

---

# Optimal individualized decision rules from a multi-arm trial: a comparison of methods and an application to tailoring inter-donation intervals among blood donors in the UK

Statistical Methods in Medical Research

XX(X):1–27

© The Author(s) 2019

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/ToBeAssigned

www.sagepub.com/

SAGE

Yuejia Xu<sup>1</sup>, Angela M. Wood<sup>2,3</sup>, Michael J. Sweeting<sup>4,2</sup>, David J. Roberts<sup>5,6</sup>, and Brian D. M. Tom<sup>1</sup>

## Abstract

There is a growing interest in precision medicine where individual heterogeneity is incorporated into decision-making and treatments are tailored to individuals to provide better healthcare. One important aspect of precision medicine is the estimation of the optimal individualized treatment rule (ITR) that optimizes the expected outcome. Most methods developed for this purpose are restricted to the setting with two treatments, while clinical studies with more than two treatments are common in practice. In this work, we summarize methods to estimate the optimal ITR in the multi-arm setting and compare their performance in large-scale clinical trials via simulation studies. We then illustrate their utilities with a case study using the data from the INTERVAL trial, which randomly assigned over 20000 male blood donors from England to one of the three inter-donation intervals (12-week, 10-week, and 8-week) over 2 years. We estimate the optimal individualized donation strategies under three different objectives. Our findings are fairly consistent across five different approaches that are applied: when we target the maximization of the total units of blood collected, almost all donors are assigned to the 8-week inter-donation interval, whereas if we aim at minimizing the low haemoglobin deferral rates, almost all donors are assigned to donate every 12 weeks. However, when the goal is to maximize the utility score that “discounts” the total units of blood collected by the incidences of low haemoglobin deferrals, we observe some heterogeneity in the optimal inter-donation interval across donors and the optimal donor assignment strategy is highly dependent on the trade-off parameter in the utility function.

## Keywords

Precision medicine, individualized treatment rule, multi-arm trial, utility function, blood donation.

---

<sup>1</sup>MRC Biostatistics Unit, University of Cambridge, UK

<sup>2</sup>Cardiovascular Epidemiology Unit, University of Cambridge, UK

<sup>3</sup>NIHR Blood and Transplant Research Unit in Donor Health and Genomics, UK

<sup>4</sup>Department of Health Sciences, University of Leicester, UK

<sup>5</sup>BRC Haematology Theme and Radcliffe Department of Medicine, University of Oxford, UK

<sup>6</sup>National Health Service Blood and Transplant, UK

## Corresponding author:

Yuejia Xu, MRC Biostatistics Unit, University of Cambridge, Robinson Way, Cambridge, CB2 0SR, UK

Email: yuejia.xu@mrc-bsu.cam.ac.uk

## 1 Introduction

Precision medicine is a rapidly expanding field in the new era of healthcare with major advancements having been made in technologies for collecting patient-level data and better characterizing each individual patient. The goal of precision medicine is to improve patient outcomes by tailoring treatment selection based on observed patient characteristics, for example, demographic information, clinical and laboratory measurements, medical history, and genetic data. In clinical practice, it is well-recognized that responses to treatments can vary substantially due to patient heterogeneity. Therefore, the treatment that is regarded as the best for a patient with one set of characteristics might not be the best for another, and the traditional “one-size-fits-all” approach does not lead to optimal clinical outcomes in many cases. In light of this, individualized, evidence-based clinical decision-making strategies that account for such heterogeneity are considered more desirable and are gaining much popularity in medical research.

One important aspect of precision medicine is the estimation of the optimal individualized treatment rule (ITR) by mapping patient information onto the set of treatment options. There has been a considerable amount of literature focusing on this matter. We refer readers to Lipkovich et al.<sup>1</sup> and Kosorok and Laber<sup>2</sup> for a comprehensive review of existing methods.

