

ARTICLE

Heterogeneous treatment effect analysis based on machine-learning methodology

Xiajing Gong | Meng Hu | Mahashweta Basu | Liang Zhao

Office of Research and Standards,
Office of Generic Drugs, Center for
Drug Evaluation and Research, Food
and Drug Administration, Silver Spring,
Maryland, USA

Correspondence

Meng Hu, 10903 New Hampshire Ave.,
Bldg 75, Room 4649, Silver Spring, MD
20993-0002, USA.
Email: meng.hu@fda.hhs.gov

Funding information

No funding was received for this work.

Abstract

Heterogeneous treatment effect (HTE) analysis focuses on examining varying treatment effects for individuals or subgroups in a population. For example, an HTE-informed understanding can critically guide physicians to individualize the medical treatment for a certain disease. However, HTE analysis has not been widely recognized and used, even given the explosive increase of data availability attributed to the arrival of the Big Data era. Part of the reason behind its underuse is that data are often of high dimension and high complexity, which pose significant challenges for applying conventional HTE analysis methods. To meet these challenges, a newly developed causal forest HTE method has been derived from the random forest machine-learning algorithm. We conducted a systematic performance evaluation for the causal forest method against the conventional two-step method by simulating scenarios with different levels of complexity for the analysis. Our results show that causal forest outperforms the conventional HTE method in assessing treatment effect, especially when data are complex (e.g., nonlinear) and high dimensional, suggesting that causal forest is a promising tool for real-world applications of HTE analysis.

Study Highlights**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**

The arrival of the Big Data era brings explosive increases of data availability, but also imposes significant challenges for conventional heterogeneous treatment effect (HTE) analysis because of the equally increased data complexity. HTE methods based on machine-learning (ML) that have superior performance in handling complex data have not been introduced to the clinical pharmacology community.

WHAT QUESTION DID THIS STUDY ADDRESS?

What advantages can ML-based methods bring for HTE analysis when compared with conventional HTE methods? Conventional HTE methods construct separate models for different treatment groups and then estimate treatment effects by calculating the difference in the predicted responses from the separately built models.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

ML-based HTE methods were introduced to the community and showed superior performance over the conventional HTE method in (i) estimating treatment effects when covariates manifest nonlinear relationships and (ii) identifying influential variables of high-dimensional data with less sensitivity to data sizes and noise level.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

ML-based HTE methods are a promising tool to assess and predict heterogeneity for treatment effect for real-world applications, such as personalized medicine and policy making.

INTRODUCTION

Treatment effect refers to the causal effect of a treatment or intervention (e.g., administering an anticancer drug) on an outcome of interest (e.g., health or disease progression of the patient) based on the counterfactuals (e.g., difference in outcomes with/without using the drug). Treatment effects are rarely perfectly homogeneous over the population. For instance, a new treatment may perform similarly to an existing treatment in the overall population but may be extremely beneficial to a subgroup of subjects with specific characteristics. Thus, it can be difficult to apply the average treatment effect to address questions concerning individual outcome, for example, for personalized medicine.¹ As the arrival of the Big Data era brought dramatically increased data volume and complexity (e.g., nonlinear and/or high-dimensional data), handling complex data has become an important research topic across multiple disciplines.²⁻⁶ Recently, the analysis of the heterogeneous treatment effect (HTE), conducted to reflect the nonrandom variation in a treatment effect over a population, has drawn growing attention in a variety of fields from economics to medicine.^{7,8} It is worth noting that although the response analysis predicts the outcome itself, HTE analysis focuses on estimating the expected change in outcome as a result of the treatment for individuals.⁹ For example, a tree service company wants to identify a subgroup of customers who will sign a contract after receiving a phone call advertising the service but would take no action without the phone call. Considering phone advertising and signing a contract as the treatment and outcome, respectively, HTE analysis can provide information to identify the subgroup of interest, which will improve the marketing strategy in terms of cost-effectiveness with regard to the fact that it would not be cost-effective to keep advertising to a group who will sign the contract regardless of receiving the phone advertising.

Although HTE analysis has also been applied to drug development, including clinical trials,¹⁰ study design,¹¹ and personalized medicine,¹² conducting HTE analysis

can be a challenging task. One unique challenge is that the quantity to be estimated (i.e., treatment effect) is often unknown on given data, as each subject can often only be exposed to one condition of treatments, which is also known as the fundamental problem of causal inference.¹³ Previously, an intuitive two-step model was developed to conduct HTE analysis. The two-step model first builds separate models for different treatment groups (e.g., treatment vs. control), and the treatment effect for each individual is then estimated by calculating the difference in the predicted responses from the separately built models. However, despite its intuitiveness and simplicity, the two-step model suffers from several drawbacks.¹⁴ The most important drawback is that the difference between two independent, accurate models does not necessarily result in an accurate model. In addition, the separately built models are often based on ordinary regression models, and thus the model performance of the two-step method could be compromised when dealing with nonlinear and/or high-dimensional data. Alternatively, to assess HTE, a regression model could be built containing prespecified interactions between treatment and covariate(s), considering that the interaction has been acknowledged as a major source contributing to HTE.¹⁵ However, to implement this method, sufficient knowledge is needed to predefine potential interactions, and it is an almost impossible task in the case of high-dimensional data.¹⁶ Recently, machine-learning (ML) methodologies have been employed in HTE analysis, especially with tree-based approaches. Tree-based models refer to a family of ML models based on binary trees obtained with the classification and regression tree algorithm.¹⁷ In such trees, binary splits recursively partition a full data set into homogeneous or near-homogeneous subsets (dubbed as “leaf” of tree). As such, tree-based models can serve as a natural solution to estimate HTE if appropriate split criteria can be designed to reflect population subgroups in terms of treatment effect.^{9,14} Causal forest, an HTE method based on random forest, is one of the most recent advances in tree-based

HTE method. It has been developed to overcome potential issues observed in HTE analysis with the use of the single-tree method.¹⁸ More important, causal forest is ML based and has no assumption on the data (e.g., linear relationship between covariates) and thus has flexibility to handle complex practical problems.¹⁰

