the abstract ethical principles established by the Final Rule into concrete policy, the transplant community is confronted with many ethical dilemmas. For example, to which of two equally sick candidates needing an organ should it be offered first? What if one candidate would be transplanted at a hospital closer to the donor’s location, reducing potentially detrimental effects of organ ischemic time and increasing organ placement efficiency, but the more distant candidate has been waiting longer? Adjudicating such dilemmas in a systematic, objective, and repeatable manner every time a donated organ becomes available (24 × 7, 365 days a year) is the role of the organ allocation policies operationalized by UNOS.

Crucially, UNOS does not unilaterally develop policy, but rather serves as a convener of the transplant community to help develop new and refine existing policies. This community consists of a host of stakeholders including transplant surgeons, physicians, Organ Procurement Organization (OPO) professionals, transplant candidates and recipients, living donors and donor families, as well as the general public. The OPTN policy development process is designed to foster careful, wide-ranging discus-

\textsuperscript{6} Jon J. Snyder, et al., Organ Distribution without Geographic Boundaries: A Possible Framework for Organ Allocation, 18 AM J TRANSPLANT 2635 (2018).

\textsuperscript{7} Organ Procurement and Transplantation Network, Key Initiatives: Continuous Distribution, https://optn.transplant.hrsa.gov/governance/key-initiatives/continuous-distribution/ (Accessed April 10, 2022).

\textsuperscript{8} Allocation of organs. 42 C.F.R. § 121.8, https://www.ecfr.gov/cgi-bin/text-idx?SID=dea21abb91c032be0bd4a42b85fd8c7&mc=true&node=se42.1.121_18&rgn=div8 (Accessed April 10, 2022).
sion and deliberation based on input from these various constituencies.\textsuperscript{9,10} Transplant professionals and patients provide input by serving on OPTN committees that meet regularly to discuss clinical and practical details involved in developing policy and monitoring policy impacts post implementation.

In these discussions, the OPTN employs an evidence-based approach to policy development, guided by subject-matter expertise from transplant professionals and patients and backstopped by public comment and Board approval steps incorporated into the process. Evidence is typically derived from both retrospective analyses of the rich OPTN transplant database (e.g., identifying inequities in access to organs from historical data) and allocation simulation modeling performed by the Scientific Registry of Transplant Recipients (SRTR) contractor to predict outcomes of proposed policy changes.\textsuperscript{11} In recent years, more sophisticated approaches incorporating mathematical optimization and simulation modeling have at times played an integral role in the policy development process.\textsuperscript{12,13}

II.B. Continuous Distribution

The tension inherent in translating the guidance of the Final Rule into concrete policy is nowhere more evident than in the longstanding policy debates surrounding the role of geography in allocating organs.\textsuperscript{14–20} To what degree should proximity to the donor hospital be considered vis-a-vis a candidate’s medical urgency, especially given the historic precedent of prioritizing ‘local’ candidates (those registered at a transplant hospital within the administrative boundary assigned to OPOs)? With the central aim of increasing transparency and removing hard geographic boundaries that sometimes preclude organs from going to candidates most in need, the OPTN recently embarked

\textsuperscript{9} Organ Procurement and Transplantation Network, \textit{Policy Development}, https://optn.transplant.hrsa.gov/media/3115/optn-policy-development-process-explanatory-document.pdf (Accessed April 10, 2022).

\textsuperscript{10} David L. Weimer, \textit{Medical Governance: Values, Expertise, and Interests in Organ Transplantation} (Georgetown University Press, 2010).

\textsuperscript{11} Scientific Registry of Transplant Recipients, \textit{Simulated Allocation Models}, https://www.srtr.org/requesting-srtr-data/simulated-allocation-models/ (Accessed April 10, 2022).

\textsuperscript{12} Sommer E. Gentry, et al., \textit{Addressing Geographic Disparities in Liver Transplantation Through Redistricting}, 13 \textit{Am J Transplant} 2052 (2013).

\textsuperscript{13} Sanjay Mehrotra, et al., \textit{A Concentric Neighborhood Solution to Disparity in Liver Access that Contains Current UNOS Districts}, 102 \textit{Transplantation} 255 (2018).

\textsuperscript{14} Alexandra K. Glazier, \textit{The Lung Lawsuit: A Case Study in Organ Allocation Policy and Administrative Law}, 14 \textit{J Health Biomed Law} (2018).

