

Selecting Optimal Subgroups for Treatment Using Many Covariates

Tyler J. VanderWeele,^a Alex R. Luedtke,^b Mark J. van der Laan,^c and Ronald C. Kessler^d

Abstract: We consider the problem of selecting the optimal subgroup to treat when data on covariates are available from a randomized trial or observational study. We distinguish between four different settings including: (1) treatment selection when resources are constrained; (2) treatment selection when resources are not constrained; (3) treatment selection in the presence of side effects and costs; and (4) treatment selection to maximize effect heterogeneity. We show that, in each of these cases, the optimal treatment selection rule involves treating those for whom the predicted mean difference in outcomes comparing those with versus without treatment, conditional on covariates, exceeds a certain threshold. The threshold varies across these four scenarios, but the form of the optimal treatment selection rule does not. The results suggest a move away from the traditional subgroup analysis for personalized medicine. New randomized trial designs are proposed so as to implement and make use of optimal treatment selection rules in healthcare practice.

Keywords: Effect modification; Interaction; Optimal treatment selection; Precision medicine; Personalized treatment; Randomized trial; Subgroup

(*Epidemiology* 2019;30: 334–341)

Biomedical researchers and social scientists are often interested in identifying the subgroups that would benefit most from a particular treatment or intervention. In randomized trials, subgroup analyses are often used to compare the effect of

treatment across subgroups defined by various pretreatment covariates.^{1–6} Such analyses can help give insight into whether a treatment might be more effective for men versus women, or for younger versus older persons, or for any other characteristic or variable defined before the receipt of treatment. These types of analyses are relevant if the effect of treatment might vary across individuals in a population, a phenomenon often referred to as “effect heterogeneity.” Such analyses can be useful in deciding who to treat, or who to treat first, if resources are limited. They can also be useful when deciding which of two treatments to give to whom.

While well-established methodology has been used for decades to carry out such subgroup analyses across strata defined by a single covariate,^{1,7–11} in actual practice it would be more desirable to make use of data on numerous covariates. Viewed from the individual perspective, we are interested in knowing how to best choose the appropriate treatment for an individual with a particular set of characteristics. This task is sometimes now described as “personalized medicine” or “precision medicine.” It is the optimal selection of treatment for the individual.^{12–14} However, viewed from the perspective of a population, if we optimize the treatment for each individual, we are also optimizing the outcomes for the population and are thus interested in which subgroups to give which treatment to maximize the outcomes within a population of interest, possibly subject to resource constraints.

To make progress with multiple covariates for this task, it is not uncommon in the biomedical or social sciences to form a “prognostic score.”^{4,15–18} In a randomized trial with treatment and control, this prognostic score is defined as the predicted value of the outcome, conditional on an individual’s covariates, if that person were not given treatment. The prognostic score is often obtained by first fitting a regression model of the outcome on the pretreatment covariates among the control arm of the randomized trial. Using the estimates of the regression parameters of this model, one can then obtain predicted outcomes under the absence of treatment for each individual in the study to give the prognostic score. The prognostic score itself is then typically taken as the variable by which subgroups are formed. An analyst might, for example, subsequently analyze the data within tertiles, quartiles, or quintiles of the prognostic score. If those with low prognostic scores would benefit most from treatment, then this might be the group for which it would be the best to target treatment. This approach is used with some

Editor’s Note: A related commentary appears on p. 342.

Submitted February 26, 2018; accepted January 30, 2019.

From the ^aDepartments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; ^bFred Hutchinson Cancer Research Center, Seattle, WA; ^cDepartment of Biostatistics, University of California, Berkeley, CA; and ^dDepartment of Health Care Policy, Harvard Medical School, Boston, MA.

Supported by National Institutes of Health grant CA222147.

Disclosure: The authors report no conflicts of interest.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Tyler J. VanderWeele, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115. E-mail: tvanderw@hsph.harvard.edu.

Copyright © 2019 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 1044-3983/19/3003-0334

DOI: 10.1097/EDE.0000000000000991

frequency in the biomedical and social sciences.^{18–22} It is sometimes also referred to as “risk stratification”¹⁶ or “endogenous stratification.”¹⁸ While such procedures theoretically are effective with very large sample sizes, recent evidence suggests that in most practical settings, even with thousands of study participants,¹⁸ biases from this sort of approach result from overfitting if the same data are used to form the prognostic score and to run the subgroup analyses.^{18,23,24} Various techniques, using cross-validation, have been proposed to address these biases.¹⁸

However, a more fundamental problem with the approach is that even if such biases were absent, using the prognostic score or individual covariate subgroup analysis does not in fact identify the optimal treatment allocation rule. There are better ways to use the covariate data available to optimize an individual’s outcome and the mean outcomes for the population. A growing literature has begun to explore statistical approaches for more effective treatment selection rules.^{25–29} In fact, the optimal rule depends subtly on precisely what question the analysis is intended to address.