Overall, the treatment effect information obtained by HTE analysis can significantly improve the trial study design in the drug development process and potentially guide personalized medicine.¹⁹ This study has the following two main aims: (1) highlight the advantageous benefits associated with HTE analysis over the conventional response analysis on average effects and (2) perform a systematic performance analysis for the conventional two-step and causal forest HTE methods. To the best of our knowledge, no comprehensive performance evaluation has been conducted for these methods. We therefore simulated scenarios with different levels of complexity in terms of HTE to fully characterize the ability of the two methods to identify effect heterogeneity. The simulation approach was used because it allows the explicit specification of HTE and maintains ground truth information for a model performance check. Of note, unfoundedness (i.e., randomized treatment assignment) was one key assumption when causal forest was developed for HTE analysis.⁹ For observational studies that often retain confounding factors, data adjustment for case-control comparisons (e.g., matching)²⁰ that adjust original observational data to obtain a relatively balanced treatment assignment, as expected in a randomized study, can be applied before conducting HTE analysis using causal forest. In this study, one simulation was provided to mimic observational studies with confounding factors to demonstrate the use of causal forest for observational data.

METHODS

In this section, after describing the basic principle of treatment effect, we introduce the concept of HTE and

graphically illustrate the differences between no treatment heterogeneity and HTE. Subsequently, we report the methodology development for HTE analysis from the conventional two-step method through a tree-based approach to the causal forest. Lastly, we describe the simulation models and performance evaluation methods.

Heterogeneous treatment effect

Treatment effect refers to counterfactual effect of a treatment on an outcome. Without loss of generality, for individual i , define T_i as the binary treatment indicator (e.g., 1 = treatment; 0 = control) and Y_i as the outcome (e.g., real values). That is, $Y_i(0)$ and $Y_i(1)$ correspond to the outcomes from different treatments (0 or 1, respectively). Thus, the treatment effect for individual i can be represented as $Y_i(1) - Y_i(0)$, and the average treatment effect is denoted as $E[Y_i(1) - Y_i(0)]$.

It is well known that outcomes of a treatment are dependent on the individual characteristics (covariates), such as a patient's medical history and demographic information (e.g., sex, age, and ethnicity). It is natural to infer that the treatment effect is usually not homogeneous among individuals. As a demonstration of the HTE concept, Figure 1 shows illustrative examples of homogeneous treatment effect (Figure 1a) and HTE (Figure 1b) among the population. In the homogeneous treatment effect, although the outcomes show variation across the individuals for each treatment group and diverge between treatment groups ($T = 1$ vs. $T = 0$), the treatment effect $Y_i(1) - Y_i(0)$ is the same for every individual and identical to the average treatment effect (Figure 1a). In HTE, the population shows significant heterogeneity in response to treatments, with some individuals benefiting more (i.e., responders), some less, and some not benefiting at all (i.e., nonresponders) from the treatment (Figure 1b). Thus, the average treatment effect is of limited value to provide

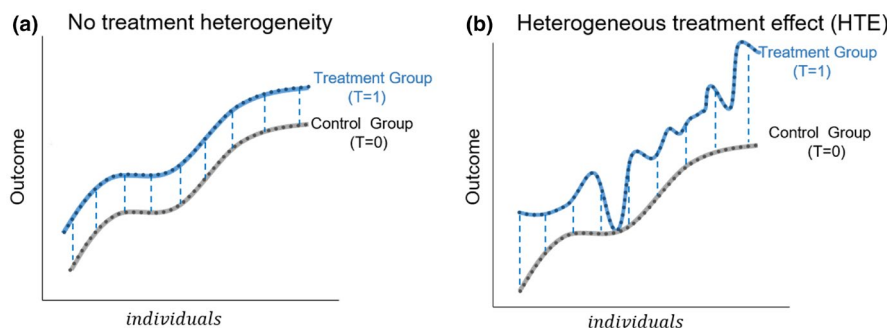


FIGURE 1 (a) Homogenous treatment effect (no treatment heterogeneity): the outcome of the treatment shows variation across the individuals and between treatment groups, but the treatment effect (i.e., the difference of the outcomes depicted by the dotted lines between the two treatment outcome curves) is the same for every individual. (b) Heterogeneous treatment effect (HTE): treatment effect varies among individuals. Some individuals benefit more, some less, and some might not benefit at all from the treatment

information for individuals, and HTE analysis is needed to understand how the treatment effects vary among the whole population.

For many applications, HTE analysis shares the same fundamental challenge with the causal inference method, as only one of two potential outcomes—either $Y_i(0)$ or $Y_i(1)$ —is observable for individuals; that is, the treatment effect ($Y_i(1) - Y_i(0)$) is not explicitly provided by the original data. As such, because of the unique challenge, HTE analysis methods must be specifically developed to address such issues.

Methods to estimate HTE

Two-step method

One of the commonly used approaches to estimate HTE is a two-step method,²¹ which builds separate regression models for the treatment and control groups. This counterfactual model consisting of the two constructed regression models is then used to estimate the counterfactual differences in individual outcomes to infer the individual treatment effects. Specifically, for an individual with distinct covariate values, each regression model can project outcome values, and the difference between the two outcomes will represent the predicted treatment effect. The two-step method has been conventionally used in many fields, such as econometrics,²² social science,²³ epidemiology,²⁴ and medical science.²⁵ Despite being intuitive and straightforward to implement, this method can be constrained by the nature of linear regression, which imposes linear relationships unless more complex relationships are explicitly predefined in the model. As such, its performance can be significantly compromised in the presence of model misspecification for complex relationships. Another intrinsic drawback is that the difference between the two independent “accurate” models does not necessarily lead to an accurate HTE estimate.

ML method: causal forest

Several ML approaches have been developed to estimate HTE.^{14,26–29} Among them, the decision tree–based HTE method was first developed and has been widely recognized.^{30,31} The essence of a decision tree, featured by partitioning full data into subgroups, makes it perfectly suitable for HTE analysis aiming to find subgroups (or individuals) with a distinct treatment effect.