\textsuperscript{15} Ranjit Deshpande, et al., \textit{Liver Allocation and Distribution: Time for a Change}, 22 \textit{Curr Opin Organ Transplant} 162 (2017).

\textsuperscript{16} Allan B. Massie & John Paul Roberts, \textit{Geographic Disparity in Liver Allocation: Time to Act or Have Others Act for US}, 102 \textit{Transplantation} 189 (2018).

\textsuperscript{17} Logan Patrick Moore & David L. Weimer, \textit{The Geography of Life and Death: Evidence and Values in the Evolution of US Liver Transplant Rules}, 13 \textit{World Med Health Policy} S26 (2021).

\textsuperscript{18} Zhizhou Yang, et al., \textit{Shipping Lungs Greater Distances Increases Costs Without Cutting Waitlist Mortality}, 110 \textit{Ann Thorac Surg} 1691 (2020).

\textsuperscript{19} Rebecca R. Lehman & Kevin M. Chan, \textit{Elimination of the Donor Service Area (DSA) from Lung Allocation: No Turning Back} (Wiley Online Library 2019).

\textsuperscript{20} Daniela Lamas & Lisa Rosenbaum, \textit{Very Complicated Math—Reconfiguring Organ Allocation}, 371 \textit{NEngl J Med} 2447 (2014).
on a large-scale initiative to migrate all allocation policies, starting with lung, to the borderless CD framework.\textsuperscript{21}

Although under current policy candidates are prioritized by way of an ordered list of groups of ‘similar’ candidates, under the CD framework candidates are prioritized by a mathematical formula. In its simplest form, all candidates on the lung waitlist, for example, could be ranked by a weighted sum of their medical priority (quantified by their Lung Allocation Score) and placement efficiency (using distance to the donor hospital as a proxy). The relative weight of these two components would govern how widely organs are offered to those in need, thereby codifying the tradeoff between placement efficiency and need-based equity. Converging on the ‘right’ balance between efficiency and fairness is exactly the sort of ethical dilemma the OPTN seeks to reconcile through the deliberative policy development process.

Converting to the new CD framework also provides an opportunity to revisit value judgements embedded in existing policies.\textsuperscript{22} For example, what influence should a candidate attribute like waiting time, which codifies the ‘first come, first serve’ ethic, play vis-a-vis other policy objectives like ensuring equitable access to patients with biological disadvantages (e.g. a harder-to-match blood type)? Or what role should reducing waitlist mortality play compared with maximizing survival time among those who receive a transplant? Additional attributes can be incorporated into the composite allocation score to adjust candidate priority accordingly, and relative weights of each component chosen to strike a defensible balance between the OPTN’s many efficiency, utility, and fairness objectives. Value judgements (weights), as well as attributes used to build the score, will almost certainly differ depending on the organ (kidney, liver, heart, lung, etc.), as each organ-specific policy is expected to have its own formula.

Simulation allocation modeling provides crucial input to the OPTN’s evaluation process by predicting system-wide outcomes of proposed policies. The SRTR’s simulation models use historical waitlist and transplant data to simulate counterfactual allocation under a proposed prioritization scheme and predict aggregated outcomes (e.g., overall mortality rates, transplant rates for different candidate subpopulations, transport metrics, etc.). The ability to compare policies holistically across multiple objective dimensions facilitates ‘outcome-driven’ policy discussions and helps to identify pain points to be accounted for (e.g., increased inefficiencies, such as more organ shipments expected to require a flight, or unintended inequities in access to organs under a proposed policy).

Consequently, simulation modeling has typically been conducted via trial-and-error, with simulated outcomes from initial policy ideas used to iteratively refine those ideas until acceptable predicted outcomes are achieved. With CD, the process might work as follows. Given attributes to be included in the priority formula, an initial set of policies (i.e., corresponding weights) would be proposed, simulated, and their predicted utility/efficiency/fairness outcomes reviewed by stakeholders. The results and committee discussions would inform the selection of a subsequent round of policy

\textsuperscript{21} Organ Procurement and Transplantation Network, Executive Summary of the OPTN/UNOS Board of Directors Meeting, December 3-4, 2018, https://optn.transplant.hrsa.gov/media/2787/boards_executivesummary_201812.pdf (Accessed April 10, 2022).