In this article, we will present four settings in which optimal subgroup selection is of interest. We will describe these settings and the optimal treatment rule in each. We will discuss how the approaches in this article relate to what is typically done in practice and how might be best to proceed in the subsequent research when selecting optimal subgroups for treatment is of interest. New randomized trial designs are further proposed so as to implement and make use of optimal treatment selection rules in practice. The focus of this article is conceptual. Our goal here is to more clearly consider the types of questions that arise with subgroup selection and to relate that to how subgroups are to be optimally formed. We then compare this with what is done in practice. The focus of the article will not be statistical methods. Statistical methods are available to carry out some of this work and are described elsewhere,^{25–29} and one approach developed by Luedtke and van der Laan^{27–29} is summarized in the eAppendix, <http://links.lww.com/EDE/B466>, but our focus here is on concepts and how we ought to think about subgroup selection within epidemiology and within the biomedical and social sciences more generally.

Notation

We will let A denote a treatment or intervention under study. We will assume that receipt of treatment has been randomized with probability 1/2, but we will comment later in the article on how the methodology described here is also potentially applicable to observational studies. We will let Y denote an outcome of interest. Finally, we will let C denote a set of pretreatment covariates that are available for each individual in the study. We will let Y_1 denote the potential outcome³⁰ that would have occurred for each individual if they had received treatment, and we will let Y_0 denote the potential outcome that would have occurred under control. We only get to observe one of Y_1 and Y_0 : we observe Y_1 for those who

actually received treatment and Y_0 for those who were actually in the control arm. We do not in general know the potential outcome if an individual had been in the other arm of the trial.

In what follows, the task of treatment selection will essentially be to partition the population into two groups, which we will call “T” and “S,” those receiving the treatment, and those not receiving the treatment, respectively. The goal will be, in each setting, to decide on how to partition the population into those who do versus do not receive treatment to maximize mean outcomes. We will refer to this partition of individuals who do and do not receive treatment as the optimal treatment rule.

We will, for simplicity, assume here that treatment A is binary with 1 denoting treatment and 0 denoting control. However, the ideas that are developed below are also applicable if we are comparing two different treatments so that $A = 1$ denotes one treatment and $A = 0$ denotes another. Although we will generally use of “treatment” and “control,” the same methods and ideas described below are applicable also in the setting of comparing two treatments with “selecting who gets treatment” simply interpreted as “selecting who gets the first treatment” and “control” interpreted as “those receiving the second treatment.” Later in the paper, we will also comment on how the ideas potentially extend to settings when more than two treatments are being considered.

In what follows, we will provide an overview of the relevant concepts and methods. We will state results that are precise under some technical conditions. More formal statements and proofs are given in the eAppendix, <http://links.lww.com/EDE/B466>.^{27–29}

Four Questions Relevant to Optimal Subgroup Selection

We will consider four settings that may be of interest in selecting optimal subgroups for treatment. Stated intuitively, these settings are as follows:

1. Who do we treat if resources are limited so that we can only treat $q\%$ of the population?
2. Who do we treat if resources are not limited so that we could potentially treat everyone and are simply deciding who would benefit from treatment?
3. Who do we treat if resources are not limited, but are subject to costs or side effects?
4. How do we select subgroups to maximize the effect heterogeneity across subgroups?

We will address each question in turn.

Setting 1. Subgroup Selection Under Resource Constraints

First, let us suppose that owing to some form of resource constraints (e.g., costs, doses available, etc.), we are only able to treat at most $q\%$ of the population. We have data from a randomized trial of treatment A where we have collected outcome Y and pretreatment covariates C . We want to use the

covariates C , and the outcome data from our randomized trial to determine a treatment rule to partition the population into those that we should treat so as to maximize the expected outcome for the population, subject to the constraint that we can only treat $q\%$ of the population. Once we decide on these two sets, T , the treated, and S , the untreated, then the expected outcome for the population under this treatment rule is:

$$\frac{q}{100}E[Y | A = 1, T] + (1 - \frac{q}{100})E[Y | A = 0, S]$$

In other words, for $q\%$ of the population, we get the average outcome under treatment for the subgroup T that we selected for treatment, and for $(100 - q)\%$ of the population, we get the average outcome under control for the subgroup S that we selected not to receive treatment.