One milestone for the method development of decision trees is the emergence of the random forest algorithm.³² Considering the greedy nature of one-step-at-a-time node

splitting in binary trees, random forest attempts to mitigate the “overfitting” issue (i.e., inability to generalize unseen data) of a single binary tree by implementing a randomization procedure. Randomization is carried out in the following two forms: (1) a collection of binary decision trees that independently grew based on the bootstrap sample of the original data and (2) a randomly selected subset of variables that are chosen as candidate variables for splitting at each node of the tree.³² The random forest combines hundreds or thousands of trained decision trees and makes its final predictions by averaging the predictions of each individual tree. Recently, random forest has also been extended to HTE analysis, specifically with the causal forest method.⁹ Briefly, the causal forest method keeps the main structure of random forest such as the recursive partitioning, subsampling, and random split selection, but the tree-splitting criteria are modified to suit the goal of HTE analysis, that is, maximizing the treatment effect heterogeneity—the difference of estimated treatment effect between daughter nodes. It is worth mentioning that, although named as causal forest by the method developers,⁹ this method performs HTE analysis based on the estimated counterfactuals within nodes rather than carrying out a standard causal inference that requires specific designs of questions, studies, and analysis.¹³ Because of the word limit, detailed descriptions of the causal forest are provided in the Supplementary Information.

Models for simulation

Interactions between treatment and covariates of subjects can lead to HTE among the study population, which lays the basis for our simulations. Define $X_i = (x_{i1}, \dots, x_{ik})$ as a vector of observable covariates for subject i and $\tau(x) = E[Y_i(1) - Y_i(0) | X_i = x]$ as the treatment effect for a given set of covariates x . The HTE scenario can be simulated by developing models,¹⁶ such as:

$$Y_i(T) = f(X_i) + g(X_i)T + \varepsilon \quad (1)$$

where $f(\cdot)$ refers to the direct impact of covariates (with no interaction with treatment indicator) on the outcome Y_i , $g(\cdot)$ is a function of covariates X_i that interact with treatment, and the interaction term $g(X_i)T$ specifies the HTE.

For the scenario with homogeneous treatment effect, the data can be simulated by denoting outcome (Y) as:

$$Y_i(T) = f(X_i) + \delta T + \varepsilon \quad (2)$$

where the treatment effect of each individual is the same as $E[Y_i(1) - Y_i(0)] = \delta$, independent of the covariates.

Equation (2) is a special case of Equation (1) without interactions between treatment and covariates. As such, Equation (1) can simulate different HTE scenarios by varying the form of the function g . In this study, we developed the following four models with increasing complexity of HTE (Table 1): (I) no heterogeneity covariates (i.e., no HTE), that is, all observations have the same treatment effect; (II) a linear relationship between heterogeneity covariates; (III) a nonlinear relationship between heterogeneity covariates; and (IV) high-dimensional data where the number of covariates exceeds the number of individuals/observations. For each simulation model, the covariate of the individuals X_i were generated randomly from a mean-zero multivariate normal distribution with covariance matrix $(1 - \rho) \mathbf{I}_k + \rho \mathbf{1} \mathbf{1}^T$. Treatments $T \in \{-1, 1\}$ were randomly assigned to the whole population, and random errors were defined as $\epsilon \sim N(0, \sigma_0^2)$. No correlation was assumed among the covariates, that is, covariance $\rho = 0$ and $\sigma_0 = 0.5$ for the noise level. We set the sample size as $n = 2000$ and the number of covariates as $k = 10$ in Models I, II, and III, and $n = 500$ and $k = 1000$ in the high-dimensional case of Model IV.

Using these simulated scenarios, we conducted a systematic performance evaluation for causal forest and compared it with the two-step method. For each scenario (Models I–IV), we independently generated 200 data sets. Each data set consisted of training and testing data independently generated from the given model. The training data were used to build the predictive HTE model using the two-step method or causal forest, whereas the testing data were used to examine the predictive ability of the established model. Given the predicted and true treatment effects, the predictive ability was evaluated, and the performance evaluation metrics were averaged over the 200 simulation replications. Both the root mean square error (RMSE) and the incremental gains curve¹⁴ were used to evaluate model performance (see the Supplementary Information for detailed descriptions).

Moreover, for observational studies where the unconfoundedness assumption may be violated, the approach of using estimated propensity scores^{33,34} can be applied with

causal forest for its robustness to confounding. A scenario with confounding factors was simulated to demonstrate the use of causal forest for observational data (see the Supplementary Information).

RESULTS

Case example

In this section, a case example with nonlinear HTE was simulated to demonstrate HTE analysis conducted by causal forest and the two-step method. By stipulating Equation (1), the example case was simulated based on the model $Y = \beta_0 + \sum_{k=1}^{10} \beta_k x_k + (\gamma_0 + \gamma_1 x_1^5 + \gamma_2 e^{x_2} + \gamma_{12} x_1 x_2) T + \epsilon$, where $\beta_0 = 0.4$, $\beta_k = 0.5$, and heterogeneity covariates x_1 and x_2 have interactions with the treatment (T) in a nonlinear form, $\gamma_0 + \gamma_1 x_1^5 + \gamma_2 e^{x_2} + \gamma_{12} x_1 x_2$ ($\gamma_0 = 0.8$, $\gamma_1 = \gamma_{12} = 1.6$, $\gamma_2 = -0.64$). As such, a nonlinear relationship between treatment effects and covariates was created in the model. Figure 2a illustrates the distribution of the model-defined true treatment effects as a function of covariates x_1 and x_2 (x and y axes). In Figure 2a, a color gradient corresponding to the values of treatment effects is used, with red indicating the highest treatment effect (lower right corner) and blue indicating the most unexpected treatment effects (the leftmost region in the middle). A nonlinear transition from the highest to the lowest treatment effect can be observed.

