Balancing learning and targeting in predictive allocation

BRYAN WILDER
Carnegie Mellon University
and
PIM WELLE
Allegheny County Department of Human Services

1. INTRODUCTION

This letter provides an overview of our recent work on "Learning Treatment Effects While Treating Those in Need" (published at the 2025 ACM Conference on Economics and Computation) as well as a more general perspective on design goals for algorithmic systems that are used to allocate limited resources in policy settings. Our motivation is the kind of algorithms that are used widely at present to prioritize candidates for various kinds of social interventions: public housing assistance, drop-out prevention programs in education, unconditional cash transfers in development, or a variety of other social services. By far the most common way of constructing such systems is the lens of *predictive allocation*: the algorithm designer identifies an outcome that the program seeks to alter (long-term homelessness, dropping out of school, etc) and constructs a predictive model for that outcome [Vaithianathan and Kithulgoda 2020; Aiken et al. 2022; Pan et al. 2017; Toros and Flaming 2017]. Candidates are ranked by predictions of risk so that, e.g., limited spots in a housing program might be offered to those at greatest predicted risk of long-term homelessness.

We argue that algorithm designers should take a more expansive view of the design goals that such systems aim to optimize. Our work focuses specifically on the tradeoff between predictive targeting and causal learning: if resources are allocated strictly according to predicted risk, it becomes difficult to tease out to what extent the program improves outcomes for its beneficiaries. Learning such causal effects requires experimentation, where otherwise similar candidates sometimes receive different allocation decisions. However, randomized experimentation would depart from the goal of strictly aligning allocations with predictions of risk. The fear of denying services to higher-need applicants is one of the main reasons that randomized controlled trials are relatively uncommon in many policy settings. However, high-quality evidence about program effectiveness is absolutely critical in order to improve services over time.

Our work asks how stark the tradeoffs really are. We construct a multiobjecitve

Authors' addresses: `bwilder@cmu.edu`, `Paul.Welle@alleghenycounty.us`

framework for optimizing a policy for allocating a limited resource. The goal is to optimally trade off between the goals of (1) allocating resources to individuals that have been identified as having higher levels of risk or need; and, (2) introducing enough random variation to accurately estimate the average treatment effect of the intervention. Our method allows policymakers to explore the Pareto frontier between these goals. We apply this framework to data on human service programs in Allegheny County, Pennsylvania and find that the tradeoffs between these goals are empirically quite tractable: with careful design, policymakers can get most of one without giving up too much of the other.

More broadly, we suggest that algorithm designers should incorporate other goals (like causal evaluation) explicitly into the design of allocative systems. Framing allocation as a prediction problem naturally suggests that the best policy is one that aligns with the most accurate predictive model. Prediction may often be a reasonable way to construct a proxy for an individual's need for a service. However, it is still only a proxy (and one which we have shown in other work may not always align well with other goals, like benefit from an intervention [Sharma and Wilder 2025]). It may be very worthwhile to give up a bit of predictive performance in order to ensure that algorithmic systems reflect the much broader array of challenges for decision making in social systems.

2. A MULTIOBJECTIVE FRAMEWORK FOR LEARNING AND TARGETING

The specific setting studied in our paper concerns the allocation of a single limited resource. The policymaker chooses a function $p : \mathcal{X} \to [0, 1]$ which maps an individuals' features $X \in \mathcal{X}$ to the probability with which they are allocated treatment. The policymaker specifies upfront a function $u(X)$ which specifies their preferences for which individuals should be allocated the resource; we refer to u as their *targeting utility function*. In our running example, u is the output of a predictive model for some adverse outcome: the policymaker prefers that the resource be allocated to individuals with higher levels of predicted risk. However, our algorithmic framework is agnostic as to how u is constructed. Individuals have potential outcomes $(Y(0), Y(1))$, where outcome $Y(1)$ is realized if they receive the intervention and $Y(0)$ is realized if they do not. We assume that $(X, Y(0), Y(1))$ are drawn iid from some joint distribution. The average treatment effect of the intervention is given by $\tau = \mathbb{E}[Y(1) - Y(0)]$. Our goal is to design an allocation p which will enable accurate estimates of τ while still scoring highly according to u. Allocations are subject to a budget constraint that at most a b fraction of the population can receive the intervention, i.e., that $\mathbb{E}[p(X)] \leq b$.