It is shown in the eAppendix, <http://links.lww.com/EDE/B466>, that if we knew the potential outcomes, Y_1 and Y_0 , for each individual in the population, then the optimal treatment rule to maximize the expected outcome for the population would simply be to treat those for whom $\{Y_1 - Y_0 > k\}$, where k is determined so that exactly $q\%$ are treated. In other words, if we knew the potential outcomes for each individual, so that we knew the actual effect, $Y_1 - Y_0$, of treatment for each individual, we would simply treat the $q\%$ for which the effect of treatment itself was the largest. In actual fact, however, we do not know both potential outcomes for every individual in the population. We only have our randomized trial data, our outcomes Y , and our covariates C . Therefore, we want to use C to partition individuals into those who we do or do not treat to maximize outcomes. It is again shown in the eAppendix, <http://links.lww.com/EDE/B466>, that, to maximize outcomes, using covariates C , the optimal treatment rule is to treat those with covariate values c such that

$$\{E[Y | A = 1, C = c] - E[Y | A = 0, C = c] > k\}$$

where the cutoff k is again determined, so that exactly $q\%$ is treated. In other words, the optimal treatment rule is to treat the $q\%$ with the highest expected treatment effect conditional on their covariates. The expected treatment effect for each individual conditional on their covariates is something that can be estimated from the data in a randomized trial, and thus, this treatment rule can be implemented in practice. We could, for example, fit regression models for the expected outcome under treatment $E[Y | A = 1, C = c]$ and under control $E[Y | A = 0, C = c]$, (or, more directly, their difference) conditional on covariates, to obtain estimates. However, again, it can be shown that the best we can do in terms of maximizing outcomes for the population using just the covariates C is to treat those with the highest expected treatment effect, $E[Y | A = 1, C = c] - E[Y | A = 0, C = c]$, conditional on their covariates. With this treatment rule, the expected outcome for the population is again then $\frac{q}{100}E[Y | A = 1, T] + (1 - \frac{q}{100})E[Y | A = 0, S]$.

The expected outcome under the treatment rule will not be as high as we could have obtained had we known both

potential outcomes for all individuals, but again this is the best we can do with the measured covariates C . We could compare the expected outcome under the treatment rule with what we would obtain if we simply randomly selected $q\%$ of the population for treatment, in which case we would have an expected outcome of:

$$\frac{q}{100}E[Y | A = 1] + (1 - \frac{q}{100})E[Y | A = 0]$$

How much better we do under the treatment rule using the covariates C will depend in part on how predictive are the measured covariates with respect to the association between treatment and the outcome of interest, and also how well we statistically model the expected outcomes $E[Y | A = 1, C = c]$ and $E[Y | A = 0, C = c]$, or, more directly, their difference. We could compare the expected population outcomes in under different estimates of the optimal treatment rule using different modeling techniques. Intuitively, how well we improve on the outcomes by selecting subgroups for treatment using covariates C , instead of randomly allocating treatment, will effectively depend on how well we can use the covariates C and statistical modeling to predict the potential outcomes, i.e., how well we estimate the true $E[Y | A = 1, C] - E[Y | A = 0, C]$.

Setting 2. Subgroup Selection Under Unconstrained Resources

We will now turn to a different setting in which resources are not constrained, so that we could potentially treat anyone who might benefit from the treatment. Once again our objective is to determine the treatment rule that partitions individuals into two sets: T , those who do receive treatment, and S , those who do not; so as to maximize the average outcome for the population, which is then:

$$E[Y | A = 1, T]P(T) + E[Y | A = 0, S]P(S)$$

It is shown in the eAppendix, <http://links.lww.com/EDE/B466>, that if we knew the potential outcomes, Y_1 and Y_0 , for each individual in the population, then the optimal treatment rule to maximize the expected outcome for the population would simply be to treat those for whom $\{Y_1 - Y_0 > 0\}$. In other words, if we knew the potential outcomes for each individual, we would simply treat those for whom the effect of treatment itself was positive. This is, of course, relatively intuitive. In actual fact, we do not, of course, know both potential outcomes for every individual; we only have our covariates C . With covariates C , it shown in the eAppendix, <http://links.lww.com/EDE/B466>, that to maximize outcomes, using covariates C , the optimal treatment rule is to treat those with covariate values c such that

$$\{E[Y | A = 1, C = c] - E[Y | A = 0, C = c] > 0\}.$$

In other words, we treat those who have, conditional on their covariates, a positive expected treatment effect. We can again estimate this from the data from our randomized trial.