Clinical trials with multiple treatment arms are common in practice.<sup>3</sup> However, most existing statistical methods for estimating the optimal ITR were developed for binary treatment arm settings, and methods that directly aggregate multiple pairwise comparison results such as one-versus-rest or one-versus-one can be problematic: there are situations where the two-way preferences may be non-transitive and thus a consensus on the final decision of the original multi-category comparison problem cannot be reached. In addition, this may result in suboptimal decisions (simulated examples demonstrating the problem can be found in Lou et al.<sup>4</sup> and Zhou et al.<sup>5</sup>). For example, in the context of multi-class classification, it has been shown that the optimal solution obtained from the one-versus-rest approach can be different from the Bayes decision rule when there is no dominating class.<sup>6</sup> In this paper, we describe several methodological options that can be used to identify the optimal ITR in clinical trials with more than two treatment arms and are computationally feasible for large-scale trials, including  $l_1$ -penalized least squares,<sup>7</sup> adaptive contrast weighted learning,<sup>8</sup> direct learning,<sup>9</sup> and a Bayesian approach that is based on Bayesian additive regression trees.<sup>10</sup> Our motivating example, the INTERVAL trial,<sup>11,12</sup> had more than 20000 observations on each gender, and we are in particular interested in the performance of these methods on large datasets similar in size to the INTERVAL trial. However, to our knowledge, there is a lack of studies evaluating their performance in datasets with sample sizes larger than 2000. To fill this gap, we compare the aforementioned multi-arm ITR estimation methods in large-scale trials via simulation studies. We then conduct a case study, applying these methods to the data from the INTERVAL trial to estimate the optimal inter-donation interval among three possible options (12-week, 10-week, and 8-week) for male donors. We estimate the proportion of blood donors allocated to each inter-donation interval and quantify the gain (or loss) in outcomes when assigning donors according to the optimal ITRs inferred using these methods under three different targets: (1) maximizing the total units of whole blood collected by the blood service, (2) minimizing the risk of deferral for low haemoglobin (Hb), and (3) maximizing a utility score that seeks to balance the amount of blood collected against the number of deferrals for low Hb. We note from the INTERVAL trial that only considering maximizing the blood collection (target 1) or minimizing deferrals for low Hb (target 2) results in donor assignment strategies that are not particularly surprising and

quite “polar” in nature.<sup>12</sup> The purpose of investigating these two targets by applying precision medicine-based approaches is to examine in extreme cases where the universal rule (i.e., non-personalized strategy) should lead to nearly optimal outcomes, whether ITRs estimated using precision medicine-based methods are “almost non-personalized” and assign almost all donors to the marginally best inter-donation interval option, or instead “falsely” discover a fair amount of heterogeneity in the optimal inter-donation interval across different donors.<sup>13</sup> To our knowledge, this aspect has been rarely explored in the area of precision medicine as most researchers begin by assuming that patient subgroups exist and different patients would benefit from different treatment strategies. For male blood donors in the INTERVAL trial, results suggest that different statistical methods to estimate the optimal ITR lead to fairly similar recommendations on the optimal donation strategies: optimal ITRs estimated under the first two targets are very close to “one-size-fits-all” strategies, while optimal ITRs estimated under the third target suggest some heterogeneity in different blood donors’ optimal inter-donation intervals. In addition, the optimal ITR is highly dependent on how the utility score is constructed, i.e., the trade-off between the blood collection and the deferrals for low Hb.

The rest of the paper is organized as follows. In Section 2, we introduce the INTERVAL trial. In Section 3, we discuss the statistical framework and methods for estimating the optimal ITR when the treatment is binary. In Section 4, we review a selection of methods for estimating the optimal ITR when more than two treatment options are available. In Section 5, we conduct simulation studies to evaluate the performance of the multi-arm ITR estimation methods described in Section 4 under different scenarios. Section 6 presents results for an application of these methods to the INTERVAL trial dataset. We conclude with the discussion in Section 7.

