DOI: 10.1111/1475-6773.13586

BEST OF ARM 2020

HSR Health Services Research

Predicting preventable hospital readmissions with causal machine learning

Ben J. Marafino BS[1] | Alejandro Schuler PhD[2] | Vincent X. Liu MD, MS[2,3] | Gabriel J. Escobar MD[2] | Mike Baiocchi PhD[4]

[1]Biomedical Informatics Training Program, Department of Biomedical Data Science, School of Medicine, Stanford University, Stanford, California, USA

[2]Systems Research Initiative, Kaiser Permanente Division of Research, Oakland, California, USA

[3]Critical Care Medicine, Kaiser Permanente Medical Center, Santa Clara, California, USA

[4]Departments of Epidemiology & Population Health, Department of Medicine, Stanford University, Stanford, California, USA

Correspondence
Ben J. Marafino, BS, Biomedical Informatics Training Program, Department of Biomedical Data Science, School of Medicine, Stanford University, Stanford, CA 94305, USA.
Email: marafino@stanford.edu

Funding information
The Permanente Medical Group; Stanford University School of Medicine; National Institute of General Medical Sciences, Grant/Award Number: R35GM128672; Kaiser Foundation Hospitals; Agency for Healthcare Research and Quality, Grant/Award Number: KHS022192A; U.S. National Library of Medicine, Grant/Award Number: T15LM007033

Abstract

Objective: To assess both the feasibility and potential impact of predicting preventable hospital readmissions using causal machine learning applied to data from the implementation of a readmissions prevention intervention (the *Transitions Program*).

Data Sources: Electronic health records maintained by Kaiser Permanente Northern California (KPNC).

Study Design: Retrospective causal forest analysis of postdischarge outcomes among KPNC inpatients. Using data from both before and after implementation, we apply causal forests to estimate individual-level treatment effects of the Transitions Program intervention on 30-day readmission. These estimates are used to characterize treatment effect heterogeneity and to assess the notional impacts of alternative targeting strategies in terms of the number of readmissions prevented.

Data Collection: 1 539 285 index hospitalizations meeting the inclusion criteria and occurring between June 2010 and December 2018 at 21 KPNC hospitals.

Principal Findings: There appears to be substantial heterogeneity in patients' responses to the intervention (omnibus test for heterogeneity $p = 2.23 \times 10^{-7}$), particularly across levels of predicted risk. Notably, predicted treatment effects become more positive as predicted risk increases; patients at somewhat lower risk appear to have the largest predicted effects. Moreover, these estimates appear to be well calibrated, yielding the same estimate of annual readmissions prevented in the actual treatment subgroup (1246, 95% confidence interval [CI] 1110-1381) as did a formal evaluation of the Transitions Program (1210, 95% CI 990-1430). Estimates of the impacts of alternative targeting strategies suggest that as many as 4458 (95% CI 3925-4990) readmissions could be prevented annually, while decreasing the number needed to treat from 33 to 23, by targeting patients with the largest predicted effects rather than those at highest risk.

Conclusions: Causal machine learning can be used to identify preventable hospital readmissions, if the requisite interventional data are available. Moreover, our results suggest a mismatch between risk and treatment effects.

KEYWORDS
patient readmission, risk assessment, machine learning, clinical decision rules

1 | INTRODUCTION

Unplanned hospital readmissions represent an undesirable outcome following a hospitalization, but are common, costly, and associated with substantial morbidity and mortality, occurring within 30 days following nearly 20% of hospitalizations by Medicare beneficiaries.[1] In 2011, 3.3 million patients in the United States were readmitted to the hospital within 30 days, incurring costs of $41 billion.[2] In 2012, responding the growing awareness of the toll of readmissions, the Centers for Medicare and Medicaid Services introduced the Hospital Readmissions Reduction Program (HRRP), which penalizes hospitals with risk-adjusted 30-day readmission rates higher than the average. As a consequence of the HRRP and other value-based care initiatives, many hospitals and health care systems in the United States have since implemented quality improvement (QI) initiatives and population health management programs relying on risk assessment tools to identify hospitalized patients at high risk of readmission. Tailored interventions can be then targeted to these patients immediately following discharge, with the goal of preventing their readmission. The effectiveness of these interventions in preventing readmissions has been mixed, and the precise mechanisms through which they do so remain unclear.[3-9]

Many risk assessment tools used in these efforts apply statistical modeling or supervised machine learning to estimate readmission risk among hospitalized patients based on data prior to discharge.[8-13] Stakeholders select a risk threshold with respect to resource constraints, so that an intervention is to be delivered to all patients above this threshold, while those below it receive usual care. Underlying many population health management and QI efforts aimed at reducing readmissions is the implicit assumption that the patients most at risk are also those most likely to benefit from the intervention.[8,14-16] Ostensibly, this assumption has intuitive appeal, given that higher-risk patients appear to have "more room to move the needle," but it is not guaranteed to hold in practice[17,18], especially in the context of readmissions[7,19] and other settings where treatment effect heterogeneity may exist.[18]

The need for analytical approaches that estimate patient-level benefit—referred to in some contexts as *impactibility*[20-23] and falling under the umbrella of *precision medicine* more generally—is beginning to be recognized, particularly for readmission reduction programs.[22] However, the distinction between benefit and risk may currently be overlooked in the development and application of risk assessment tools. Individual benefit is often expressed in terms of treatment effects, which cannot be estimated by modeling outcome risk. Predicting, for example, a readmission risk of 60% for a patient provides no information on their counterfactual risk if they were to receive a certain readmissions reduction intervention. The actual counterfactual risk for this hypothetical patient could be unchanged, on average, corresponding to no effect for the intervention. On the other hand, the effect of this intervention may be heterogeneous across levels of predicted risk, so that, for example, this patient experiences an absolute risk reduction (ARR) of 10% as a result of the intervention, while another patient at a predicted risk of 30% experiences an ARR of 20%. Given limited resources, a decision maker may wish to give the intervention to the latter patient. Indeed, when it comes to preventing readmissions, there is growing evidence that higher-risk patients—referred to in some contexts as "super-utilizers"[7]—may be less sensitive to a class of care coordination interventions relative to those at lower risk.[19,22,24]