Both causal forest and the two-step methods were applied to the simulated data for HTE analysis. Specifically, for the two-step method, a separate regression models were established for the treatment and control groups. When predicting treatment effect for an individual, the data of this individual were fed into the two separately established models, and the difference between the estimated outcomes was calculated to represent the treatment effect. For the causal forest, each tree grew (split) according to the splitting criteria. The established causal forest model can then provide the predicted treatment effect for individuals. Figure 2b displays the distribution of treatment effects as estimated by the causal forest method, which almost matches the true treatment

TABLE 1 Summary of the four simulation mathematical models generated with increasing heterogeneous treatment effect complexity

Model	Description of relationships between heterogeneity covariates	Outcome model
I	No heterogeneous treatment effect	$Y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \delta T + \epsilon$
II	Linear	$Y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + (\gamma_0 + \sum_{k=1}^p \gamma_k x_{ik}) T + \epsilon$
III	Nonlinear + interactive	$Y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + (\gamma_0 + \gamma_1 x_{i1}^3 + \gamma_{23} \cos(x_{i2}) x_{i3}) T + \epsilon$
IV	High-dimensional covariates	Model II

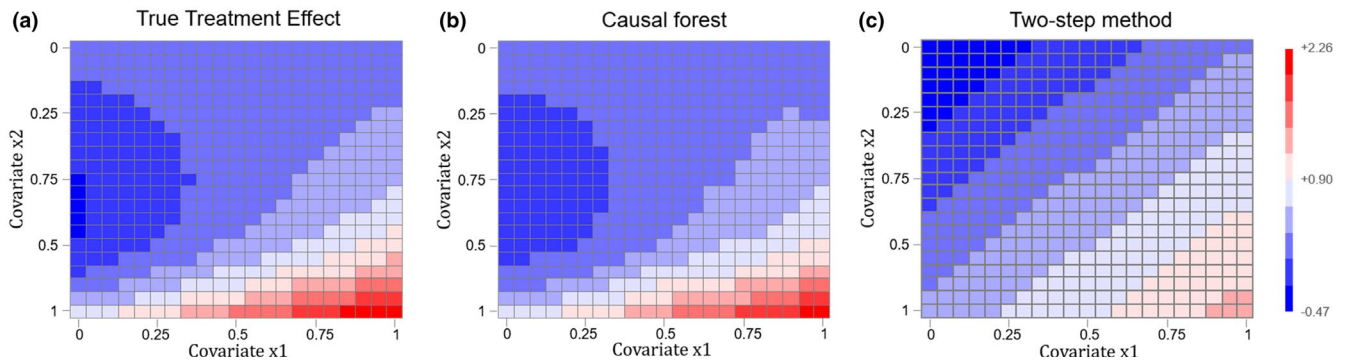


FIGURE 2 Treatment effect in a data set where the relationship between the two covariates $\times 1$ and treatment effect is nonlinear. The (a) “true” treatment effect with varying values of $\times 1$ and $\times 2$ and the predicted treatment effect using (b) causal forest and (c) the two-step method. The treatment effect is denoted by color from blue (low) to red (high)

effects shown in Figure 2a. In contrast, Figure 2c shows the biased treatment effect estimation with a linear pattern provided by the two-step method, which is restricted by its explicitly linear additive regression model. Although the estimation reflects the general trend of the true treatment effects, the nonlinear pattern between the treatment and covariate cannot be recovered (Figure 2c). These results demonstrate that the causal forest method can detect the underlying HTE even when a complex interaction relationship exists between the treatment and covariates.

Simulations based on mathematical Models I–IV

To conduct a systematic performance check on the causal forest and the conventional two-step method for HTE analysis, we simulated four scenarios based on hypothetical mathematical models with various relationships among the covariates and treatment effect with progressively increasing complexity of HTE (see the Methods section and Table 1). Model I was intended as a “baseline” case without interaction between the covariates and treatment, and therefore no HTE. In Model II, two covariates interacted with treatment in a linear form. Model III assumed a case where the covariates interacted with the treatment in a nonlinear form. We designed Model IV to represent the high-dimensional data scenario, in which 1000 covariates were sampled from multivariate normal distribution with 500 observations. In this model, the 1000 covariates followed a linear additive relationship with respect to treatment effect, and the covariate coefficients were set to zero except for the first 5 covariates. This allowed us to examine if an HTE model can correctly describe such high-dimensional data and offer insights into the preset important covariates even when they are sparse (i.e., 5/1000).

Performance evaluation

For each scenario (Models I–IV), we generated 200 data sets, each of which include one training and one testing data independently generated from the given model. The training data were used to build the predictive HTE model using the two-step or causal forest method, whereas the testing data were used to examine the predictive ability of the established model. Given the predicted and true treatment effects, the predictive ability was then evaluated by the RMSE and the incremental gains curve (see the Methods section).

Figure 3 shows the RMSE results calculated from each testing data for the four scenarios. The figure shows the mean and standard deviation of RMSE values across 200 simulation replications. No significant difference can be seen in the prediction performance between the causal forest and the two-step method for Model I (no HTE) and Model II (linear interaction between treatment and covariates). For these two scenarios, the two-step method is the correct modeling method and is thus expected to provide accurate predictions. However, the causal forest method performs equally well as the two-step method for these scenarios. For Model III (nonlinear or additive interaction between treatment and covariates), causal forest provides a more accurate treatment effect prediction than the two-step method, reflected by the significantly lower RMSE ($p < 0.01$). Such a finding further highlights that causal forest can be used to conduct an analysis for complex/real questions.