To formalize the goal of accurately estimating treatment effects, we turn to the semiparametric efficiency bound for the average treatment effect [Hahn 1998]. Suppose that we construct an estimate $\hat{\tau}$ for the ATE using any of the many methods developed in the causal inference literature. We would like to minimize the expected error in the estimate, $\mathbb{E}[(\hat{\tau} - \tau)^2]$. Since standard estimators are unbiased, the entirety of the mean-squared error in an estimate is given by the variance of the estimator. This is precisely the quantity that the efficiency bound gives: the smallest possible variance for any estimator.

We show that, under many circumstances, the efficiency bound is itself upper bounded by a function of the form

$$\mathbb{E}\left[\frac{a_1(X)}{p(X)} + \frac{a_0(X)}{1-p(X)}\right]$$

for a pair of functions $a_1(X)$ and $a_0(X)$. Intuitively, the performance of estimators for the treatment effect degrades rapidly as $p(X)$ gets closer to 0 or 1, i.e., as allocations get closer to deterministic. The efficiency bound is exactly given by taking a_0 and a_1 to be the conditional outcome variances, $a_0(X) = \mathrm{Var}(Y(0)|X)$ and $a_1(X) = \mathrm{Var}(Y(1)|X)$ [Hahn 1998]. Since the variances are typically unknown before collecting experimental data, we show how structural properties, side information, or assumptions about the distribution can be used to construct an upper bound on the worst-case variance. The particular values of a_0 and a_1 for the optimization problem then arise from such choices (e.g., stipulating that outcomes are binary, or using historical data to learn about $Y(0)$).

One possible allocation policy is to spend the entire budget on the b-fraction of the population with the highest value of $u(X)$. This corresponds to the predictive allocation strategy often employed in practice. However, since $p(X)$ is either 0 or 1 for every X, estimating treatment effects is impossible. At the other end of the spectrum is a completely uniform allocation policy, $p(X) = b$ for all X, which corresponds to a randomized controlled trial. This maximizes the power to estimate treatment effects (formally, minimizes the efficiency bound when a_1 and a_0 are constant) but entirely sacrifices the targeting goal.

While these extremes are the strategies typically employed in practice (predictive allocation most frequently, randomized trials rarely), our paper proposes an entire spectrum of allocation policies in between. Formally, these are derived as solutions to a family of multiobjective optimization problems that balance the targeting and causal learning goals. In particular, we aim to find allocation probabilities p which solve an optimization problem of the form

$$\begin{aligned} &\min_p \mathbb{E}\left[\frac{a_1(X)}{p(X)} + \frac{a_0(X)}{1-p(X)}\right] \\ &\mathbb{E}[p(X)u(X)] \geq c \\ &\mathbb{E}[p(X)] \leq b \\ &p(X) \in [\gamma, 1-\gamma] \quad \forall X \in \mathcal{X}. \end{aligned} \tag{1}$$

where the parameter c is chosen by the policymaker to control the extent to which allocation should align with their targeting utility function u. As discussed in our work, this core formulation is also easily extensible to include constraints enforcing additional goals. The last constraint exerts limited influence on the actual solution since optimal policies will never have $p(X)$ at 0 or 1 by virtue of the objective function. However, the parameter γ measuring closeness to deterministic solutions appears in the exact statement of our sample complexity bounds, with more samples required when the designer wishes to optimize over closer-to-deterministic policies. A complete discussion of this topic can be found in our full paper.

This is a strictly convex optimization problem in p, subject to linear constraints. The challenge is that this optimization problem is over the space of policies, i.e., functions $p : \mathcal{X} \to [0, 1]$. In other policy learning settings, a typical approach is to optimize over a parameterized class of policies (e.g., linear models, decision trees, etc.), with sample complexity bounds for optimal policy learning then depending on a measure of the complexity of this class like the VC dimension [Athey and Wager 2021; Chernozhukov et al. 2019; Swaminathan and Joachims 2015].

In our work, we show that solutions to the optimization problem above have a special structure. Effectively, the function p can be parameterized very concisely, just via a handful of parameters that correspond to the dual variables. This viewpoint enables efficient optimization and learning. To formalize this idea, we take the dual of the population optimization problem. Let the constraints be collectively denoted as functions g_j, $j = 1...J$. The dual is

$$\max_{\lambda \geq 0} \min_{p \in [\gamma, 1-\gamma]^{\mathcal{X}}} \mathbb{E}\left[\frac{a_1(X)}{p(X)} + \frac{a_0(X)}{1-p(X)} + \sum_{j=1}^{J} \lambda_j (g(p(X), X) - c_j)\right]$$

where λ_j is the dual variable associated constraint j. The dual has the attractive property that the inner objective is separable across X, allowing us to push the min inside the expectation:

$$\max_{\lambda \geq 0} \mathbb{E}\left[\min_{p(X) \in [\gamma, 1-\gamma]} \frac{a_1(X)}{p(X)} + \frac{a_0(X)}{1-p(X)} + \sum_{j=1}^{J} \lambda_j (g(p(X), X) - c_j)\right].$$