Under this treatment rule, the expected outcome will simply be $E[Y|A = 1, T]P(T) + E[Y|A = 0, S]P(S)$. We could compare this expected outcome under the optimal treatment rule with the expected outcome if we treated everyone in the population, $E[Y|A = 1]$, or if we treated no one, $E[Y|A = 0]$. Once again, how well we could optimize outcomes would depend on how predictive were the covariates C with respect to the association between treatment and outcome.

An interesting feature of this second setting of unconstrained optimal treatment selection is that the tasks of individual decision-making and maximizing population outcomes in fact coincide. The approach to maximize population outcomes is simply to assign treatment to anyone who would benefit from it. The perspectives of the individual and the policymaker coincide. This was not the case in the first setting, wherein an individual might have a positive expected treatment effect and therefore, from an individual perspective, have expected benefit from treatment, whereas a policymaker, to maximize population outcomes, might choose not to treat that individual because others have higher expected treatment effects and resources are limited.

Setting 3. Subgroup Selection Under Costs and Side Effects

Now let us turn to a setting in which resources are not constrained so that we could once again, in principle, treat everyone, but now suppose the treatment itself has a cost that we want to take into account, and/or has side effects that we want to weigh against the potentially beneficial effects on our outcome of interest Y . Because of costs or side effects, we might, for example, only want to treat those with treatment effects larger than some level δ . Or more generally, that level might depend on a person's covariates c so that we only want to treat those with the treatment effect greater than some level $\delta(c)$. The optimal rule (see eAppendix, <http://links.lww.com/EDE/B466>) if we knew both potential outcomes for all individuals would then simply be to treat those with $\{Y_1 - Y_0 > \delta(c)\}$ and the optimal rule with the actual trial data and measured covariates C would be to treat those with $\{E[Y|A = 1, C = c] - E[Y|A = 0, C = c] > \delta(c)\}$. Once again, how well we could optimize outcomes would depend on how predictive was the covariate C with the association between treatment and outcome.

Setting 4. Maximizing Effect Heterogeneity

When one reads through the subgroup analyses of many randomized trials, in which subgroup analyses are undertaken one covariate at a time, it often seems that the goal is to find a covariate, often dichotomous or dichotomized, such that the effect heterogeneity across subgroups defined by the covariate is as large as possible. When the effect estimate in one subgroup is much larger than that of the other, then the subgroup analysis is considered a success and that covariate defining the subgroups is subsequently considered important. In fact, we could carry out a similar exercise using data on multiple covariates.

In this case, we would want to use covariates C to partition the population into two subsets, T and S , such that the effect in the subgroup T , $E[Y|A = 1, T] - E[Y|A = 0, T]$, was much larger than the effect in subgroup S , $E[Y|A = 1, S] - E[Y|A = 0, S]$. In other words, we would want to maximize effect heterogeneity by maximizing the difference between the effects in these two subgroups:

$$E[Y|A = 1, T] - E[Y|A = 0, T] - \{E[Y|A = 1, S] - E[Y|A = 0, S]\}$$

This is in some sense a generalization of what seems to be the traditional subgroup task but extended to multiple covariates simultaneously. It is shown in the eAppendix, <http://links.lww.com/EDE/B466>, that once again the solution to this maximization takes the form of selecting T to be those with an actual treatment effect, $Y_1 - Y_0$ (if the potential outcomes were known), or expected treatment effect conditional covariates C , $E[Y|A = 1, C = c] - E[Y|A = 0, C = c]$, above some threshold k' , where k' can be determined numerically as described in the eAppendix, <http://links.lww.com/EDE/B466>. But once again, it is the expected conditional treatment effect, $E[Y|A = 1, C = c] - E[Y|A = 0, C = c]$, which is utilized in the criterion by which treatment decisions are to be made in this setting as well. Note, however, that although this treatment rule maximizes effect heterogeneity, the average outcome under this treatment rule will generally be worse than that selected by the treatment rule that maximizes the outcome itself as in Setting 2: Subgroup Selection Under Unconstrained Resources. It is thus not clear that this treatment rule that maximizes effect heterogeneity is of particular use in decision-making, unlike those in Settings 1, 2, and 3. We will return to this point in the discussion.