## 2 Motivating example: The INTERVAL trial

Limits on the frequency of whole blood donation exist primarily to safeguard donor health. However, there is substantial variation across blood services in the maximum frequency of donations allowed. The National Health Service Blood and Transplant (NHSBT) in England currently allows a minimum inter-donation interval of 12 weeks for males and 16 weeks for females, with shorter inter-donation intervals used in other countries (e.g. the USA and France).<sup>14–16</sup> INTERVAL was the first randomized trial to evaluate the efficiency and safety of varying the frequency of whole blood donation. It randomly assigned over 45000 blood donors recruited across England to different inter-donation intervals (8, 10, and 12 weeks for men, and 12, 14, and 16 weeks for women) over a period of 2 years. During that time, there was a substantial increase in the amount of blood collected by reducing the inter-donation intervals without any detectable effects on overall quality of life, physical activity, or cognitive function.<sup>11,12,17</sup> However, increased donation frequency resulted in a greater number of deferrals (temporary suspension of donors from giving blood) due to low Hb, lower average Hb and ferritin concentrations, and more self-reported symptoms.

INTERVAL trial participants were well-characterized at baseline,<sup>17</sup> providing an opportunity to further explore personalized donation strategies whereby the amount of blood collected is optimized while controlling for the number of deferrals for low Hb. We expect differential relationships between the amount of blood collected and the number of deferrals for low Hb by donors’ individual characteristics. For example, young, female donors with low body mass index may be less able to donate more frequently without a rise in the number of deferrals for low Hb, whereas donors with a long, successful donation history may be able to donate more frequently with no increase in the number of deferrals for low Hb.<sup>11</sup> We compare the personalized

donation strategies against assignments of the same inter-donation interval for all donors (current clinical practice).

### 3 Preliminaries

#### 3.1 Notations and statistical frameworks

We consider a clinical trial where  $n$  subjects are sampled from a population of interest. Let  $Y$  be the outcome of interest,  $A \in \mathcal{A} = \{1, \dots, K\}$  denote the treatment assignment, and  $\mathbf{X} \in \mathcal{X}$  be a  $p$ -dimensional covariate (feature) vector. Without loss of generality, we assume a larger value of  $Y$  is preferred. We observe the triplet  $(Y_i, \mathbf{X}_i, A_i)$ , for  $i = 1, \dots, n$ , which are independent and identically distributed across  $i$ . The individualized treatment rule (ITR), denoted by  $\mathcal{D}$ , is a map from the space of feature variables,  $\mathcal{X}$ , to the domain of treatment assignments,  $\mathcal{A}$ . We assume the propensity score  $P(A = a|\mathbf{X})$  is bounded strictly away from 0, i.e.,  $P(A = a|\mathbf{X} = \mathbf{x}) > 0$  for all pairs  $(\mathbf{x}, a) \in \mathcal{X} \times \mathcal{A}$ . In clinical trials, the no unmeasured confounders assumption is always satisfied (for the intention-to-treat analysis).<sup>18</sup> Assuming that the consistency assumption holds,<sup>19</sup> the “value function” associated with the treatment rule  $\mathcal{D}$  can be derived as:

$$V(\mathcal{D}) = E\left\{\frac{I(A = \mathcal{D}(\mathbf{X}))}{P(A|\mathbf{X})}Y\right\}, \quad (1)$$

where  $E(\cdot)$  is the expectation and  $I(\cdot)$  is the indicator function.<sup>7</sup> The value function  $V(\mathcal{D})$  is the expected value of the outcome  $Y$  had the regime  $\mathcal{D}$  been applied to the given population. Our goal is to find the optimal ITR,  $\mathcal{D}^*$ , that maximizes the value function under  $\mathcal{D}$ , i.e.,

$$\mathcal{D}^* = \arg \max_{\mathcal{D}} V(\mathcal{D}) = \arg \max_{\mathcal{D}} E\left\{\frac{I(A = \mathcal{D}(\mathbf{X}))}{P(A|\mathbf{X})}Y\right\}. \quad (2)$$

#### 3.2 Estimation of the optimal ITR when $K = 2$

Many statistical methods have been proposed to estimate the optimal ITR in the case with two treatments ( $K = 2$ ), including indirect and direct methods.