\textsuperscript{22} Darren E. Stewart, et al., A Revealed Preference Analysis to Develop Composite Scores Approximating Lung Allocation Policy in the US, 21 BMC MED INFORM DECIS MAK 8 (2021).
options to be simulated, for example, by increasing or decreasing certain attribute weights to address a predicted inequity or inefficiency, and the process would be repeated.

Unfortunately, the computationally expensive and time-consuming nature of iteratively running simulations, reviewing and discussing results, and determining the next set of options to simulate means that this crucial evidence-generating phase of policy development can consume many months or even years. Even when a desired improvement is identified during an iteration, it is not a priori clear what, if any, proposed remedy will achieve better outcomes until new simulations are run. Time considerations may constrain the range of options tried, or many iterations may be required until a suitable option is found. The decade-long development of the kidney allocation system, in which well over 30 different policy options were simulated, is a prime example.\textsuperscript{23–25}

III. SHIFTING TO ETHICS-BY-DESIGN FOR OUTCOME-DRIVEN POLICY DESIGN

In this work, inspired by ethics-by-design principles, we introduce a general analytical framework for outcome-driven policy design when development is constrained by a burdensome evaluation process such as simulation. Our approach leverages the predictive power of machine learning to replace the bottleneck evaluation step with near-instantaneous model-based evaluation. The ability to more efficiently predict each outcome for a given policy in turn allows us to invert the problem: instead of starting with policy options and iteratively evaluating predicted outcomes, the approach starts with the desired outcomes and returns a specific policy engineered to achieve them (Figure 1). The result is an analytical tool that is ideally suited to characterizing trade-offs between multiple objectives, allowing stakeholders to iterate on what the desired objective balance should be rather what design decisions achieve it.

IIIA. Optimization Methodology

We illustrate our approach as it might be applied to the design of a CD lung allocation policy. The CD case is a particularly potent example, as the space of possible policies—that is, the space of relative attribute weights for the priority formula—is essentially infinite, and searching over it for the policy that strikes the ‘right’ balance over multiple objectives becomes prohibitively expensive when using conventional simulation. Were it possible to evaluate policies instantaneously, the full efficient frontier of outcomes could be characterized a priori, and the policy instantiations that correspond to different points on it known precisely.

To this end, our methodology seeks first to characterize, using machine learning, the full range of achievable outcomes, and subsequently, through optimization, the efficient

\textsuperscript{23} Mark D. Stegall, et al., \textit{Why Do We Have the Kidney Allocation System We Have Today? A History of the 2014 Kidney Allocation System}, 78 HUM IMMUNOL 4 (2017).
\textsuperscript{24} OPTN/UNOS Kidney Transplantation Committee, \textit{Kidney Allocation Concepts: Request for Information} (2008), https://asts.org/docs/default-source/optn-unos/proposed-kidney-allocation-concepts-rfi-se ptember-24-2008.pdf (Accessed April 10, 2022).
\textsuperscript{25} Ajay K. Israni, et al., \textit{New National Allocation Policy for Deceased Donor Kidneys in the United States and Possible Effect on Patient Outcomes}, 25 J AM SOC NEPHROL 1842 (2014).
Figure 1. One step in the iterative design process under a traditional trial-and-error approach (top) vs. our framework (bottom). In the former, a burdensome evaluation process (e.g. computationally expensive simulation) forms a bottleneck that limits the number of candidate policies stakeholders can use to assess tradeoffs. Moreover, it is not a priori clear whether a proposed remedy aimed at improving outcomes will work until the modified policy is evaluated itself. In our framework, machine learning models remove the evaluation bottleneck and allow optimization to efficiently search for a policy that best achieves a set of prespecified outcomes. Source: Author-created illustration.
frontier. The role of policymakers would then be to select which point on the frontier best achieves the stated policy objectives; that is, decide what relative attribute weights strike their desired balance in the various efficiency, fairness, and utility outcomes of interest. To be clear, the end-product of this design process is still a fully transparent CD policy that ranks candidates according to a static allocation formula and can be easily explained to patients, physicians, and other community members.

At a high level, our approach relies on specialized machine learning models to accurately and near-instantaneously predict the outcomes of any given policy instantiation. Efficient predictions in turn allow optimization algorithms to sift through the space of possible policies and find the one that best achieves a set of user specified outcomes.