Moreover, efforts targeting a preventative intervention based on predicted risk also fail to take into account that low-risk patients comprise the majority of readmissions.[25] That the majority of poor outcomes are experienced by patients at low risk, but who would not have been selected to receive an intervention, has previously been observed in a range of predictive modeling problems in population health management.[8] Thus, targeting a preventative intervention so as to include lower-risk patients among whom they may be effective, rather than targeting them only to high-risk patients, may potentially prevent more readmissions than the latter strategy.[26-28]

Few, if any, analytical approaches to identify "care-sensitive" patients, or those whose outcomes may be most "impactible," currently exist, despite a clear need for such methods.[20,23] Existing approaches based on off-the-shelf supervised machine learning methods, despite their flexibility and potential predictive power, cannot meet

What is known on this topic

- Readmission risk assessments are widely used by hospitals and health systems to target readmission prevention interventions to inpatients immediately postdischarge, with a focus on those at highest risk.

What this study adds

- Using data from 1.5 million hospital discharges in an integrated health system from before and after the implementation of a readmission prevention intervention, we find evidence for risk treatment effect mismatch in this setting: patients at high predicted risk of 30-day readmission appeared to derive less benefit compared to low-risk patients.
- Our results may have implications for the design of readmission prevention programs and related initiatives: targeting preventative and quality improvement interventions based on estimated benefit, and not estimated risk, may maximize aggregate benefit.

this need.[18] In this study, we demonstrate the feasibility of applying causal machine learning to identify preventable hospital readmissions with respect to a readmission prevention intervention via modeling its heterogeneous treatment effects. In our setting, the "preventability" of a readmission is not based on predefined, qualitative criteria, as in prior work.[29,30] Rather, it is expressed in quantitative terms: the greater the treatment effect on readmission estimated for a patient, the more preventable their potential readmission may be. To accomplish this, we leverage a rich set of data drawn from before and after the implementation of a comprehensive readmissions prevention intervention in an integrated health system1).

2 | METHODS

2.1 | Data and context

The data consist of 1 584 902 hospitalizations taking place at the 21 hospitals within Kaiser Permanente's Northern California region (hereafter KPNC) between June 2010 and December 2018. In particular, these data include patient demographics, diagnosis codes, laboratory-based severity of illness scores at admission and at discharge, and a comorbidity burden score that is updated monthly (Table 1). A subset of these data, which span from June 2010 to December 2017, have previously been described in greater detail.[31]

These data encompass a period where a comprehensive readmissions prevention intervention, known as the *Transitions Program*, began and completed implementation at all 21 KPNC hospitals from January 2016 to May 2017. The Transitions Program had two goals: (1) to standardize postdischarge care by consolidating a range of preexisting care coordination programs for patients with complex care needs and (2) to improve the efficiency of this standardized intervention by targeting it to the patients at highest risk of the composite outcome of postdischarge readmission and/or death. The Transitions Program intervention is principally a care coordination intervention centered around early primary care physician follow-up, shortly after discharge, and ongoing nursing assessment delivered by telephone in this 30-day period. A full description of the intervention is available in the Appendix S1.

As currently implemented, the Transitions Program relies on a validated predictive model for the risk of this composite outcome,[12]

TABLE 1 The covariates used in this study and their d

Covariate	Description	Mean (median; IQR)
AGE	Patient age in years, recorded at admission	65.0 (67; 54-78)
MALE	Male gender indicator	47.5% (–)
DCO 4	Code status at discharge (4 categories)	84.3% (–)
HOSP PRIOR7 CT	Count of hospitalizations in the last 7 d prior to the current admission	0.05 (0; 0-0)
HOSP PRIOR8 30 CT	Count of hospitalizations in the last 8 to 30 d prior to the current admission	0.11 (0; 0-0)
LOS 30	Length of stay, in days (with stays above 30 d truncated at 30 d)	4.6 (3; 2-5)
MEDICARE	Indicator for Medicare Advantage status	58.8% (–)
DISCHDISP	Discharge disposition (home, skilled nursing, home health)	72.7% (–)
LAPS2	Laboratory-based acuity of illness score, recorded at admission	55.7 (49; 16-84)
LAPS2DC	Laboratory-based acuity of illness score, recorded at discharge	44.5 (40; 24-60)
COPS2	Comorbidity and chronic condition score, updated monthly	44.7 (24; 10-66)
HCUPSGDC	Diagnosis supergroup classification	–
W (or W_i)	Treatment: Transitions Program intervention	5.2% (–)
Y (or Y_i)	Outcome: Nonelective readmission within 30 d postdischarge	12.4% (–)

Note: A more comprehensive listing of characteristics, stratified respective to the implementation of the Transitions Program, as well as definitions of the HCUPSGDC variables, can be found in the Appendix S1. For binary variables, only means are presented; for DCO_4 and DISCHDISP, the quantities presented correspond to the proportion of discharges who were full code, and those disharged to home, respectively.

Abbreviations: COPS2, COmorbidity Point Score, version 2; HCUPSGDC, Health Care and Utilization Project Super Group at discharge; IQR, interquartile range; LAPS2, Laboratory-based Acute Physiology Score, version 2.

which was developed using historical data from between June 2010 and December 2013. Following development and validation of this model by teams at KPNC's Division of Research, it was subsequently integrated into KP HealthConnect, KPNC's electronic health record (EHR) system to produce continuous risk scores, ranging from 0 to 100%, at 6:00 AM on the planned discharge day.

These risk scores are used to automatically assign inpatients awaiting discharge to be followed by the Transitions Program over the 30 days postdischarge. Inpatients with a predicted risk of ≥25% (medium or high) are assigned to be followed by the Transitions Program, and are considered to have received the Transitions Program intervention in this analysis. Inpatients with a predicted risk below 25% instead received usual postdischarge care, at the discretion of the discharging physician.