Q-learning and A-learning are two main examples of indirect estimation methods, where the estimation of the optimal regime relies on modelling conditional mean outcomes (Q-learning) or modelling treatment contrasts (A-learning). The optimal ITR is then inferred (indirectly) based on the predicted means from the posited model. In the Q-learning paradigm, a model including treatment-covariate interactions is postulated for  $E(Y|\mathbf{X}, A = a)$ . It can be shown from (2) that  $\mathcal{D}^*(\mathbf{x}) = \arg \max_a E(Y|\mathbf{X} = \mathbf{x}, A = a)$ , and thus the decision follows naturally by maximizing the outcome from the fitted model, i.e.,  $\widehat{\mathcal{D}}^*(\mathbf{x}) = \arg \max_a \widehat{E}(Y|\mathbf{X} = \mathbf{x}, A = a)$ .<sup>20</sup> Alternatively, A-learning methods posit a model on the contrast function  $C(\mathbf{x})$  between two treatment options, and the final decision is based on the sign of the estimated contrast  $\widehat{C}(\mathbf{x})$ .<sup>21</sup> Compared to Q-learning, A-learning avoids the modelling of marginal covariate effects. Other indirect methods include G-estimation,<sup>22</sup> regret regression,<sup>23</sup> and weighted ordinary least squares.<sup>24</sup> Those indirect approaches focus on building a prediction model for either the conditional mean outcome or the treatment contrast and they target prediction accuracy instead of the direct maximization of the value function. We also note that methods such

as G-estimation and weighted ordinary least squares are doubly-robust. In randomized trials, the treatment assignment model is known by design, and thus these methods can be preferable to Q-learning as they offer protection against the misspecification of the outcome model.<sup>24,25</sup>

Alternatively, Zhao et al.<sup>26</sup> proposed a direct approach, the outcome weighted learning (OWL), to estimate the optimal ITR (directly) by maximizing the value function  $V(\mathcal{D})$  rather than (indirectly) by inverting estimates from prediction models. Since (2) indicates that  $\mathcal{D}^* = \arg \min_{\mathcal{D}} E\{I(A \neq \mathcal{D}(\mathbf{X}))Y/P(A|\mathbf{X})\}$ , Zhao et al.<sup>26</sup> reformulated the problem of maximizing the value function into a weighted classification problem (with the aim of minimizing the weighted misclassification error), where they classified  $A$  based on  $\mathbf{X}$  and weighted the misclassification by  $Y/P(A|\mathbf{X})$ . This could then be solved using support vector machines (SVM). To further improve the performance of OWL, Zhou et al.<sup>27</sup> and Liu et al.<sup>28</sup> introduced residual weighted learning (RWL) and augmented outcome weighted learning (AOL) respectively by using different weights in the objective function.

Most of the aforementioned methods, except for Q-learning, were developed for the case of two treatments. A direct extension to the multi-arm setting by combining pairwise decision rules using methods such as one-versus-rest or one-versus-one may lead to suboptimal results.<sup>5,6</sup>

## 4 Review of a selection of methods for $K > 2$

Recently, some new methods have been proposed to deal with the case of more than two treatments. In this section, we review (in publication order) a selection of approaches that can be used to estimate the optimal ITR in multi-arm trials and scale well for large datasets similar in size to the INTERVAL trial, including  $l_1$ -penalized least squares,<sup>7</sup> adaptive contrast weighted learning,<sup>8</sup> direct learning,<sup>9</sup> and a Bayesian approach that is based on Bayesian additive regression trees.<sup>10</sup> We note that there are other recently-proposed methods for learning the optimal ITR with multiple treatment options that we do not discuss in detail in this paper, for example, the multi-treatment outcome weighted learning,<sup>29</sup> and the sequential outcome weighted multicategory learning,<sup>5</sup> since they involve solving weighted support vector machine problems, which can impose high computational costs when applied to large datasets.<sup>30</sup>