Extensions to Observational Studies

Our discussion thus far has been within the context of a randomized trial. However, as discussed further in the eAppendix, <http://links.lww.com/EDE/B466>, all of the discussion above pertains also to optimal subgroup selection and treatment decisions from data arising from an observational study as well, provided that the covariates C suffice to control for confounding of the effect of treatment A on outcome Y , though the formulae for the optimized outcome need to be modified (see eAppendix, <http://links.lww.com/EDE/B466>). With data from observational studies, an additional context that may be of interest is if, in data from the study, there are available covariates C that suffice to control for confounding for the effect of treatment A on outcomes Y , but if, when treatment decisions are made subsequently, only data on some subset W of the covariates C will be available. Methodology for this setting has been developed and is described elsewhere.^{27–29,31} Further discussion of statistical approaches is given in the Appendix and eAppendix, <http://links.lww.com/EDE/B466>.

IMPLICATIONS FOR CURRENT PRACTICES

We have shown that under a wide range of different goals and settings, including making treatment decisions with or without resource constraints, and with or without side effects, or even when trying to maximize effect heterogeneity, the correct approach to find the optimal treatment rule is to estimate expected treatment effects for each individual conditional on the covariates. In each of the settings described above, the optimal treatment rule involved treating those above some threshold of the conditional expected treatment effect. The threshold differed according to whether there were or were not resource constraints, or whether there were or were not costs or side effects, or whether we wanted to maximize effect heterogeneity, but the form of the treatment rule did not vary across these contexts. In each case, the form of the optimal treatment rule was simply to treat those with conditional expected treatment effects above a specific threshold. The results are summarized in the Table. This has a number of important implications for the actual practice of subgroup analysis, treatment selection, precision medicine, and the modeling of interactions.

Subgroup Analysis

One fundamental insight from our discussion above is that for treatment selection and decisions, our discussion suggests a need to move away from subgroup analyses conducted one covariate at a time. The problems with this approach are numerous. First, subgroups may come into conflict: if subgroup analyses indicate that treatment A is better for women and treatment B is better for men, and also indicate that A is better for older and B better for younger persons, and we want to make treatment decisions for a younger woman, the subgroup analyses conflict. Second, the subgroup analyses often fail to answer the scientific question of interest. As they are typically carried out, they tend to be aimed at maximizing effect heterogeneity, whereas what is actually of interest is maximizing population outcomes or individual treatment decision-making. The optimal treatment rule for maximizing population outcomes or individual treatment decision-making is not the same as for maximizing effect heterogeneity. Finally, compared with individual covariate subgroup analyses, we can in fact do better at maximizing mean outcomes by making simultaneous use of all covariates, rather than running analyses one covariate at a time. It is conceivable, of course, that the optimal treatment selection in some rare cases might involve only a single dichotomous covariate, or in some settings, a

single dichotomous covariate may constitute the decision to be made (e.g., resources are limited so we can only intervene in city 1 or city 2), but in general, the optimal decision-making rule will make fuller use of covariate data.

The need to move away from one-covariate-at-a-time approaches in optimizing population outcomes or individual treatment decision-making is relevant not just to traditional subgroup analyses when we are looking at whether the treatment effect is larger in one group versus another, but this same point is also relevant to the analysis of so-called qualitative or crossover interactions,^{32–36} in which the treatment has a positive effect in one subgroup and a harmful effect in another. The analysis of such crossover interactions is again often done one covariate at a time, but for the purposes of decision-making, it ought to be done using all available relevant covariate data. In fact, the methodology described in Setting 2. Subgroup Selection Under Unconstrained Resources is doing precisely that.

Prognostic Scores

A second important implication of the discussion in this article is that, for optimizing population outcomes or individual treatment decision-making, we should move away from the “prognostic score,” which is often employed in both the biomedical and social sciences.^{16–22} The practice of stratifying on prognostic scores in small- or medium-sized trials has numerous statistical problems with “overfitting” documented elsewhere.¹⁸ At a more fundamental level, though, it gets the objective wrong because the patients at the greatest risk of bad outcomes in the absence of treatment are not necessarily the same patients who will profit most from intervention. While stratifying the results of randomized trials using the predicted outcome under control can provide some insight into who might be considered to have greatest need for treatment (and this may itself be of interest if there are questions of equity), it is not the correct approach to optimize population outcomes or individual treatment decision-making. To optimize population outcomes or individual treatment decision-making, one stratifies, not by predicted outcome under control, but by the expected effect of treatment; that is, the difference between the predicted outcome under treatment and the predicted outcome under control, conditional on covariates. It is this stratification that gives one insight into optimal treatment decisions either with or without resource constraints.