We used a subset of 1 539 285 hospitalizations taking place at 21 KPNC hospitals that meet a set of eligibility criteria. These criteria include: the patient was discharged alive from the hospital; age ≥18 years at admission; and their admission was not for childbirth (although postdelivery complications were included) nor for same-day surgery. Moreover, a readmission was considered nonelective if it began in the emergency department; if the principal diagnosis was an ambulatory care-sensitive condition[32]; or if the episode of care began in an outpatient clinic, and the patient had elevated severity of illness, based on a mortality risk of ≥7.2% as predicted by their laboratory-based acuity score (LAPS2) alone.

This project was approved by the KPNC Institutional Review Board for the Protection of Human Subjects, which has jurisdiction over all the study hospitals and waived the requirement for individual informed consent.

2.2 | From observational data to predicted treatment effects: causal forests

To identify potentially preventable readmissions, we undertake a causal machine learning approach using data taken from before and after the implementation of the Transitions Program at KPNC. Our causal machine learning approach is distinct from supervised machine learning as it is commonly applied in that it seeks to estimate individual *treatment effects* (or *lift*), and not outcome risk. Compared to other methods for studying treatment effect heterogeneity (eg, subgroup analyses), causal machine learning methods afford a major advantage in that they avoid potentially restrictive parametric assumptions, allowing a data-driven approach, while guarding against overfitting via regularization.

We express these individual treatment effects of the Transitions Program intervention in terms of the predicted conditional average treatment effects (CATEs), $\hat{\tau}_i$, which estimate

$$\tau_i(x) = E\left[Y_i(1) - Y_i(0) | X_i = x\right], \quad (1)$$

where $Y_i(1)$ and $Y_i(0)$ represent potential outcomes[33] of a 30-day readmission or no readmission within 30 days, respectively; E is the expectation operator; and X_i denotes the covariates x associated with patient i. Importantly, this quantity can be interpreted as the absolute risk reduction (ARR). It is through the sign and magnitude of these estimated CATEs that we consider a readmission potentially preventable: a $\hat{\tau}_i < 0$ denotes that the intervention would be expected to lower 30-day readmission risk for that patient, while a $\hat{\tau}_i > 0$ suggests that the intervention would be more likely to result in readmission within 30 days. A larger (more negative) CATE suggests a greater extent of preventability: that is, $\hat{\tau}_j < \hat{\tau}_i < 0$ implies that patient j's readmission is more "preventable"—their risk is more modifiable by the intervention—compared to patient i's.

To estimate these CATEs, we apply causal forests to the KPNC data described in the previous subsection. Causal forests[34] represent a special case of generalized random forests[35]; causal forests use a different loss function for placing splits and allow treatment effect estimates to be computed individually by each tree (see Section 6.2 of this work[35]). Our overall approach resembles that in prior work[36], which used them to study treatment effect heterogeneity in an observational setting. Causal forests can be viewed as a form of adaptive, data-driven subgroup analysis, and can be applied in either observational or randomized settings. As such, they do not make parametric assumptions regarding the relationships between the covariates and the treatment effect, which may give them more power to detect heterogeneity, if it exists, and to allow individual effects to be more accurately estimated.[37] In the Appendix S1, we describe some necessary assumptions that are required in order to identify the CATEs in our setting, and an omnibus test for heterogeneity[38] which we apply to establish quantitative evidence for treatment effect heterogeneity.

All analyses were performed in R (version 3.6.2); causal forests and the omnibus test for heterogeneity were implemented using the grf package (version 0.10.4). Causal forests were fit using default settings (except for the minimum node size, which was set to 10) with n = 8000 trees and per-hospital clustering. Propensity and marginal outcome models were estimated prior to fitting the causal forests and used to "orthogonalize" the forests.[39]

2.3 | Translating predictions into targeting strategies

A relevant question, once predictions have been made—be they of risk or of treatment effects—is how to translate them into treatment decisions. With predicted risk, these decisions are made with respect to some risk threshold, or to a decision-theoretic threshold (eg, as in prior work[11]). However, both approaches are potentially suboptimal in the presence of treatment effect heterogeneity, requiring strong assumptions to be made regarding the nature of the treatment effect, for example, assuming a constant effect.[11] Instead of prioritizing the patients most at risk, we prioritize those with the treatment effects $\hat{\tau}_i < 0$ that are largest in magnitude, that is, the most negative. We also describe in the Technical Appendix (S1) how estimates of "payoffs," or the gains associated with

successfully preventing a readmission, could be incorporated into this analysis.

To estimate the impacts of several notional targeting strategies that focus on treating those with the largest effects, we undertake the following approach. We begin by stratifying the patients in the dataset into 20 ventiles $V_1, \cdots, V_{20}$, of predicted risk, where V_1 denotes the lowest (0 to 5%) risk ventile, and V_{20} the highest (95% to 100%). Then, the causal forest is trained on data through the end of 2017 and used to predict CATEs for all patients discharged in 2018.

First, for all patients above a predicted risk of 25%, we compute the impact of the current risk- based targeting strategy by summing these predicted CATEs from 2018. This yields an estimate of the number of readmissions prevented, which we compare to another estimate using the average treatment effect of this intervention, from prior work.[40] This comparison serves as one check of the calibration of the predicted CATEs; the numbers of readmissions prevented should be similar. Second, we then use these same predicted CATEs from 2018 to assess three CATE-based targeting strategies, which treat the top 10%, 20%, and 50% of patients in each risk ventile V_j based on their predicted CATE. Similarly, we compute the NNT for a targeting strategy by taking the reciprocal of the average CATE of the patients notionally treated under that strategy, $1/\bar{\hat{\tau}}_i$. [Correction added on 6 November 2020, after first online publication: the mathematical expression has been amended to $1/\bar{\hat{\tau}}_i$.]

3 | RESULTS

3.1 | Overall characteristics of the cohort

From June 2010 to December 2018, 1 584 902 hospitalizations took place at the 21 KPNC hospitals represented in this sample. Further details regarding the overall cohort are presented in Table S1. Of these hospitalizations, 1 539 285 met the inclusion criteria, of which 1 127 778 (73.3%) occurred during the preimplementation period for the Transitions Program, and 411 507 (26.7%) during the postimplementation period. Among these 411 507 hospitalizations taking place postimplementation, 80 424 (19.5%) were predicted to be at risk of 30-day postdischarge mortality or readmission and were considered to have received the Transitions Program intervention postdischarge.