### 4.1 $l_1$ -penalized least squares (indirect method)

Qian and Murphy<sup>7</sup> proposed the two-stage model-based  $l_1$ -penalized least squares ( $l_1$ -PLS) method. The conditional mean response was first estimated using  $l_1$ -PLS with a rich linear model to guard against over-fitting and then the estimated means under different treatments were used to derive the optimal ITR. Specifically, we fit the model for  $E(Y|\mathbf{X}, A)$  using the basis function  $(1, \mathbf{X}, A, \mathbf{X}A)$ , in which we use  $K - 1$  dummy variables to replace  $A$ . The decision can be derived as  $\widehat{\mathcal{D}}^*(\mathbf{x}) = \arg \max_{a \in \{1, \dots, K\}} \widehat{E}(Y|\mathbf{X} = \mathbf{x}, A = a)$ .

$\mathbf{X}$  in the INTERVAL trial dataset contains several nominal categorical covariates. In order to avoid the over-selection of variables with many categories and ensure that all dummy variables encoding the same categorical covariate are either included or excluded from the model simultaneously, we use group LASSO (GL) which encourages sparsity at the factor level for variable selection.<sup>31,32</sup>

Another variable selection approach that is of interest in this context is hierarchical group LASSO (HGL). HGL was designed for learning pairwise interactions in regression models that satisfy strong hierarchy, i.e.,

if the coefficient associated with the interaction term is estimated to be non-zero, then its two associated main effects also have non-zero estimated coefficients. The assumption underlying HGL is that covariates that predict treatment effect heterogeneity (included as interactions) are also prognostic (included in main effects).

As has been pointed out by Gunter et al.,<sup>32</sup> maintaining hierarchy in variable selection avoids finding spurious interactions that may appear due to the exclusion of important main effect terms. The HGL method proposed by Lim and Hastie<sup>33</sup> picks out only main effects if the truth has no interactions, but when interactions truly exist, it allows the discovery of important interaction terms despite their weak effects by regularizing coefficients of main effects using the “glinternet” penalty. We can embed HGL in the first step of  $l_1$ -PLS to identify important treatment-covariate interactions. For ease of interpretation, we restrict the search space to be all possible two-way interactions between treatment  $A$  and covariates  $\mathbf{X}$ .

As an aside, there are two types of interactions: quantitative interactions and qualitative interactions.<sup>1</sup> A quantitative interaction between  $A$  and  $\mathbf{X}$  refers to the situation where the magnitude of the effect of  $A$  on the outcome  $Y$  depends on  $\mathbf{X}$  but the direction of this effect is consistent for all possible values of  $\mathbf{X}$ , whereas a qualitative interaction between  $A$  and  $\mathbf{X}$  indicates that both the direction and the magnitude of  $A$ 's effect on the outcome  $Y$  can depend on  $\mathbf{X}$  (Supplementary Figure 1 in the online supplemental materials Appendix A). GL and HGL interaction selection procedures are not able to automatically distinguish between qualitative and quantitative interactions. Therefore, further assessment of interactions (e.g. using the Gail-Simon test<sup>34</sup>) is required to determine whether the selected interactions are useful for achieving specific aims (e.g. the aim can either be to detect a qualitative difference in treatment effects, i.e., crossover treatment effect heterogeneity, or to detect a quantitative difference in treatment effects that informs the identification of subgroups with elevated response to a given treatment, i.e., non-crossover treatment effect heterogeneity) in the context of precision medicine.

One limitation of  $l_1$ -PLS is that the correct inference on the optimal ITR depends on a correctly specified outcome model, while the postulated outcome model is prone to misspecification (especially when  $\mathbf{X}$  is high-dimensional) and thus may result in suboptimal ITR. Unlike tree-based methods, which can capture more general forms of interactions,  $l_1$ -PLS is restricted to searching for additive-type interactions. However, the true relationship between covariates, treatment assignment, and the outcome is usually complicated in practice, and the underlying interaction structures can be much more flexible than additive. Another issue with this approach is the interference between main effects and treatment-covariate interactions, which may impair the search for the optimal ITR. In most cases, the variability in the outcome is predominantly explained by main effects rather than interactions. Therefore, interaction effects can be overlooked due to their small predictive ability when the method focuses on prediction.<sup>35–37</sup>

#### 4.2 Adaptive contrast weighted learning (direct method)

Adaptive contrast weighted learning (ACWL) was proposed by Tao and Wang<sup>8</sup> to estimate the optimal ITR in the multi-arm setting. They constructed doubly robust semi-parametric regression-based contrasts<sup>38</sup> with the adaptation of treatment effect orderings, and the adaptive contrasts simplified the optimization problem with multiple treatment comparisons into a weighted classification problem.