The first step involves using the simulation allocation model to predict the full set of outcomes for a fixed number of randomly generated policies, resulting in a dataset that maps policy instantiations (attribute weights) to each outcome of interest (mortality rate, disparities in transplant rates, transport costs, etc.). We modify the 2015 Thoracic Simulation Allocation Model (TSAM), publicly available through the SRTR, to prioritize patients according to a CD formula and simulate each policy over the 2009–2011 period. These initial simulations, though computationally expensive, need only be run once and can be parallelized across multiple machines to reduce computational overhead.

The simulated runs are used to train nonlinear regression models that serve as computationally efficient ‘approximators’ of the simulator. During this offline training process, the models use the initial runs to identify and extrapolate correlations between attribute weights and each simulated outcome; then, given a new set of attribute weights, they can predict outcomes such as mortality rate or transport cost without invoking the computationally expensive simulator. We ensure that the models accurately capture the complexity of the simulator by using piecewise-linear functions that account for nonlinearity in the relationship between attribute weights and outcomes. In validation, we found that these regressions achieved out-of-sample R² ranging from 0.90 to 0.99 depending on the outcome, able to predict the simulator’s output with surprising accuracy.

With the ability to predict any simulated outcome accurately and efficiently, we can then use optimization to find new policies that achieve any set of prespecified outcomes in near real-time. In an optimization problem, one seeks to find the values of decision variables (here, composite score attribute weights) that minimize or maximize an overarching objective function (a simulated outcome, such as mortality or posttransplant survival) subject to additional constraints (efficiency or fairness requirements, such as an upper bound on average organ transport distance or transplant rate disparities for patients with different blood types). Given such a specification, we use an open-source mixed-integer optimization solver to sift through the vast space of possible policies to find one that meets the specified criteria. If no such policy exists (e.g., because the constraints were too stringent), the optimization problem can be modified to instead find a policy that deviates minimally from the stated requirements. Once such a policy has been found, the original simulator is reinvoked to verify the regression models’ predictions and evaluate the optimization-derived policy in full.
III.B. Implications for Policy Design

Because any simulator output modeled in the regression can be designated in either the objective or constraints of an optimization problem, policymakers can quickly iterate through different optimization scenarios—that is, different sets of objectives and constraints involving different policy outcomes—and explore what policies are predicted to achieve them. One might envision a ‘slider’ based interface, as depicted in Figure 2, whereby policymakers iteratively explore scenarios by changing the outcome to optimize and adding or adjusting constraint bounds.

One might begin, for example, by looking for a policy that minimizes waitlist mortality (the overarching objective) without increasing average organ transport distance vis-à-vis current policy (a constraint). The optimization computes a conforming allocation policy within seconds, which is simulated and other outcomes shown. If the resulting policy exhibits some undesirable characteristic, say an unacceptable increase in transplant rate disparities by blood type, an additional constraint could be applied and the optimization resolved to arrive at a new policy.

A natural extension of the above framework is the ability to construct tradeoff curves for simulated outcomes, as depicted in Figure 3. Here, the optimization algorithm is invoked a number of times to generate policies on the efficient frontier of two outcomes. Each optimization run attempts to minimize waitlist mortality (y-axis), but the upper bound for median organ transport distance constraint (x-axis) is systematically varied across a range of values, resulting in policies with different placement efficiency characteristics. As the upper bound on distance increases, the optimization selects attribute weights that increasingly favor medical urgency over proximity in the composite allocation score. Initially, the added flexibility allows policies to realize a significant reduction in mortality rate as organs are transplanted to sicker patients farther away. At a certain point, however, this reduction exhibits diminishing returns, providing valuable insight into the projected benefits of broader distribution.

Although harder to visualize, the optimization methodology is able to search over the efficient frontier of any number of additional outcomes. We include in the same figure a set of optimized policies, generated as before, but with a ‘guardrail’ constraint that dictates that policies should not increase disparities in transplant rates by candidate blood type and height group vis-à-vis current policy. That the ‘No Increase’ policies fall on the same curve as the ‘Unconstrained’ ones suggests that lung CD policies are able to achieve essentially the same mortality benefit per mile of additional organ transport distance without exacerbating disparities in access to transplant by candidate blood group.