The difference in RMSE is most prominent in Model IV ($p < 0.01$) (Figure 3). The two-step method fails to yield a reasonable treatment effect estimation because of the parameter identifiability issue, considering that the number of observations (500) is less than the number (1000) of covariates. Causal forest achieves more than fivefold improvement over the two-step method in terms of RMSE. Importantly, causal forest is also able to capture influential

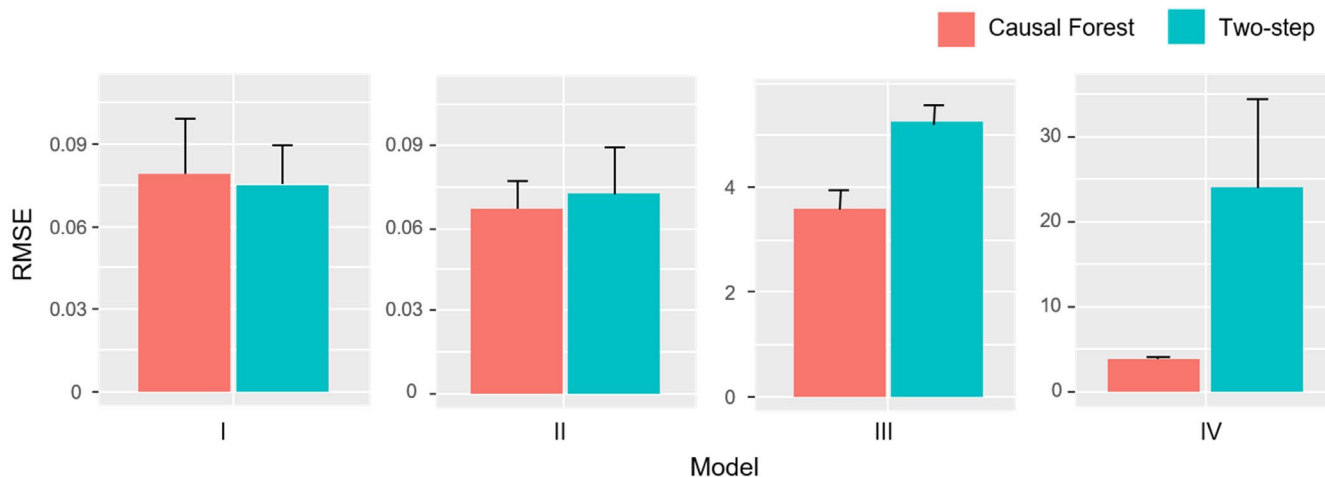


FIGURE 3 Comparison of the performance of the causal forest and two-step methods. Results are based on 200 replicated simulations. Mean (bar height) and standard deviation (error bar) of the root mean square error (RMSE) are displayed

covariates based on the variable importance, measured by a simple weighted sum of the number of times that the covariate of interest was split at each depth in the forest. Higher importance values indicate corresponding covariates with high predictive power, whereas lower to zero values indicate variables with low predictive power. Note that the variable importance as defined for causal forest is different from that for random forest (i.e., variable importance),³⁵ which is calculated by prediction error change after noising up a variable. Causal forest can provide calculated variable importance values. However, it does not include a criterion that can be used to identify whether variable importance values are statistically meaningful or just obtained from randomness. Therefore, we developed a permutation-based statistical significance test to identify statistically meaningful important covariates (refer to the Supplemental Information for a detailed description). The results from the significance test show that only the first 5 covariates have statistically meaningful variable importance values, consistent with the simulation by design (Figure 4).

In addition to RMSE, we also applied the incremental gains curve as another model assessment tool and calculated the corresponding Qini coefficient (Figure 5) based on the simulated data representing a population of individuals with HTE. Figure 5 illustrates the cumulative number of incremental responses relative to the cumulative number of targets (both expressed as a percentage of the total targets): the x-axis shows the fraction of individuals in the population in which the treatment is performed, and the y-axis shows the incremental number of positive outcomes between the treatment and control groups expressed as percentage of the size of the target (treated) population. The dashed diagonal line in the figure denotes the benchmark. The Qini coefficient (marked on each

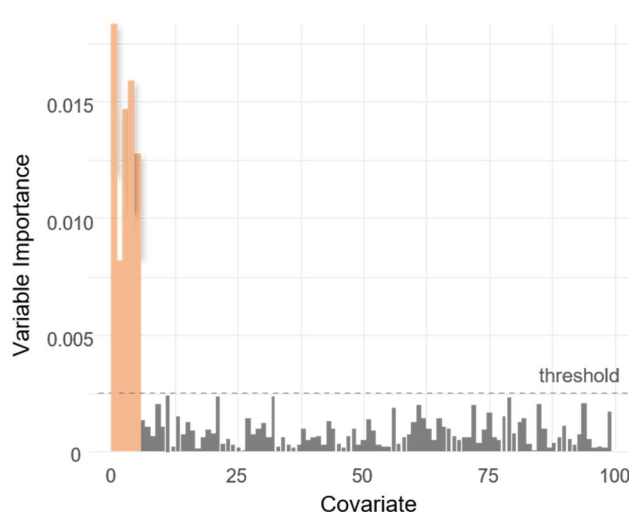


FIGURE 4 The variable importance determined by causal forest for high-dimensional simulated data based on Model IV. The preset significant covariates are shown in orange. The five preset important covariates were identified, as their variable importance values are greater than the significance threshold (dashed). Please refer to Supplementary Information for a description of the statistical significance test

panel) is a single estimate of model performance. Relative Qini values should be compared within the same data set rather than comparing absolute values among different data sets. Given that there is no HTE designed in Model I, the incremental gains curves of the causal forest and the two-step method overlap with the benchmark line. For Model II (linear HTE), both methods can capture equally well the HTE information and can regroup subjects based on the better treatment effect gain rather than the random regrouping (benchmark). For Model III (nonlinear), causal forest performs better than the two-step method for