Of the patients whose index stays were included, their mean age was 65.0 years, and 52.5% were women. The overall 30-day nonelective rehospitalization rate among these index stays was 12.4%, and 30-day postdischarge mortality was 4.0%. Notably, patients at low risk (risk score < 25%) represented 63.3% of all readmissions throughout the study period, while making up 82.9% of index stays, compared to 36.7% of all readmissions among those at risk (≥25%), which represented 17.1% of index stays. The observed-to-expected ratios of the outcome in both groups followed similar trends prior to implementation, but diverged postimplementation; these rates are presented in Figure TA1 in the Appendix S1.

3.2 | Characterizing the treatment effect heterogeneity of the Transitions Program intervention

The estimated out-of-bag conditional average treatment effects (CATEs) yielded by the causal forest are presented in Figure 1. Qualitatively, these distributions exhibit widespread and suggest some extent of heterogeneity in the treatment effect of the Transitions Program intervention. In particular, treatment effects appear to be largest for patients discharged with a predicted risk of around 15% to 35% and appeared to attenuate as risk increased. Notably, particularly among patients at higher risk, some estimated effects were greater than zero, indicating that the intervention was more likely to lead to readmission within 30 days. Finally, the CATE estimates themselves also appeared well calibrated, in the sense that we identified no cases where an individual's CATE estimate was greater than their predicted risk.

Figure 2 is similar to the previous, but stratifies the display by Clinical Classification Software (CCS) supergroups. The overall pattern is similar to that in the unstratified plot, in that treatment effects appear to be greatest for patients at low to moderate risk, All supergroups appear to exhibit heterogeneity in treatment effect both within and across ventiles, which is more pronounced for some conditions, including hip fracture, trauma, and highly malignant cancers. Qualitatively, some supergroups exhibit bimodal or even trimodal distributions in the treatment effect of the Transitions Program intervention, suggesting identification of distinct subgroups based on these effects.

Quantitatively, fitting the best linear predictor using these CATE estimates (described in the Appendix S1) yields estimates of $\hat{\alpha} = 1.16$ and $\hat{\beta} = 1.06$, with $p = 5.3 \times 10^{-8}$ and 2.23×10^{-7}, respectively. Interpreting the estimate of β as an omnibus test for the presence of heterogeneity, we can then reject the null hypothesis of no treatment effect heterogeneity.

These effects can also be evaluated on a grid of two covariates to assess how the estimated CATE function changes as these covariates vary. This yields insight into the qualitative aspects of the surface of the CATE function and may help identify subgroups among which the Transitions Program intervention may have been more or less effective. Figure 2 shows the resulting CATE functions for two choices of patient age: 50 and 80 years, where the estimated CATE ranged from −0.060 to 0.025 (−6.0 to 2.5%). The estimated CATE was generally more negative—suggesting that the Transitions Program intervention became more effective with increasing age — at age 80 compared to 50.

Moreover, the estimated CATE tended to increase with increasing LAPS2DC. This finding suggests that for patients who were more ill at discharge (the average value of LAPS2DC in 2018 was 45.5), enrolling them in the Transitions Program may have actually encouraged them to return to the hospital. While this finding may appear surprising, it is unclear if it actually represents "harm" in the sense such a positive effect is usually interpreted; we discuss this finding in more depth in the Discussion. Additional dimensions of heterogeneity are presented in Figure S1 in the Appendix S1.

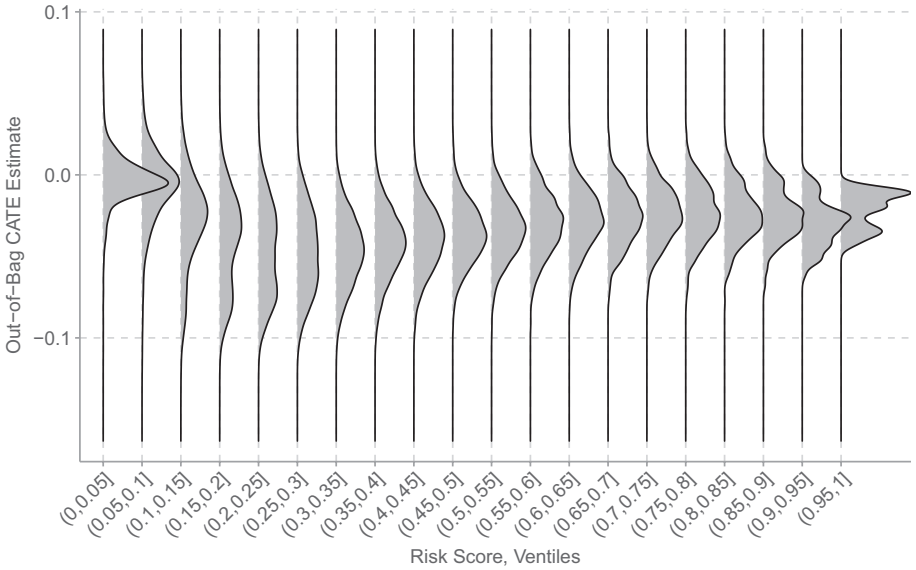

FIGURE 1 Treatment effect heterogeneity across risk score ventiles. The densities represent the distribution of estimated conditional average treatment effects within each ventile. They are drawn on a common scale, and hence do not reflect the variation in sample size across ventiles. The values in the bracket denote the risk range for that ventile; for example, (0.25, 0.3] represents all patients with predicted risk of 25 to 30%