III.C. Disclaimer

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

The data reported here have been supplied by the Hennepin Healthcare Research Institute (HHRI) as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the
Choose Outcome to Optimize

Minimize # Waitlist Deaths

| Constrained Outcomes                                      | Upper Bound Value | % relative current |
|-----------------------------------------------------------|-------------------|--------------------|
| Median Organ Transport Distance (nautical miles)          |                   |                    |
|                                                            | 75                | 201                | 380                | 120.00%            |
| Transplant rate disparities by Blood Type (TX / patient-year) | 0.02              | 0.12               | 0.67               | 100.00%            |
| Transplant rate disparities by Height Group (TX / patient-year) | 0.15              | 0.32               | 0.79               | 100.00%            |
| Transplant rate disparities by Age Group (TX / patient-year) | 0.11              | 0.36               | 0.42               | 100.00%            |

**Figure 2.** Slider based optimization interface. Any simulation outcome can be selected either as the overarching objective outcome (here, minimizing # Waitlist Deaths), or added as a constraint by providing an upper bound on its value vis-à-vis some reference policy. The optimization methodology produces a conforming allocation policy in seconds, whose predicted outcomes can be used to refine the objective and constraints further. Source: Author-created illustration.
Figure 3. Efficient frontier of mortality versus median organ transport distance in lung CD policy. Each circle represents a different optimization-derived policy, with different relative score attribute weights, selected to minimize waitlist mortality and with a varying upper bound on median organ transport distance. Triangles are similarly selected, but including additional optimization constraints to enforce that the discrepancy in transplant rates across patient blood types and height groups does not increase vis-à-vis current policy. Source: Authors’ analysis of Thoracis Simulated Allocation Model (TSAM) simulations.

author(s) and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government.

IV. DISCUSSION

As the OPTN case study shows, leveraging the predictive power of machine learning to replace the bottleneck policy evaluation step with near-instantaneous model-based evaluations opens up exciting possibilities for refining ethicists’ and other stakeholders’ input to data-driven policymaking. One of the chief advantages of the proposed approach is that it enables ethicists and communities to offer input at precisely the points at which their views have the most value. To illustrate, consider the following example of lung allocation policymaking. Asking ethicists or community members whether a lung candidate attribute estimating posttransplant survival and another attribute measuring waitlist survival should both be weighted at 34% of the composite allocation score, or one reduced by 2%, would likely elicit blank stares or weak intuitions at best. Asking instead ‘would you be willing to tolerate a 10% decrease in average years of post-transplant survival if it meant a 5% reduction in waitlist mortality,’ and showing how the tradeoff changes as a ‘slider’ moves, poses a set of considerations for which ethical theory and/or community sentiment is likely to generate more useful feedback. We view this tool as an important way to improve the ability of key stakeholders without medical expertise, including donor and recipient populations, to influence policymaking. This approach also enables ethicists and community groups to clearly prespecify ‘guardrails,’ for example, that disparities
between blood type or racial groups not increase at all, or more than a specified amount, and show in near-real time how including, removing, or modifying those guardrails will change results.

While we view this approach as a major step forward for ethically informed policymaking, it does not solve every problem. First among the difficulties that persist is that modeling is only as good as the data on which it is based and the assumptions upon which it relies. In general, our approach works best when policymakers have access to high-quality data and analytical tools to accurately predict outcomes, e.g. counterfactual simulation in the case of organ allocation. Even then, an ethicist or a community stakeholder may raise a question about the effect of a policy change—for example, effects on patients with intellectual disabilities—that is not captured in the input data and cannot be accurately modeled in a tradeoff curve. Similarly, predictions in applications areas where practitioners have little historical experience (e.g., COVID-19 vaccine allocation), are likely to rely on assumptions that should be closely examined. Recent work has shown how particular modeling choices in, say, curation of the training data or selection of objective metric, can result in biased model recommendations. Fortunately, the machine learning and optimization communities have made significant headway in developing methods to identify and mitigate such model-based biases.

In practice, data and model limitations should also be clearly communicated to stakeholders during the design process, to avoid overreliance on imperfect predictions and to allow clinical and subject-matter expertise to influence conclusions drawn from the analysis. The ideal solution is to use representative, high-quality data sets, when that is not possible to be transparent about the limitations of the data sets being used, and when those limitations are serious enough to reconsider whether the model can do the work we want out of it. That being said, we note that such limitations also beset older, simpler approaches to policymaking, and one advantage of our framework is that it streamlines determining where current modeling practices fall short in addressing ethically important questions; this in turns sets up the possibility of changing those practices to fill the gaps.