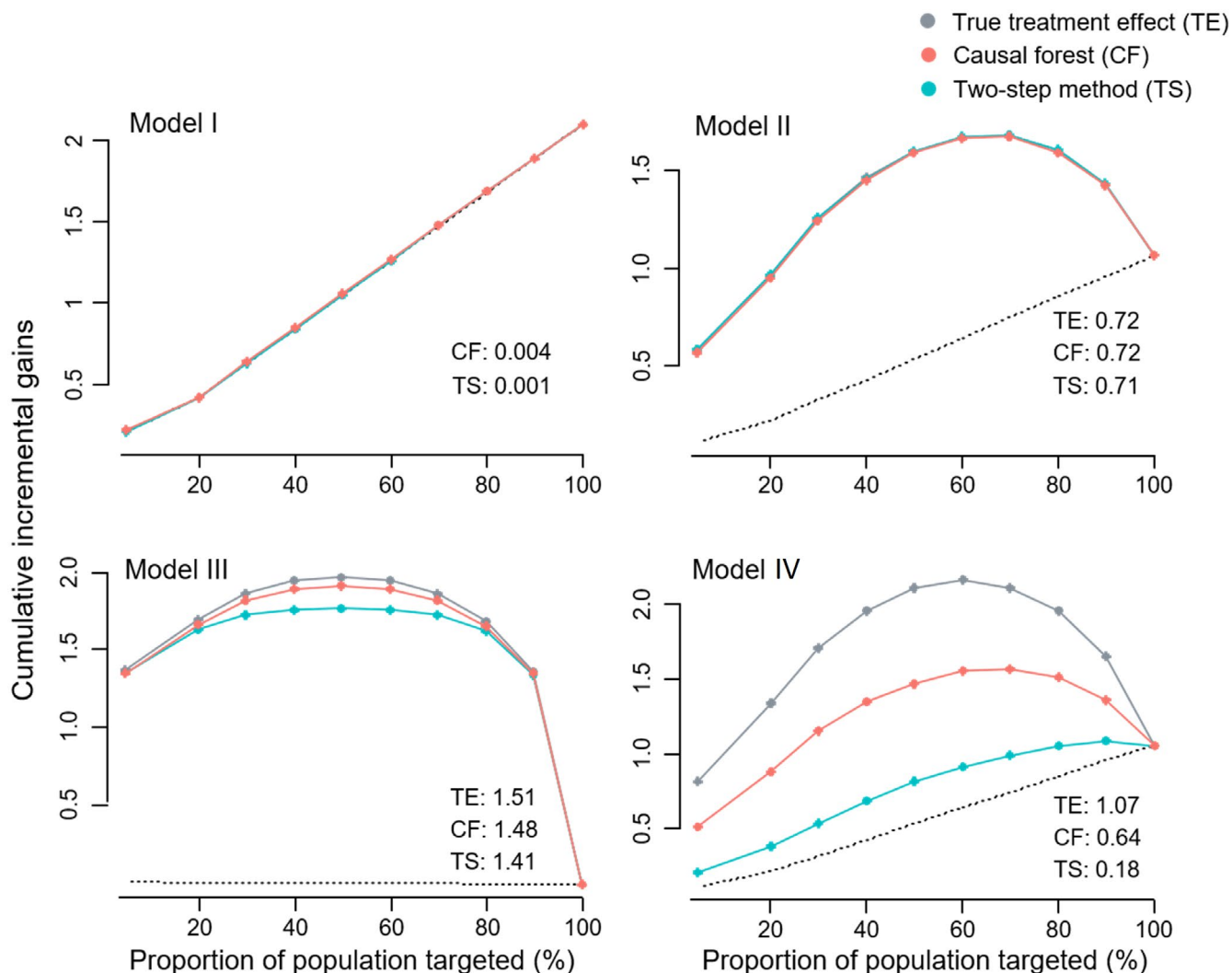


FIGURE 5 Incremental gains curves (or Qini curves) from each model. This curve shows the cumulative number of incremental individuals with positive treatment effect relative to the cumulative number of the targeted population. The dashed diagonal line depicts the theoretical incremental individuals with positive treatment effect from random targeting, whereas the gray line refers to the true treatment effect. For each mode, the incremental gain curves shown are the average of all the curves from 200 simulation replications. The Qini coefficients displayed on each panel are the average values from 200 simulation replications

up to 80% of the target population. For Model IV (high dimensionality), causal forest shows overall better performance at every level (percentage) of the target population (Figure 5). For both Models III and IV, the incremental gains curves of the causal forest appear closer to the curves for model-defined true treatment effect compared with the two-step method. The Qini coefficients reflect the previously described observations. Consistent with the RMSE results, the Qini coefficient from causal forest is higher than that from the two-step method in Models III and IV ($p < 0.01$), with an even larger relative difference exhibited in Model IV (high dimensionality).

The overall findings from our simulation studies clearly show that the ML-based causal forest method can outperform the linear regression-based two-step method for HTE analysis with nonlinear and high-dimensional

data. In addition, data sensitivity testing was conducted to examine the robustness of the ML-based causal forest method to the sample size and noise level of data (see the Supplementary Information).

In addition, results from the simulation with confounding factors shows improved predictability of causal forests using estimated propensity scores, indicating that it can be useful in observational studies with confounding factors (see the Supplementary Information).

DISCUSSION

This study introduced an ML-based HTE analysis and presented a systematic evaluation of HTE analysis methods. HTE analysis demonstrated its ability to address questions

on the individualized counterfactual treatment effects. Thus, HTE analysis can be employed in a wide range of applications, including those related to clinical trial design and personalized medicine. The HTE analysis method has evolved to meet the challenges posed by the diverse data sources with increased data volume and complexity accompanying the Big Data era. In this study, causal forest, a recently developed HTE method that is based on a well-established ML-method random forest, was introduced to the clinical pharmacology community, and its performance was systematically assessed by predicting treatment effects under several simulated scenarios. Our results showed that causal forest outperforms the conventional two-step HTE method, especially for nonlinear or high-dimensional data.