Let  $\mu_a(\mathbf{X})$  denote the conditional mean  $E(Y|\mathbf{X}, A = a)$ . Tao and Wang,<sup>8</sup> following Zhang et al.,<sup>38</sup> employed the doubly robust augmented inverse probability weighted estimator (AIPWE) for  $\mu_a(\mathbf{X})$ . Let  $\mu_{(1)}(\mathbf{X}) \leq \dots \leq \mu_{(K)}(\mathbf{X})$  be the order statistics of  $\mu_1(\mathbf{X}), \dots, \mu_K(\mathbf{X})$ , and  $l_a$  be the treatment effect order

with  $\mu_{(a)}(\mathbf{X}) = \mu_{l_a}(\mathbf{X})$ . It follows that

$$\begin{aligned}\mathcal{D}^*(\mathbf{X}) &= \arg \max_{\mathcal{D}} \sum_{a=1}^K E \left[ \mu_{(a)}(\mathbf{X}) I\{\mathcal{D}(\mathbf{X}) = l_a(\mathbf{X})\} \right] \\ &= \arg \min_{\mathcal{D}} \sum_{a=1}^{K-1} E \left[ \{\mu_{(K)}(\mathbf{X}) - \mu_{(a)}(\mathbf{X})\} I\{\mathcal{D}(\mathbf{X}) = l_a(\mathbf{X})\} \right].\end{aligned}\quad (3)$$

The optimal ITR derived from (3) can be interpreted as the rule that minimizes the expected loss in the outcome due to suboptimal treatments in the entire population of interest, i.e., the ITR classifies as many subjects as possible to their corresponding optimal treatment  $l_K$  and puts more “penalties” on subjects with larger contrasts. In practice, it can be challenging to utilize all  $K - 1$  contrasts as weights to classify treatment. Tao and Wang<sup>8</sup> addressed this by constructing the lower and upper bounds of the expected loss in the outcome due to suboptimal treatments as follows:

$$\begin{aligned}\sum_{a=1}^{K-1} E \left[ \{\mu_{(K)}(\mathbf{X}) - \mu_{(a)}(\mathbf{X})\} I\{\mathcal{D}(\mathbf{X}) = l_a(\mathbf{X})\} \right] &\geq \sum_{a=1}^{K-1} E \left[ \{\mu_{(K)}(\mathbf{X}) - \mu_{(K-1)}(\mathbf{X})\} I\{\mathcal{D}(\mathbf{X}) = l_a(\mathbf{X})\} \right] \\ &= E \left[ C_1(\mathbf{X}) I\{\mathcal{D}(\mathbf{X}) \neq \underbrace{l_K(\mathbf{X})}_{\text{optimal}}\} \right],\end{aligned}\quad (4)$$

and

$$\begin{aligned}\sum_{a=1}^{K-1} E \left[ \{\mu_{(K)}(\mathbf{X}) - \mu_{(a)}(\mathbf{X})\} I\{\mathcal{D}(\mathbf{X}) = l_a(\mathbf{X})\} \right] &\leq \sum_{a=1}^{K-1} E \left[ \{\mu_{(K)}(\mathbf{X}) - \mu_{(1)}(\mathbf{X})\} I\{\mathcal{D}(\mathbf{X}) = l_a(\mathbf{X})\} \right] \\ &= E \left[ C_2(\mathbf{X}) I\{\mathcal{D}(\mathbf{X}) \neq \underbrace{l_K(\mathbf{X})}_{\text{optimal}}\} \right],\end{aligned}\quad (5)$$