The strategy proposed in our work is to estimate the dual parameters using a sample of X's drawn from the distribution of interest by solving the *sample* problem

$$\max_{\lambda \geq 0} \frac{1}{n} \sum_{i=1}^{n} \left[\min_{p(X_i) \in [\gamma, 1-\gamma]} \frac{a_1(X_i)}{p(X_i)} + \frac{a_0(X_i)}{1-p(X_i)} + \sum_{j=1}^{J} \lambda_j (g(p(X_i), X_i) - c_j)\right]. \quad (2)$$

This is a strictly convex optimization problem in n variables that can easily be solved using standard methods. Then, to compute the optimal allocation probability $p(X)$ for any given individual, we solving the inner minimization problem at the estimated dual parameters, i.e., we compute

$$\widehat{p}(X) = \text{argmin}_{p(X) \in [\gamma, 1-\gamma]} \frac{a_1(X)}{p(X)} + \frac{a_0(X)}{1-p(X)} + \sum_{j=1}^{J} \widehat{\lambda}_j (g(p(X), X) - c_j) \quad (3)$$

separately for each individual and randomize them to be treated with probability $\widehat{p}(X)$.

Theoretically, we establish that this strategy enjoys strong sample complexity guarantees. Intuitively, the class of policies that we optimize over has only J parameters, one per constraint (of which there were two in the core formulation discussed above). We show that the number of samples from the covariate distribution required in order to produce an approximately optimal policy scales at parametric, root-n rates, with the complexity of the policy class represented by a function of J:

Proposition 1 *[informal]: In order to obtain policy $\hat{p}$ that is within ϵ of both optimality and feasibility for Problem 1, it suffices to have $n = O\left(\frac{J^3}{\epsilon^2}\right)$ samples.*

Empirically, we simulate the policies output by our method on historical data

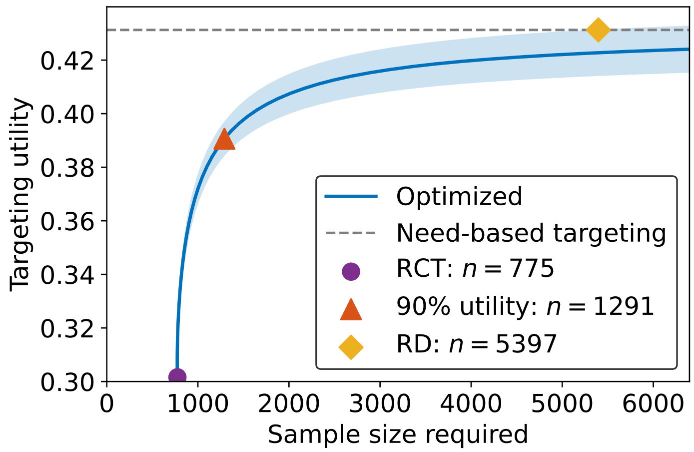

Fig. 1. Tradeoff between predictive targeting and causal learning on historical data from the Allegheny County Human Services Department. The horizontal axis shows the number of samples that would be required to power an estimate of the average treatment effect. The vertical axis shows the fraction of individuals who would reenter jail in the next year that are targeted for the program. "Optimized" refers to the Pareto frontier of designs output by our method while "need-based targeting" refers to allocating entirely to individuals with the highested predicted risk. "'RCT" gives the performance of a randomized controlled trial that uniformly allocates the treatment. "90% utility" labels the point on the Pareto curve output by our method that achieves 90% of the targeting utility of need-based targeting. Finally, "RD" shows the number of samples required to power a regression discontinuity estimate of the local treatment effect.

from the Allegheny County Human Services Department. In this letter, we reproduce one example result, which uses administrative data on the outcomes of formerly incarcerated people reentering the community. We mimic standard practice by fitting a machine learning model to predict the probability that a given individual will reenter jail within the next month. Our evaluation models a hypothetical program which offers an intervention to avert jail reentry. The policymaker would like to offer this program preferentially to individuals with a higher probability of reentering jail and we characterize the tradeoff between this targeting goal and learning causal effects. To provide a concrete interpretation for power to learn causal effects, we rescale asymptotic variance into the number of samples that would be required to estimate of the average treatment effect with 5% type-1 error and 80% power.