Interaction Analysis

A third important implication, related somewhat to the first, concerns the modeling of interactions. In reading the

Table. Summary of Optimal Subgroup Selection Settings and Optimal Treatment Selection Rules

Setting	Optimal Treatment Rule	Threshold
Resource constraints (can only treat $q\%$)	$E[Y A = 1, C = c] - E[Y A = 0, C = c] > k$	k is selected, so $q\%$ are treated
Unconstrained resources	$E[Y A = 1, C = c] - E[Y A = 0, C = c] > 0$	Treat all with positive expected treatment effect
Unconstrained resource with costs or side effects	$E[Y A = 1, C = c] - E[Y A = 0, C = c] > \delta(c)$	Treat all with expected treatment effect above costs
Maximizing effect heterogeneity	$E[Y A = 1, C = c] - E[Y A = 0, C = c] > k'$	k' is determined by numerical optimization

literature, one is often left with the impression that the principal goal of interaction analysis is to determine whether, in a given statistical model, a product term involving two variables is “statistically significant” or nonzero. Methodology to detect to “interactions” or nonzero product terms has become increasingly advanced.^{37–40} However, once again, if the purpose of the analysis is optimizing population outcomes or individual decision-making, the question as to whether a specific product term in a particular statistical model may be present is, in fact, secondary. All that matters for the task of optimizing population outcomes or individual decision-making is having predictive covariates and having statistical models that give good predictions of expected outcomes conditional on those covariates. If the product terms help in a particular model, they can be included; if not, they can be omitted. In either case, though, their presence or absence is secondary in having a good predictive model so as to make optimal treatment decisions. Indeed, using models both with and without product terms and, more generally, numerous models and machine learning algorithms, to generate predicted outcomes, and possibly ensemble methods to average over, or choose among them, as discussed in the Appendix and eAppendix, <http://links.lww.com/EDE/B466>, is probably a preferable way to proceed.

It might be thought that subgroup analyses one-covariate-at-a-time or the analysis of individual product terms in statistical models may still be of interest for the purposes of understanding or explanation. While this may be true to some degree, it is important to clarify the goal of such understanding or the form of explanation that is in view.^{41–43} If what is thought to be of importance is to understand which covariates in fact are most relevant in decision-making (e.g., because it was thought undesirable to measure all of the covariates subsequently in treatment decision-making), then one could instead consider the result of optimal treatment rules on the maximized population outcome when only certain subsets of the covariates C are considered. On the other hand, if one simply wanted to assess which covariates in some sense seemed most “responsible” for the effect heterogeneity, one might instead still model the outcome with all covariates simultaneously, and then consider, for example, what a one-unit shift in any given covariate for all individuals would do in changing expected treatment effects. In linear models for the expected outcomes in each treatment arm, this would simply be the difference between the covariate coefficient in the model under treatment $E[Y|A = 1, C = c]$ and the covariate coefficient in the model under control $E[Y|A = 0, C = c]$. But the approach of considering a one-unit shift in a particular covariate across all individuals could also be employed in nonlinear models as well. Other metrics could also potentially be developed. Finally, sometimes analyses of interactions are undertaken for the purpose of understanding the joint effects of the treatment and a particular covariate, or to gain mechanistic insight.^{41–43} In this case, it may be appropriate to assess the joint effects of one covariate at a time, but in this case, if the *effect*

of the covariate is in view, then confounding control must be made for the association between that covariate and the outcome^{42–44} and what additional variables are needed to control for such confounding will vary depending on which covariate is in view. This is no longer simply a question of effect heterogeneity but of *joint effects*.^{11,42} A model that includes all of the covariates C available will not in general be adequate to provide appropriate control in addressing this type of question if the covariates themselves affect one another.

Heuristics and Multiple Treatments

Yet another argument that might be put forward for doing one-by-one subgroup analyses may involve trying to generate heuristics. A physician cannot remember the functional form of two conditional expectations but can remember that treatment A is better for women and treatment B is better for men. While such treatment heuristics can be of some value, they can, as already discussed above, come into conflict with one another. Moreover, the use of such heuristics in decision-making becomes even more complex when there are more than two potential options to choose among, which brings us to another topic of our discussion: extensions to multiple treatments.