Second, when done right, the approach discussed here seeks input from a multiplicity of stakeholders including ethicists, transplant professionals, donor communities, potential recipients, and sub-stratifications such as racial and ethnic minorities. The introduction of new analytical tools does not obviate the need for wide-ranging, cross-cutting deliberation to reach consensus, particularly as engaging with these stakeholders also means engaging with their biases. We believe our approach makes it easier for stakeholders to more meaningfully assess and weigh relevant tradeoffs, but it does not guarantee more convergence between the groups. How should a policymaker respond when various communities champion different tradeoffs? Should all views be treated equally? Many ethicists will chafe at the idea of ‘ethics by headcount’ in the sense of aggregating, without exploring the reasons behind, preferences for tradeoffs

---

26 Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF L. REV. 671 (2016).
27 Harini Suresh & John Guttag, *A Framework for Understanding Sources of Harm Throughout the Machine Learning Life Cycle*, ACM Conference on Equity and Access in Algorithms, Mechanisms and Optimization, 1-9 (2021).
28 Jon Kleinberg, et al., *Discrimination in the Age of Algorithms*, 10 J LEGAL ANAL, 113 (2018).
and potentially dismissing some as inconsistent or problematic. Though the federal regulation governing U.S. organ allocation policy provides a final safeguard to rule out legally inconsistent policy options, this Final Rule does not precisely dictate which tradeoffs are out of bounds. Again, this problem is not new; in making tradeoffs more visible and accessible, our approach may make the problem more common—though some may view this more a feature (by inviting greater engagement with the tradeoffs) than a weakness.

Finally, there often exist significant cultural and institutional hurdles that must be overcome for the adoption of advanced analytical tools like the one we propose. In the case of organ allocation, US policymakers over the past two decades have most often used trial-and-error simulation to explore policy options. The shift to include advanced machine learning and optimization methods for helping develop a lung CD policy challenged the community’s flexibility to accommodate a new approach. Despite the introduction of new methodology, many of the core components of the age-old policy-making process remained in place including, committee discussion, public comment feedback from the community, and SRTR simulation modeling. The additional analyses, particularly optimized tradeoff curves (Figure 3), provided extra scrutiny of the proposal and helped stakeholders home-in on a final set of policy options worth consideration. Leveraging these tradeoff curves also provided a gentle introduction to the use of advanced mathematical methods that could pave the way for broader community acceptance of even greater reliance on goal-driven optimization in future allocation policy development.

Although some in the transplantation community might struggle with more sophisticated analytical tools, others see their adoption as an opportunity to improve the historically time-consuming policy development process. For years, committee discussion and retrospective data analysis informed policy proposals that were then modeled by the SRTR. However, with these new tools, the community and committee can feel more confident about their chosen allocation policy options before the final, confirmatory simulation modeling is conducted. Over time, these more complex models can gain public confidence as new-and-improved policies are implemented and demonstrated to meaningfully improve outcomes for patients awaiting organ transplantation.

More generally, one might worry that, given relative unfamiliarity with machine learning among ethicists and stakeholder communities of interest, some may find it difficult to understand precisely what the approach we describe here ‘does’ or be concerned about ‘not seeing the whole picture.’ To remedy this gap will require thoughtful attempts at scientific communication that meets stakeholders ‘where they live,’ rather than a one-size-fits all strategy. Policymakers need to be sensitive to algorithmic aversion and key opportunities to manage it including the use of interactive tools like ‘sliders’ that enable stakeholders to see how AI/ML works even if they will never get ‘under the hood.’ 29 For healthcare policymakers focused on allocation controversies, it might also be worth considering, depending on the setting, combining the machine

29 Berkeley J. Dietvorst, et al., Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms if They Can (Even Slightly) Modify Them, 64 MANAGE SCI 983 (2018).
learning methods discussed here with deliberative democracy techniques, such as deliberative polling, consensus conferences, citizen juries, etc.\textsuperscript{30,31,32}

**DISCLOSURES**

IGC is a member of the OPTN/UNOS Ethics Committee. The views expressed in this paper are his own and do not represent the views of that Committee or OPTN/UNOS.

\textsuperscript{30} Archon Fung & Erik Olin Wright, *Deepening democracy: Institutional innovations in empowered participatory governance* §4 (Verso, 2003).

\textsuperscript{31} Archon Fung, *Survey Article: Recipes for Public Spheres: Eight Institutional Design Choices and their Consequences*, 11 J Polit Philos 338 (2003).

\textsuperscript{32} Michael M Burgess, *From ‘Trust Us’ to Participatory Governance: Deliberative Publics and Science Policy*, 23 Public Underst Sci 48 (2014).