Causal forest can be considered as an extension of random forest for HTE analysis, and as such inherits several appealing properties from random forest. Causal forest (1) imposes a minimal assumption on data, rendering it suitable for handling complex data; (2) uses data partition as the training process, making it a natural candidate for HTE analysis; and (3) is capable of processing “large feature” data, where the number of covariates is much greater than the number of observations. During the training process of causal forest (i.e., generation of decision trees), the tree grows from the root (whole data set) to leaves (subsets), with each leaf determined through a series of binary splitting. For each splitting, a covariate is selected from the current node to maximize the difference of treatment effects of daughter nodes. As such, each leaf represents a subgroup with a distinct treatment effect featured by a certain set of covariates. Through the training process, relationships between treatment effect and covariates are revealed, and thus HTE analysis can be conducted. Theoretically, the training process could be compromised if a considerable number of subjects share similar covariates but exhibit significantly different responses. The response difference may derive from systematic randomness, measurement error, or unobserved influential covariates that were not included in the study. If the aforementioned phenomenon occurs only for a few subjects, the training process should be still sufficient, partly because subjects sharing the same set of covariates may still be in the same node and the dominating treatment effect will still be reflected against the particular covariates selected for node splitting. Furthermore, this limitation is reasonably likely to affect HTE analysis based on a single decision tree more severely than the HTE analysis based on causal forest, as the latter retains data bagging and covariate random selection processes.³²

When the causal forest was developed, “unconfoundedness” was a key assumption for implementing HTE analysis,⁹ which indicates that treatments in a study are

randomly/unbiasedly assigned to different experiment groups with matched covariates, as expected in a randomized experiment. However, in observational studies, confounding can influence both outcomes and treatments, thus resulting in unbalanced treatment assignments. Uncontrolled confounding can lead to biased analysis. Data adjustment for case-control comparison (e.g., matching) is the common practice to address confounding.²⁰ In general, matching methods adjust original data to mitigate differences in the distributions of covariates between treatment groups, leading to a relatively balanced treatment assignment to approximate a randomized experiment. For observational data where confounding often exists, the data-matching process is warranted before conducting HTE analysis by causal forest. Although an approach of using estimated propensity scores has been proposed for causal forest,^{33,34} there is a diversity of other applicable matching methods,²⁰ including ML-based methodologies.³⁶

In addition to the two-step and causal forest methods, other HTE analysis methods have been proposed and assessed using linear models based on transformed covariates,¹⁶ adapted support vector machines,²⁷ Bayesian trees,^{26,29} and forests.³⁷ Among them, the uplift model^{14,38} was developed to maximize the return on investment (e.g., for marketing) by essentially conducting an HTE analysis with more application focus. In business settings, a random forest-based uplift model has been commonly used to optimize the selection of insurance policies¹⁴ and personalized marketing interventions.³⁹ Compared with the uplift model, causal forest can implement honest forest (also see the Supplementary Information)⁹ and reduce the potential estimation bias and therefore provide credible confidence interval estimations for the model-predicted treatment effects. Overall, both approaches are based on the random forest model and aim to address similar HTE questions (e.g., both can be used for survival analysis^{40–43}). Although our study focused on HTE analysis with a single treatment, it is worth noting that the uplift model and causal forest are applicable in circumstances with multiple possible treatments.^{44–46}

Part of the recognized merit associated with HTE analysis is that it makes the treatment effect not only measurable but also actionable. A variety of domains can benefit from the knowledge of treatment effects over the population of interest. For example, a retail store can make a more cost-effective plan to offer sales (i.e., treatment) to the identified target subpopulation to maximize transactions (i.e., outcome), and an oncologist can individualize treatment plans based on HTE results. In a regulatory setting, an agency can use HTE analysis to inform policy making and regulatory action to maximize the interest of the public. For example, one critical mission of the

Office of Generic Drug (OGD) at the U.S. Food and Drug Administration is to facilitate the generic drug development and increase drug accessibility for the US public.⁴⁷ OGD regularly releases product-specific guidance (PSG) to share the OGD's current thinking and recommendations on the development of a generic product following new drug approval.⁴⁸ The PSG release, if considered as a treatment, can be evaluated quantitatively for its impact on the timing and the number of Abbreviated New Drug Application submissions by the HTE analysis. This effort will be discussed in a separate article emphasizing real-world applications.

CONCLUSION

Given its resilience in handling complex data (e.g., nonlinear and/or high-dimensional data), the causal forest HTE method, an ML approach derived from a random forest algorithm, provides a unique opportunity for scientists to assess and predict heterogeneity for treatment effect for real-world applications. Further research is warranted to extrapolate its applications to support decision making for both medical and regulatory practices.

DISCLAIMER

The opinions expressed in this manuscript are those of the authors and should not be interpreted as the position of the US Food and Drug Administration.

ACKNOWLEDGMENTS

The authors thank Drs. Sara Lomonaco and Jing (Jenny) Wang for their support in preparing the manuscript.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

X.G., M.H., M.B., and L.Z. wrote the manuscript. M.H., X.G., and L.Z. designed the research. X.G., M.H., M.B., and L.Z. performed the research. X.G., M.H., and M.B. analyzed the data.