The setting of multiple treatment options is important in general and especially so in an era of personalized or precision medicine. Full discussion of the issue is beyond the scope of the present article, but many of the points discussed above do generalize to the multiple treatments setting. Specifically, in the task of optimizing treatment decisions without resource constraints, the solution to maximizing the population outcome, which is itself identical, in this setting, to maximizing the outcome for each individual, and involves a very similar form to what has already been discussed above. The optimal treatment rule in this setting with measured covariates C simply involves obtaining the expected outcome given an individual's covariates C under each possible treatment, $E[Y|A = a, C = c]$, $a = 0, 1, 2, \dots, N$, and then assigning to each individual the treatment that gives the highest predicted outcome. Likewise, for the same reasons as those given above, in this setting, if the goal is to maximize population outcomes or individual decision-making, there is little reason to carry out one-by-one-covariate subgroup analyses or to consider which product terms in statistical models are statistically significant.

Clinical Judgment and New Randomized Trial Designs

In the clinical setting, one might also wonder about the role of expert judgment. Are there perhaps aspects of a patient's profile which are not, or even cannot be, adequately captured by a variable that we can use in a statistical model? This of course remains a possibility. Are we to abandon clinical judgment and simply rely on statistical models to make such predictions? Are we to pit clinical judgment and modeling against one another? We would like to close this article by attempting to tackle this question head on with a

compromise, to allow both clinical judgment and predictive models, by proposing a new type of study design.

A possible design, what we will refer as an Predicted Outcomes Trial, is to first use either *prior* randomized trial data, and/or observational data, with a relatively rich set of covariates *C* to build models for the expected outcomes with and without treatment. With such models, for each study participant in the Predicted Outcomes Trial, the clinician (or patient) is randomized either to receive no further information, or to receive information on the predicted outcome given their covariates under each treatment scenario. This could include outcomes under multiple treatment options. The clinician (or patient) then decides, based on the information available and their own judgments and preferences, which treatment to select. Outcomes are measured after a suitable follow-up period to determine whether the information provided by the predictive outcome models is useful in such decision-making. A trial of this sort will allow decision makers to make use of both individually oriented outcome predictions under statistical models, and also personal judgments, in making treatment decisions. It would also preserve decisionmaker autonomy, and be more likely to be palatable to clinicians, and therefore more likely also to be scalable. The trials themselves would determine the additional utility of the information provided by the predictive models. A variation that added an additional arm in which treatment always followed the predicted maximum outcome could also be used to evaluate the role of clinical judgment, whether beneficial or harmful, above and beyond reliance on predicted probabilities. We believe that such trials will be of use in determining the utility of prediction models for personalized or precision medicine in actual practical settings.

CONCLUSIONS

In summary, we believe that careful thought as to what the correct question is in individual treatment decision-making, and careful selection of the correct optimization question and statistical method corresponding to the question of interest, will result in better patient outcomes. Current practices of one-covariate-at-a-time subgroup analysis, the use of prognostic scores, and the detection of significant interactions are simply not optimal for decision-making.