where  $C_1(\mathbf{X}) = \mu_{(K)}(\mathbf{X}) - \mu_{(K-1)}(\mathbf{X})$  and  $C_2(\mathbf{X}) = \mu_{(K)}(\mathbf{X}) - \mu_{(1)}(\mathbf{X})$ . In the least conservative case where suboptimal treatments only lead to the minimal expected loss in the outcome,

$$\mathcal{D}^*(\mathbf{X}) = \arg \min_{\mathcal{D}} E \left[ C_1(\mathbf{X}) I\{\mathcal{D}(\mathbf{X}) \neq l_K(\mathbf{X})\} \right], \quad (6)$$

while in the most conservative case where suboptimal treatments lead to the maximal expected loss in the outcome,

$$\mathcal{D}^*(\mathbf{X}) = \arg \min_{\mathcal{D}} E \left[ C_2(\mathbf{X}) I\{\mathcal{D}(\mathbf{X}) \neq l_K(\mathbf{X})\} \right]. \quad (7)$$

The optimal ITR estimated using (6) or (7) minimizes the lower or upper bounds of the expected loss in the outcome due to suboptimal treatments over the entire population, respectively. By optimizing these bounds, the optimization problem with multiple treatment comparisons is simplified to a weighted classification problem, which can be solved by classification techniques, such as classification and regression trees (CART).

ACWL is flexible and robust. However, additional uncertainties may be induced through the introduction of the “ordering labels”. Therefore, using adaptive contrasts may not be the most efficient way to avoid the multiple treatment comparison issue.

### 4.3 Direct learning (direct method)

Qi and Liu<sup>9</sup> developed the direct learning (D-learning) approach that uses regression methods to directly estimate the optimal ITR. The advantage of D-learning over Q-learning is that D-learning directly targets the maximization of the value function, thus avoiding the need to model main effects.

In the binary arm case (treatment  $A$  is encoded as  $-1$  or  $1$ ),

$$\begin{aligned}\mathcal{D}^*(\mathbf{X}) &= \arg \max_{\mathcal{D}} V(\mathcal{D}) \\ &= \text{sign}[E(Y|\mathbf{X}, A = 1) - E(Y|\mathbf{X}, A = -1)] \\ &= \text{sign}\left[E\left\{\frac{YA}{P(A|\mathbf{X})}\middle|\mathbf{X}\right\}\right] \\ &:= \text{sign}\{f^*(\mathbf{X})\},\end{aligned}\tag{8}$$

where  $f^*(\mathbf{X})$  is the optimal decision function.<sup>9</sup> According to (8),  $f^*(\mathbf{X})$  can be estimated using the fact that  $f^*(\mathbf{X}) = E\left\{\frac{YA}{P(A|\mathbf{X})}\middle|\mathbf{X}\right\}$  and Qi and Liu<sup>9</sup> showed that  $f^*(\mathbf{X}) \in \arg \min_f E[\{2YA - f(\mathbf{X})\}^2/P(A|\mathbf{X})]$ . We can estimate  $f^*(\mathbf{X})$  using regression methods for either linear or nonlinear decision rules. For example, suppose the decision function is linear and  $f(\mathbf{X}) = \mathbf{X}^T\beta$ . Then  $\beta$  can be estimated using the ordinary weighted least squares (or LASSO in the high-dimensional case). The estimated optimal linear decision function is  $\hat{f}^*(\mathbf{X}) = \mathbf{X}^T\hat{\beta}$ , and the estimated optimal ITR  $\hat{\mathcal{D}}^*(\mathbf{X})$  is the sign of  $\hat{f}^*(\mathbf{X})$ .