REFERENCES

- Schork NJ. Personalized medicine: time for one-person trials. *Nature*. 2015;520(7549):609-611.
- Garrido MM, Deb P, Burgess JF, et al. Choosing models for health care cost analyses: issues of nonlinearity and endogeneity. *Health Serv Res*. 2012;47(6):2377-2397.
- Belloni A, Chernozhukov V, Hansen C. High-dimensional methods and inference on structural and treatment effects. *J Economic Perspect*. 2014;28(2):29-50.
- Tsay RS, Chen R. *Nonlinear Time Series Analysis*. Vol. 891. John Wiley & Sons; 2018.
- Fernández-Villaverde J, Rubio-Ramírez JF. Estimating dynamic equilibrium economies: linear versus nonlinear likelihood. *J Appl Econ*. 2005;20(7):891-910.
- Hu M, Liang H. Adaptive multiscale entropy analysis of multivariate neural data. *IEEE Trans Biomed Eng*. 2011;59(1):12-15.
- Atthey S, Imbens G. Recursive partitioning for heterogeneous causal effects. *Proc Natl Acad Sci*. 2016;113(27):7353-7360.
- Guelman L. *Optimal personalized treatment learning models with insurance applications*; 2015.
- Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc*. 2018;113(523):1228-1242.
- Goldstein BA, Rigdon J. Using machine learning to identify heterogeneous effects in randomized clinical trials—moving beyond the forest plot and into the forest. *JAMA Network Open*. 2019;2(3):e190004.
- Varadhan R, Seeger JD. Estimation and reporting of heterogeneity of treatment effects. In: Velentgas P, Dreyer NA, Nourjah P, Smith S, Torchia M, eds. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Rockville, MD: Agency for Healthcare Research and Quality; 2013:35-44.
- Powers S, Qian J, Jung K, et al. Some methods for heterogeneous treatment effect estimation in high dimensions. *Stat Med*. 2018;37(11):1767-1787.
- Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc*. 2005;100(469):322-331.
- Guelman L, Guillén M, Pérez-Marín AM. Uplift random forests. *Cyber Syst*. 2015;46(3-4):230-248.
- Friedman J, Hastie T, Tibshirani R. *The Elements of Statistical Learning*. Springer; 2001. *Springer Series in Statistics*; vol. 1.
- Tian LU, Alizadeh AA, Gentles AJ, et al. A simple method for estimating interactions between a treatment and a large number of covariates. *J Am Stat Assoc*. 2014;109(508):1517-1532.
- Breiman L, Friedman J, Stone C, Olshen R. *Classification and Regression Trees*. CRC Press; 1984.
- Kingsford C, Salzberg SL. What are decision trees? *Nat Biotechnol*. 2008;26(9):1011-1013.
- Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. 2018;363:k4245.
- Colson KE, Rudolph KE, Zimmerman SC, et al. Optimizing matching and analysis combinations for estimating causal effects. *Sci Rep*. 2016;6(1):1-11.
- Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics*. 1985;41(2):361-372.
- Hsu Y-C, Shen S. Testing treatment effect heterogeneity in regression discontinuity designs. *J Econom*. 2019;208(2):468-486.
- Xie Y, Brand JE, Jann B. Estimating heterogeneous treatment effects with observational data. *Sociol Methodol*. 2012;42(1):314-347.
- Nilsson A, Bonander C, Strömberg U, et al. Assessing heterogeneous effects and their determinants via estimation of potential outcomes. *Eur J Epidemiol*. 2019;34(9):823-835.
- Raman G, Balk EM, Lai L, et al. Evaluation of person-level heterogeneity of treatment effects in published multiperson N-of-1 studies: systematic review and reanalysis. *BMJ Open*. 2018;8(5):e017641.

26. Green DP, Kern HL. Modeling heterogeneous treatment effects in survey experiments with Bayesian additive regression trees. *Public Opinion Quarterly*. 2012;76(3):491-511.
27. Imai K, Ratkovic M. Estimating treatment effect heterogeneity in randomized program evaluation. *Ann Appl Stat*. 2013;7(1):443-470.
28. Radcliffe NJ. Using control groups to target on predicted lift: Building and assessing uplift models. *Direct Market Anal J*. 2007;1(3):14-21.
29. Hill JL. Bayesian nonparametric modeling for causal inference. *J Comput Graph Stat*. 2011;20(1):217-240.
30. Hansotia B, Rukstales B. Incremental value modeling. *J Interact Market*. 2002;16(3):35.
31. Zeileis A, Hothorn T, Hornik K. Model-based recursive partitioning. *J Computat Graph Stat*. 2008;17(2):492-514.
32. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.
33. Athey S, Tibshirani J, Wager S. Generalized random forests. *Ann Stat*. 2019;47(2):1148-1178.
34. Athey S, Wager S. Estimating treatment effects with causal forests: an application. *Observat Stud*. 2019;5(2):37-51.
35. Ishwaran H. Variable importance in binary regression trees and forests. *Electron J Stat*. 2007;1:519-537.
36. Pirracchio R, Petersen ML, Van Der Laan M. Improving propensity score estimators' robustness to model misspecification using super learner. *Am J Epidemiol*. 2015;181(2):108-119.
37. Taddy M, Gardner M, Chen L, et al. A nonparametric bayesian analysis of heterogenous treatment effects in digital experimentation. *J Bus Econ Stat*. 2016;34(4):661-672.
38. Gutierrez P, Gérardy J-Y. Causal inference and uplift modelling: a review of the literature. In: International Conference on Predictive Applications and APIs; October 10-12, 2015; Boston, MA.
39. Guelman L, Guillén M, Pérez-Marín AM. A decision support framework to implement optimal personalized marketing interventions. *Decis Support Syst*. 2015;72:24-32.
40. Jaroszewicz S, Rzepakowski P. Uplift modeling with survival data. In: *ACM SIGKDD Workshop on Health Informatics (HI-KDD-14)*; August 24, 2014; New York, NY.
41. Ramachandra V. *Causal Inference for Survival Analysis*. arXiv 2018;1803.08218.
42. Cui Y, et al. *Estimating heterogeneous treatment effects with right-censored data via causal survival forests*. arXiv 2020;2001.09887.
43. Zhu J, Gallego B. Targeted estimation of heterogeneous treatment effect in observational survival analysis. *J Biomed Inform*. 2020;107:103474.
44. Zhao Y, Fang X, Simchi-Levi D. Uplift modeling with multiple treatments and general response types. In: *Proceedings of the 2017 SIAM International Conference on Data Mining*; April 27-29, 2017; Houston, TX.
45. Olaya D, Coussement K, Verbeke W. A survey and benchmarking study of multitreatment uplift modeling. *Data Min Knowl Disc*. 2020;34(2):273-308.
46. Lechner M. *Modified causal forests for estimating heterogeneous causal effects*. arXiv 2018;1812.09487.
47. US Food and Drug Administration *Office of Generic Drugs*. <https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/office-generic-drugs>. Accessed June 3, 2021.
48. US Food and Drug Administration. *Product-specific guidances for generic drug development*. <https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm075207.htm>. Accessed June 3, 2021.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of the article at the publisher's website.

How to cite this article: Gong X, Hu M, Basu M, Zhao L. Heterogeneous treatment effect analysis based on machine-learning methodology. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:1433–1443. doi:[10.1002/psp4.12715](https://doi.org/10.1002/psp4.12715)