REFERENCES

1. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266:93–98.
2. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355:1064–1069.
3. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21:2917–2930.
4. Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet*. 1995;345:1616–1619.
5. Lagakos SW. The challenge of subgroup analyses—reporting without distorting. *N Engl J Med*. 2006;354:1667–1669.
6. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med*. 2007;357:2189–2194.
7. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol*. 1980;112:467–470.
8. Rothman KJ. *Modern Epidemiology*. Little Brown & Co., 1986.
9. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology*. 1992;3:452–456.
10. Li R, Chambless L. Test for additive interaction in proportional hazards models. *Ann Epidemiol*. 2007;17:227–236.
11. VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiologic Methods*. 2014;3:33–72.
12. Murphy SA. Optimal dynamic treatment regimes. *J R Stat Soc Series B*. 2003;65:331–336.
13. Robins JM. Optimal structural nested models for optimal sequential decisions. In: Lin DY, Heagerty, H eds., *Proceedings of the Second Seattle Symposium in Biostatistics*. 2004;179:189–326.
14. Chakraborty B, Moodie EE. *Statistical Methods for Dynamic Treatment Regimes*. Berlin, Heidelberg, New York: Springer; 2013.
15. Hayward RA, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med Res Methodol*. 2006;6:18.
16. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA*. 2007;298:1209–1212.
17. Pocock SJ, Lubsen J. More on subgroup analyses in clinical trials. *N Engl J Med*. 2008;358:2076; author reply 2076–2077.
18. Abadie A, Chingos MM, West MR. *Endogenous Stratification in Randomized Experiments*. Review of Economics and Statistics. 2018;100:567–580.
19. Kent DM, Hayward RA, Griffith JL, et al. An independently derived and validated predictive model for selecting patients with myocardial infarction who are likely to benefit from tissue plasminogen activator compared with streptokinase. *Am J Med*. 2002;113:104–111.
20. Fox KA, Poole-Wilson P, Clayton TC, et al. 5-year outcome of an interventional strategy in non-ST-elevation acute coronary syndrome: the British Heart Foundation RITA 3 randomised trial. *Lancet*. 2005;366:914–920.
21. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005;365:176–186.
22. Pane JF, Griffin BA, McCaffrey DF, Karam R. Effectiveness of cognitive tutor algebra I at scale. *Educ Eval Policy Anal*. 2014;36:127–144.
23. Peck LR. Subgroup analysis in social experiments: measuring program impacts based on post-treatment choice. *Am J Eval*. 2003;24:157–187.
24. Hansen BB. The prognostic analogue of the propensity score. *Biometrika* 2008;95:481–488.
25. Cai T, Tian L, Wong PH, Wei LJ. Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics*. 2011;12:270–282.
26. Zhao L, Tian L, Cai T, Claggett B, Wei LJ. Effectively selecting a target population for a future comparative study. *J Am Stat Assoc*. 2013;108:527–539.
27. Luedtke AR, van der Laan MJ. Targeted learning of the mean outcome under an optimal dynamic treatment rule. *J Causal Inference*. 2015;3:61–95.
28. Luedtke AR, van der Laan MJ. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann Stat*. 2016;44:713–742.
29. Luedtke AR, van der Laan MJ. Optimal dynamic treatments in resource-limited settings. *Int J Biostat*. 2016;12:283–303.
30. Rubin D. Estimating causal effects of treatments in randomized and non-randomized studies. *J Edu Psychol*. 1974;66:688–701.
31. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6:Article25.
32. Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics*. 1985;41:361–372.
33. Piantadosi S, Gail MH. A comparison of the power of two tests for qualitative interactions. *Stat Med*. 1993;12:1239–1248.
34. Pan G, Wolfe DA. Test for qualitative interaction of clinical significance. *Stat Med*. 1997;16:1645–1652.
35. Silvapulle MJ. Tests against qualitative interaction: exact critical values and robust tests. *Biometrics*. 2001;57:1157–1165.
36. Li J, Chan IS. Detecting qualitative interactions in clinical trials: an extension of range test. *J Biopharm Stat*. 2006;16:831–841.

37. Moore JH, Gilbert JC, Tsai CT, et al. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J Theor Biol.* 2006;241:252–261.
38. Green DP, Kern HL. Modeling heterogeneous treatment effects in survey experiments with Bayesian additive regression trees. *Public Opin Q.* 2012;76:491–511.
39. Imai K, Ratkovic M. Estimating treatment effect heterogeneity in randomized program evaluation. *Ann Appl Stat.* 2013;7:443–470.
40. Berger JO, Wang X, Shen L. A Bayesian approach to subgroup identification. *J Biopharm Stat.* 2015;24:110–129.
41. VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component-cause framework. *Epidemiology.* 2007;18:329–339.
42. VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology.* 2009;20:863–871.
43. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction.* New York: Oxford University Press; 2015.
44. VanderWeele TJ, Knol MJ. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Ann Intern Med.* 2011;154:680–683.
45. van der Laan MJ, Luedtke AR. Targeted learning of the mean outcome under an optimal dynamic treatment rule. *J Causal Inference.* 2014;3:61–95.

APPENDIX: STATISTICAL ANALYSIS

In the eAppendix, <http://links.lww.com/EDE/B466>, we describe methods and formal statistical inference for estimating the optimal treatment rule and the outcome under it, and software to do so. While there are many ways to go about estimation, the methods described in the eAppendix, <http://links.lww.com/EDE/B466>, flexibly model the difference in observed outcomes across treatment groups conditional on the covariates and use an ensemble technique called “super-learner”³¹ that considers numerous different possible models or algorithms for the conditional outcome differences and then weights these according to their mean square error predictive value using cross-validation. Statistical inference for the optimal treatment rule and for the outcomes under it is challenging because the same data are being used to estimate

the treatment rule and the expected outcomes. Sample splitting can potentially be used but is not efficient, and averaging across split samples does not yield valid inference.⁴⁵ The eAppendix, <http://links.lww.com/EDE/B466>, describes a cross-validated targeted minimum loss-based approach to estimate the optimal treatment rule and the outcome under it. While the approach described in the eAppendix, <http://links.lww.com/EDE/B466>, has some desirable theoretical properties, considerable work remains to be done in assessing the sample sizes that are needed for these techniques to be useful and how the various methods that have been proposed in the literature compare with one another in actual practice. While the theoretical methodologic development has come a long way in the past decade, much remains to be learned about the application of these methods.