Our main empirical result is that the tradeoff between these goals is substantially mitigated by optimal policy design. Figure 1 shows the tradeoff in the form the Pareto curve output by our method, obtained by varying the value of the targeting utility constraint c. At one end of the spectrum we have a randomized controlled trial that allocates the intervention uniformly, without consideration for risk. As the value of c increases, the allocation converges towards thresholding on the predicted risk. However, the shape of the curve in between is highly concave, indicating that the policymaker can obtain most of one goal without giving up too much of the other. For example, the allocation policies output by our method achieve 90% of the maximum possible performance at targeting high-risk individuals while requiring less than twice the number of samples in order to estimate treatment effects as the ideal randomized trial. Similar findings surface across a number of other programmatic settings and simulations. At the far end of the spectrum, we compare to the number of samples that would be required to power an estimate of a local average treatment effect via a regression discontinuity design if the policymaker instead used pure predictive allocation. The number of samples is much larger, several times that required by our method at the elbow of the curve. This indicates that the price (in terms of causal learning) required to eliminate the last bit of randomization is very high, imposing a sample size which is infeasible for many programmatic settings.

3. DISCUSSION AND CONCLUSION

We have shown how the targeting interventions to people at higher levels of present need can be reconciled with the goal of introducing randomization to learn whether these interventions produce real benefits. There is still research to be done in order to bring these ideas to the operational reality of public services. For example, in this work, we focused on designing policies that adhere to a set of constraints in-expectation on a fixed distribution. In reality, allocations are subject to a variety of other complications in implementation, like hard budget constraints, dynamic arrivals of applicants and resources, drift in covariate distributions, and so on. In subsequent work, we showed how techniques from combinatorial optimization allow us to enforce hard budget constraints (which requires dependent assignments) while still providing statistical valid estimates of treatment effects [Yamin et al. 2025]. In fact, we find that doing so actually provides substantial variance reduction benefits and may be useful in other experimental design settings as well. However, a variety of more complex settings remain open.

Broadly though, the takeaway remains: by integrating causal learning as a goal alongside prediction in designing allocation policies, it is often possible to get much of the best of both worlds in practice. Our hope is that these kinds of techniques enable much more widespread use of rigorous program evaluation strategies by lowering the cost (in terms of foregone targeting) to running randomized trials. While prediction is one way of encoding goals for allocation, we believe that many settings will benefit at least as much from better, faster, and more widely available evidence about "what works". Is is tempting for algorithm designers to focus on the details of allocation, trying to optimize the individual-level match between people and potential interventions. But this focus should not crowd out our ability to learn and improve interventions over time, ultimately improving the quality of services available to everyone.

4. ACKNOWLEDGMENTS

This material is based upon work supported by the AI Research Institutes Program funded by the National Science Foundation under AI Institute for Societal Decision Making (AI-SDM), Award No. 2229881, and the AI2050 program at Schmidt Sciences (Grant #G-22-64474).

REFERENCES

AIKEN, E., BELLUE, S., KARLAN, D., UDRY, C., AND BLUMENSTOCK, J. E. 2022. Machine learning and phone data can improve targeting of humanitarian aid. *Nature 603,* 7903, 864–870.

ATHEY, S. AND WAGER, S. 2021. Policy learning with observational data. *Econometrica 89,* 1, 133–161.

CHERNOZHUKOV, V., DEMIRER, M., LEWIS, G., AND SYRGKANIS, V. 2019. Semi-parametric efficient policy learning with continuous actions. *Advances in Neural Information Processing Systems 32.*

HAHN, J. 1998. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 315–331.

PAN, I., NOLAN, L. B., BROWN, R. R., KHAN, R., VAN DER BOOR, P., HARRIS, D. G., AND GHANI, R. 2017. Machine learning for social services: a study of prenatal case management in illinois. *American journal of public health 107,* 6, 938–944.

SHARMA, V. AND WILDER, B. 2025. Comparing targeting strategies for maximizing social welfare with limited resources. In *The Thirteenth International Conference on Learning Representations.*

SWAMINATHAN, A. AND JOACHIMS, T. 2015. Batch learning from logged bandit feedback through counterfactual risk minimization. *The Journal of Machine Learning Research 16,* 1, 1731–1755.

TOROS, H. AND FLAMING, D. 2017. Prioritizing which homeless people get housing using predictive algorithms. *Available at SSRN 2960410*.

VAITHIANATHAN, R. AND KITHULGODA, C. I. 2020. Using predictive risk modeling to prioritize services for people experiencing homelessness in allegheny county: Methodology update.

YAMIN, K., KENNEDY, E., AND WILDER, B. 2025. Dependent randomized rounding for budget constrained experimental design. In *The 41st Conference on Uncertainty in Artificial Intelligence.*