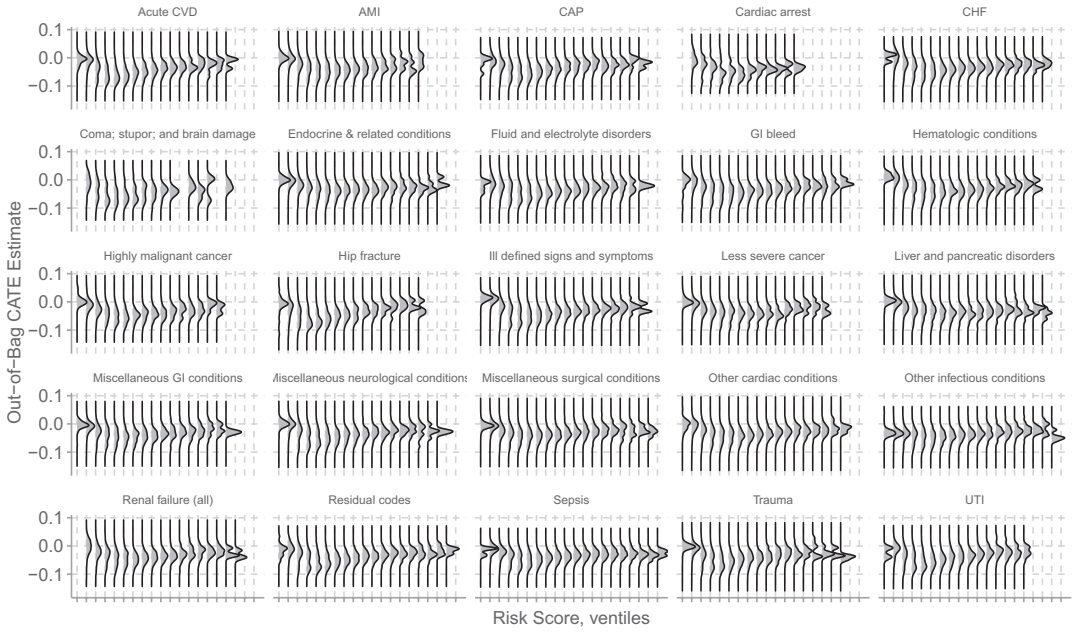

FIGURE 2 Treatment effect heterogeneity, stratified by discharge diagnosis supergroup. Treatment effect heterogeneity across risk score ventiles, stratified by Clinical Classification Software (CCS) supergroups based on the principal diagnosis code at discharge. A full listing of the definitions of these supergroups is given in Table S2 in the Appendix S1. Some ventiles are blank for some supergroups, because there were no patients belonging to those supergroups with predicted risks falling within those ranges. Abbreviations: AMI, acute myocardial infarction; CAP, community-acquired pneumonia; CHF, congestive heart failure; CVD, cerebrovascular disease; GI, gastrointestinal; UTI, urinary tract infection

3.3 | Notional estimates of overall impact under different targeting strategies

Based on these individual CATE estimates, we compute the potential impacts of several notional targeting strategies using these estimated effects, and not predicted risk, to target the Transitions Program intervention. These are given in Table 2. We first confirm the calibration of the effect estimates by taking the same group of patients who were intervened upon under the current risk-based strategy and use this group to estimate the number of readmissions

prevented, with the aim of comparing this number to a previous estimate of the impact of this policy using the average treatment effect.[40] This results in an estimate of 1246 (95% confidence interval [CI] 1110-1381) readmissions prevented annually, which compares favorably to the previous estimate of 1210 (95% CI 990-1430), representing further evidence that these estimates are well calibrated.

Next, computing the impacts of the CATE-based strategies, we find that all three strategies are estimated to result in greater potential reductions in the absolute number of readmissions prevented (Table 2). The strategy that treats the top 10% CATEs of each ventile may prevent 1461 (95% CI 1294-1628) readmissions annually, and does so more efficiently, as implied by its NNT of 13. Moreover, the top-20% strategy requires the same total number of interventions as the existing risk-based strategy (39 648 vs 39 985), yet is estimated to double the number of annual readmissions prevented, at 2478 (95% CI 2262-2694), with an NNT of 16.

Even under the most expansive strategy, which treats the top 50% of each risk ventile and requires 250% of the total interventions compared to the risk-based strategy, also represents an improvement in the NNT (23 vs 33). This strategy is estimated prevent 4458 (95% CI 3925-4990) readmissions annually, nearly four times as many as the existing risk-based strategy. Finally, we also note that while there appears to exist a tradeoff in terms of absolute impact and efficiency, all CATE-based strategies substantially improved upon the risk-based targeting strategy in terms of the NNT.

4 | DISCUSSION

Here, we have shown the feasibility of estimating individual treatment effects for a comprehensive readmissions prevention intervention using data on over 1.5 million hospitalizations, representing an example of an "impactibility" model.[20] Even though our analysis used observational data, we found that these individual estimates were well calibrated, in that none of the individual estimates were greater than the predicted risk. Moreover, these estimates, when used to compute the impact of the risk-based targeting policy, substantially agreed with a separate estimate computed via a difference-in-differences analysis.[40] Notably, our results suggest that strategies targeting similar population health management and quality improvement (QI) interventions based on these individual effects may lead to far greater aggregate benefit compared to targeting based on risk. Here, the difference translated to nearly as many as four times the number of readmissions prevented annually over the current risk-based approach.

To the best of our knowledge, this work is the first to apply causal machine learning together to estimate the treatment effect heterogeneity of a population health management intervention, and as such, may represent the first example of an end-to-end "impactibility" model.[20] Our analysis also found both qualitative and quantitative evidence for treatment effect heterogeneity, particularly across levels of predicted risk: the Transitions Program intervention appeared to be less effective as predicted risk increased. The extent of this mismatch between treatment effect and predicted risk appeared substantial and may have implications for the design of readmission reduction initiatives and related population health management programs. Notably, our finding of a risk-treatment effect mismatch is in line with some prior work in the readmissions prevention literature.[7,19,22] Further afield, a study of treatment effect heterogeneity among patients undergoing antihypertensive therapy, using a similar causal machine learning approach (the X-learner[41]), also found a risk-treatment effect mismatch in that setting.[42]

A notable finding is that some patients had a predicted CATE greater than zero, indicating that the Transitions Program intervention may have encouraged them to return to the hospital. In other settings, a positive treatment effect would often be interpreted as harm, suggesting that the intervention should be withheld from these patients. However, we argue that our finding does not readily

TABLE 2 Estimates of overall impacts of risk-based and CATE-based targeting strategies

Treatment strategy	Annual readmissions prevented, *n*	Total interventions, *n*	NNT
Risk-based targeting			
Target to ≥25% (DiD estimate)	1210 (990-1430)	39 985	33
Target to ≥25% (CF estimate)	1246 (1110-1381)	39 985	33
CATE-based targeting			
Targeting top 10%	1461 (1294-1628)	18 993	13
Targeting top 20%	2478 (2262-2694)	39 648	16
Targeting top 50%	4458 (3925-4990)	102 534	23

Note: These impacts are expressed in terms of the annual numbers of readmissions prevented as well as the numbers needed to treat (NNTs) under each targeting strategy, based on the estimates for index admissions in 2018. The first quantity—the difference-in-differences (DiD) estimate—is based on the results of a prior study.[40] All quantities are rounded to the nearest integer. Parentheses represent 95% confidence intervals.

Abbreviations: CATE, conditional average treatment effect; CF, causal forest; DiD, difference-in-differences.

admit such an interpretation. To see why, we note that this subgroup of patients with a positive CATE appeared to be those who were more acutely ill at discharge, as evidenced by their higher LAPS2DC scores (Figure 3). Hence, an alternative interpretation of these positive CATEs is that they represent readmissions which may have been necessary, and which perhaps may have been facilitated by aspects of the Transitions Program intervention, including instructions to patients outlining the circumstances (eg, new or worsening symptoms) under which they should seek further care. This finding holds particular relevance given increasing concern that readmission prevention programs, in responding to regulatory incentives, may be reducing 30-day hospitalization rates at the expense of increased short- and long-run mortality.[5,43,44]

In particular, this finding also suggests that the CATE estimates may be insufficient to capture the full impact of the Transitions Program intervention on patient outcomes, meaning that the estimated effect of the intervention on readmission alone may not represent a sufficient basis for future targeting strategies. It is plausible that intervening in a patient with a positive estimated effect may be warranted if the readmission would have a positive effect on other outcomes, despite the current emphasis of value-based purchasing programs on penalizing excess 30-day readmissions. For example, in fiscal year 2016, the maximum penalty for excess 30-day mortality was 0.2% of a hospital's diagnosis-related group (DRG) payments under the Hospital Value-Based Purchasing program, while the maximum penalty for excess 30-day readmission was 3.0% of DRG payments under the HRRP.[45] Hence, a more holistic targeting strategy would incorporate estimates of the intervention's effect on short- and long-run mortality and other outcomes. Selecting patients for treatment could then be formulated as an optimization problem that attempts to balance regulatory incentives, organizational priorities, patient welfare, and resource constraints.

Our approach could be used to re-target other population health management and QI interventions currently based on risk assessment tools. Patients expected to benefit could be prioritized to receive such an intervention, while patients unlikely to benefit could instead receive more targeted care that better meets their needs, including specific subspecialty care, and in some cases, palliative care. However, these individual estimates were derived from observational data, and not from data generated via a randomized experiment—the latter which represents the ideal substrate for estimating treatment effects, insofar as randomization is able to mitigate the effects of confounding.[46] Furthermore, our approach requires interventional data, unlike those used to develop traditional risk assessment tools, which use historical data. Hence, to implement our approach may require rethinking how predictive algorithm-driven interventions (or "prediction-action dyads"[47]) are deployed within health systems, particularly in relation to existing digital infrastructure and institutional oversight processes.

One starting point for doing so is to first deploy a new predictive algorithm-driven intervention as part of a two-arm randomized trial which compares that intervention to usual care. This trial

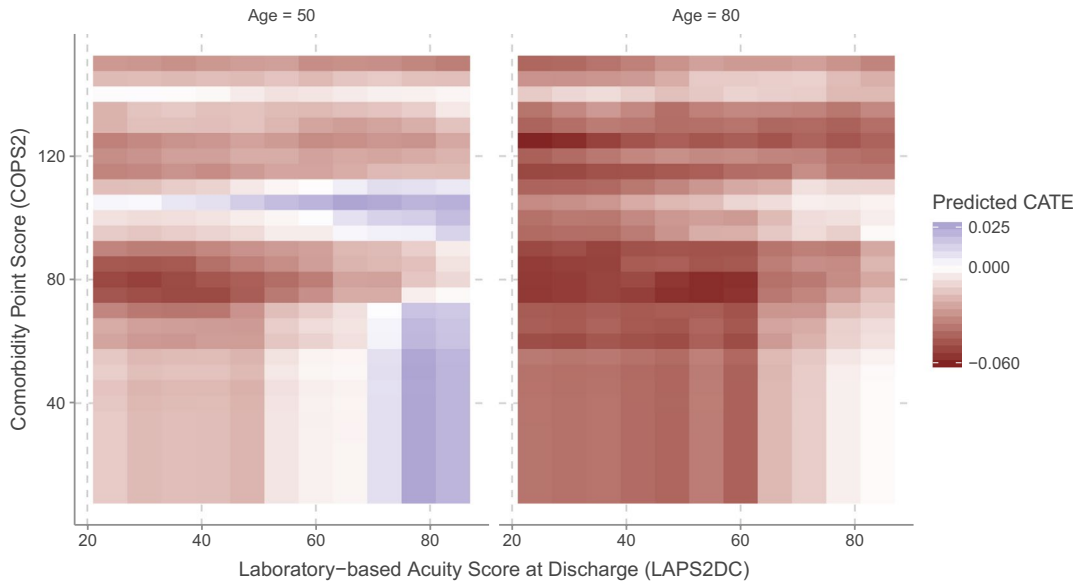

FIGURE 3 Visualization of the estimated conditional average treatment effect function. This figure presents the estimated CATE function as it varies in the dimensions of Laboratory-based Acuity Score at discharge (LAPS2DC) and Comorbidity Point Score (COPS2), for a patient with chronic heart failure at ages 50 and 80. Here, we vary LAPS2DC and COPS2 while holding all other continuous covariates at their median values, except for age, which we set to 50 and 80. Categorical covariates were held at their mode, except for the supergroup, which we set to chronic heart failure (CHF). We plot the CATE function from the 10th to 90th percentiles of LAPS2DC and from the 0th to 95th percentiles of COPS2. This is akin to evaluating the CATE function for a set of pseudo-patients with CHF having these values of COPS2 and LAPS2DC. In this region, the estimated CATE ranged from −0.060 to 0.025 (−6.0% to 2.5%), meaning that the estimated absolute risk reduction of the Transitions Program intervention was as large as −6% for some patients, while for others, their readmission risk was increased by as much as 2.5% [Color figure can be viewed at wileyonlinelibrary.com]

represents a pilot phase, generating data that are used to derive an impactibility model. Following this pilot phase, two options are possible: (1) based on this impactibility model, the intervention could be re-targeted to the patients most expected to benefit; or (2) alternatively, another two-arm randomized trial comparing risk-based to impactibility-based targeting could be carried out. In the latter option, patients would be randomized to either of the risk or impactibility arms, and based on their covariates would either receive or not receive the intervention according to their risk or benefit estimate. Using the results of this second trial, whichever targeting approach proved more effective could then be operationalized.

This proposal represents a shift from how deployments of predictive algorithm-driven interventions are usually carried out in health systems. New overnight processes[48] and digital infrastructure would be required in order to realize the full potential of these approaches. Many such interventions, if deemed to create only minimal risk, often fall under the umbrella of routine quality improvement (QI) studies, which exempts them from ongoing independent oversight.[49] However, incorporating randomization may shift these implementations from the category of routine QI to nonroutine QI or research. These latter two categories of studies often require independent oversight by, for example, an institutional review board (IRB), and may need to incorporate additional ethical considerations, for example, requiring informed consent.

This study has several limitations. First, as it is observational in nature, our analysis necessarily relies on certain assumptions, which, while we believe are plausible, are unverifiable. The unconfoundedness assumption that we make presumes no unmeasured confounding, and cannot be verified through inspection of the data nor via statistical tests. Second, these estimates of benefit must be computed with data that include the intervention, and not with historical data, as with risk assessment tools. Third, although we incorporated it into our analysis, the HCUPSGDC variable is not always available at discharge. For a retrospective study that principally seeks to characterize heterogeneous treatment effects, this does not constitute a limitation. However, prospective applications of this model would have to take this into account. Finally, it is possible that patients deprioritized under an impactibility modeling approach, but who might still be at high risk for the outcome, may still require alternative interventions better tailored to their needs.[20]

5 | CONCLUSION

Causal machine learning can be used to identify preventable hospital readmissions, if the requisite interventional data are available. Moreover, our results point to a mismatch between readmission risk and treatment effect, which is consistent with suggestions in prior work. Here, the extent of this mismatch was substantial, suggesting that many preventable readmissions may be being "left on the table" with current approaches based on risk assessment. Our proposed framework is also generalizable to the study of a range of population health management and quality improvement interventions currently driven by risk prediction models.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors are immensely grateful to Colleen Plimier of the Division of Research, Kaiser Permanente Northern California, for assistance with data preparation, as well as to Dr. Tracy Lieu, also of the Division of Research, for reviewing the manuscript. In addition, the authors wish to thank Minh Nguyen, Stephen Pfohl, Rachael Aikens, and Scotty Fleming for valuable feedback on earlier versions of this work. Mr. Marafino was supported by a predoctoral fellowship from the National Library of Medicine of the National Institutes of Health under Award Number T15LM007033, as well as by funding from a Stanford School of Medicine Dean's Fellowship. Dr. Baiocchi was also supported by grant KHS022192A from the Agency for Healthcare Research and Quality. Dr. Vincent Liu was also supported by NIH grant R35GM128672 from the National Institute of General Medical Sciences. Portions of this work were also funded by The Permanente Medical Group, Inc., and Kaiser Foundation Hospitals, Inc. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors have no conflicts of interest to disclose. The funders played no role in the study design, data collection, analysis, reporting of the data, writing of the report, nor the decision to submit the article for publication.

ORCID

Ben J. Marafino https://orcid.org/0000-0003-2385-7376
Alejandro Schuler https://orcid.org/0000-0003-4853-6130
Vincent X. Liu https://orcid.org/0000-0001-6899-9998
Gabriel J. Escobar https://orcid.org/0000-0003-2540-3327
Mike Baiocchi https://orcid.org/0000-0002-7571-5268

REFERENCES

1. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare Fee-for-Service Program. *N Engl J Med*. 2009;360:1418-1428.
2. Hines AL, Barrett ML, Jiang HJ, et al.Conditions With the Largest Number of Adult Hospital Readmissions by Payer. Statistical Brief #172. Healthcare Cost and Utilization Project (HCUP). May 2016. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/reports/statbriefs/sb172-Conditions-Readmissions-Payer.jsp.
3. Leppin AL, Gionfriddo MR, Kessler M, et al. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. *JAMA Int Med*. 2014;174:1095-1107.
4. Hansen LO, Young RS, Hinami K, Leung A, Williams MV. Interventions to reduce 30-day rehospitalization: a systematic review. *Ann Intern Med*. 2011;155:520-528.
5. Wadhera RK, Joynt-Maddox KE, Wasfy JH, et al. Association of the Hospital Readmissions Reduction Program with mortality among medicare beneficiaries hospitalized for heart failure, acute myocardial infarction, and pneumonia. *JAMA*. 2018;320:2542.
6. Kansagara D, Chiovaro JC, Kagen D, et al. So many options, where do we start? An overview of the care transitions literature. *J Hosp Med*. 2016;11(3):221-230.

7. Finkelstein A, Zhou A, Taubman S, et al. Health care hotspotting—a randomized, controlled trial. *N Engl J Med*. 2020;382:152-162.
8. Bates DW, Suchi Saria, Ohno-Machado L, et al. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*. 2014;33:1123-1131.
9. Berkowitz SA, Parashuram S, Rowan K, et al. Association of a Care Coordination Model with health care costs and utilization: the Johns Hopkins Community Health Partnership (J-CHiP). *JAMA Network Open*. 2018;1:e184273.
10. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA*. 2011;306:1688-1698.
11. Bayati M, Braverman M, Gillam M, et al. Data-driven decisions for reducing readmissions for heart failure: general methodology and case study. *PLoS ONE*. 2014;9:e109264.
12. Escobar GJ, Ragins A, Scheirer P, et al. Nonelective rehospitalizations and postdischarge mortality. *Medical Care*. 2015;53:916-923.
13. Billings J, Dixon J, Mijanovich T, et al. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *Br Med J*. 2006;333:327-330.
14. Fihn SD, Francis J, Clancy C, et al. Insights from advanced analytics at the Veterans Health Administration. *Health Affairs*. 2014;33:1203-1211.
15. Health IT Analytics. Using risk scores, stratification for population health management. 2016. https://healthitanalytics.com/features/using-risk-scores-stratification-for-population-health-management.
16. Health IT Analytics. Top 4 Big data analytics strategies to reduce hospital readmissions. 2018. https://healthitanalytics.com/news/top-4-big-data-analytics-strategies-to-reduce-hospital-readmissions.
17. Ascarza E. Retention futility: targeting high-risk customers might be ineffective. *Journal of Marketing Research*. 2018;55:80-98.
18. Athey S. Beyond prediction: using big data for policy problems. *Science*. 2017;355:483-485.
19. Lindquist LA, Baker DW. Understanding preventable hospital readmissions: Masqueraders, markers, and true causal factors. *J Hosp Med*. 2011;6:51-53.
20. Lewis GH. "Impactibility models": identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs. *Milbank Quarter*. 2010;88:240-255.
21. Freund T, Mahler C, Erler A, et al. Identification of patients likely to benefit from care management programs. *Am J Managed Care*. 2011;17:345-352.
22. Steventon A, Billings J. Preventing hospital readmissions: the importance of considering 'impactibility', not just predicted risk. *BMJ Qual Safety*. 2017;26:782-785.
23. Flaks-Manov N, Srulovici E, Yahalom R, et al. Preventing Hospital Readmissions: Healthcare Providers' Perspectives on "Impactibility" Beyond EHR 30-Day Readmission Risk Prediction. *J Gen Int Med*. 2020;35:1484-1489.
24. Rich MW, Vinson JM, Sperry JC, et al. Prevention of readmission in elderly patients with congestive heart failure: results of a prospective, randomized pilot study. *J Gen Int Med*. 1993;8:585-590.
25. Roland M, Abel G. Reducing emergency admissions: are we on the right track? *BMJ*. 2012;345:e6017.
26. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 1985;14:32-38.
27. Chiolero A, Paradis G, Paccaud F. The pseudo-high-risk prevention strategy. *Int J Epidemiol*. 2015;44:1469-1473.
28. McWilliams JM, Schwartz AL. Focusing on high-cost patients—the key to addressing high costs? *N Engl J Med*. 2017;376:807-809.
29. Goldfield NI, McCullough EC, Hughes JS, et al. Identifying potentially preventable readmissions. *Health Care Finan Rev*. 2008;30:75-91.
30. Auerbach AD, Kripalani S, Vasilevskis EE, et al. Preventability and causes of readmissions in a national cohort of general medicine patients. *JAMA Int Med*. 2016;176:484-493.
31. Escobar GJ, Plimier C, Greene JD, et al. Multiyear rehospitalization rates and hospital outcomes in an integrated health care system. *JAMA Network Open*. 2019;2:e1916769.
32. Agency for Healthcare Research and Quality. AHRQ quality indicators—guide to prevention quality indicators: hospital admission for ambulatory care sensitive conditions. 2002. Agency for Healthcare Research and Quality, Rockville, MD. https://www.ahrq.gov/downloads/pub/ahrqqi/pqiguide.pdf
33. Rubin DB. Causal inference using potential outcomes. *J Am Stat Assoc*. 2005;100:322-331.
34. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc*. 2018;1459:1-15.
35. Athey S, Tibshirani J, Wager S. Generalized Random Forests. *Ann Stat*. 2019;47:1148-1178.
36. Athey S, Wager S. Estimating treatment effects with causal forests: an application. arXiv. 2019. https://arxiv.org/pdf/1902.07409.pdf
37. Shalit U. Can we learn individual-level treatment policies from clinical data? *Biostatistics*. 2020;21:359-362.
38. Chernozhukov V, Demirer M, Duflo E, et al. Generic machine learning inference on heterogenous treatment effects in randomized experiments. 2017. https://arxiv.org/abs/1712.04802
39. Robinson PM. Root-*N*-consistent Semiparametric Regression. *Econometrica*. 1988;56:931.
40. Marafino BJ, Escobar GJ, Liu VX, et al.A comprehensive readmissions prevention intervention enabled by predictive analytics in an integrated health system. 2020. To appear.
41. Kunzel SR, Sekhon JS, Bickel PJ, et al. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc Natl Acad Sci USA*. 2019;116:4156-4165.
42. Duan T, Rajpurkar P, Laird D, et al. Clinical value of predicting individual treatment effects for intensive blood pressure therapy. *Circulat Cardiovasc Quality Outcomes*. 2019;12:5010.
43. Fonarow GC, Konstam MA, Yancy CW. The Hospital Readmission Reduction Program is associated with fewer readmissions, more deaths: time to reconsider. *J Am Coll Cardiol*. 2017;70:1931-1934.
44. Gupta A, Fonarow GC. The Hospital Readmissions Reduction Program-learning from failure of a healthcare policy. *Eur J Heart Fail*. 2018;20:1169-1174.
45. Abdul-Aziz AA, Hayward RA, Aaronson KD, et al. Association between Medicare hospital readmission penalties and 30-day combined excess readmission and mortality. *JAMA Cardiol*. 2017;2:200-203.
46. Kent DM, Steyerberg E, Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. 2018;363:k4245.
47. Liu VX, Bates DW, Wiens J, et al. The number needed to benefit: estimating the value of predictive analytics in healthcare. *J Am Med Inform Assoc*. 2019;13:1655-1659.
48. Faden RR, Kass NE, Goodman SN, et al. An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics. *Hastings Center Rep*. 2013;43:S16-S27.
49. Finkelstein JA, Brickman AL, Capron A, et al. Oversight on the borderline: quality improvement and pragmatic research. *Clinical Trials*. 2015;12:457-466.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Marafino BJ, Schuler A, Liu VX, Escobar GJ, Baiocchi M. Predicting preventable hospital readmissions with causal machine learning. *Health Serv Res*. 2020;55:993–1002. https://doi.org/10.1111/1475-6773.13586