D-learning was extended to the case with more than two treatments. Qi and Liu<sup>9</sup> showed that the optimal ITR in the multi-arm setting can be written as

$$\begin{aligned}\mathcal{D}^*(\mathbf{X}) &= \arg \max_{a \in \{1, \dots, K\}} \sum_{i \neq a}^K E\left\{\frac{YA_{ai}}{P(A_{ai}|\mathbf{X})}\middle|\mathbf{X}, A = a \text{ or } i\right\} \\ &:= \arg \max_{a \in \{1, \dots, K\}} \sum_{i \neq a}^K f_{ai}^*(\mathbf{X}),\end{aligned}\tag{9}$$

where  $A_{ai} \in \{-1, 1\}$  denotes a binary treatment indicator for whether a patient is on treatment  $i$  ( $A_{ai} = -1$ ) or on treatment  $a$  ( $A_{ai} = 1$ ) for  $i \neq a$ , and  $f_{ai}^*(\mathbf{X})$  denotes the optimal decision function between treatment  $a$  and  $i$ . Each pairwise decision function  $f_{ai}^*(\mathbf{X})$ , for  $a, i = 1, \dots, K, i \neq a$ , can be estimated using regression methods as in the binary setting.

We note that even though multi-category D-learning builds on multiple pairwise comparisons, it is different from the one-versus-one approach<sup>39</sup> in that one-versus-one aggregates multiple pairwise decisions based on the majority voting rule, whereas multi-category D-learning measures the effect of each treatment based on the sum of pairwise contrasts and then picks the treatment with the largest effect measure.



#### 4.4 Bayesian additive regression trees (indirect method)

Bayesian additive regression trees (BART) is a flexible nonparametric prediction model that was first introduced by Chipman et al.<sup>40</sup> and has gained much popularity with numerous applications in recent years.<sup>41,42</sup> The BART model is an additive ensemble of many single regression trees with each tree explaining a small portion of the outcome, and it can accommodate nonlinear main effects as well as complex interaction effects without the need to specify their functional forms. We refer readers to Tan and Roy<sup>41</sup> and Hill et al.<sup>42</sup> for a walk-through of the details on BART.

Logan et al.<sup>10</sup> incorporated the idea of BART into the estimation of the optimal ITR by using BART to model the dependency structure of the response and covariates. In the first stage, a BART model for  $E(Y|\mathbf{X}, A)$  was built using the treatment and all covariates as input variables. In the second stage, the optimal arm was chosen as the one that maximizes  $E(Y|\mathbf{X} = \mathbf{x}, A = a)$ , which can be approximated by the average over posterior draws. Under this framework, the posterior distribution of the value function of an ITR can be obtained using posterior samples straightforwardly, and uncertainties about the value of an ITR can be captured by posterior samples from the prediction model directly.

One potential issue of this method is that the prediction model constructed with BART is “black-box”, which impairs the interpretability of the model and the decision rules. In the two-arm setting, Logan et al.<sup>10</sup> proposed a way to facilitate the interpretation by fitting the posterior mean differences into a tree structure. However, the extension to the multi-arm case is not straightforward given that there are multiple pairwise contrasts when more than two treatment options are available. In addition, BART is more computationally demanding than the aforementioned non-Bayesian approaches due to its reliance on Markov chain Monte Carlo (MCMC).

## 5 Simulation studies

To our knowledge, there is a lack of studies comparing and reporting the performance of different multi-arm ITR estimation methods under scenarios where the training sample size is larger than 2000. It is reasonable to expect that the relative performance of different methods may vary with the training sample size, and thus findings in previous works may not apply to the INTERVAL trial, where the number of male/female participants is more than 20000. In addition, the BART ITR estimation method<sup>10</sup> discussed in Section 4.4 has not been compared with other methods in multi-arm trial settings before. In this section, we conduct simulation studies to compare the performance of the following five methods in large samples:

- (i)  $l_1$ -PLS-HGL:  $l_1$ -penalized least squares with hierarchical group LASSO variable selection using the basis function  $(1, \mathbf{X}, A, \mathbf{X}A)$ .<sup>7,33</sup>
- (ii)  $l_1$ -PLS-GL:  $l_1$ -penalized least squares with group LASSO variable selection using the basis function  $(1, \mathbf{X}, A, \mathbf{X}A)$ .<sup>7,31</sup>
- (iii) ACWL: Adaptive contrast weighted learning using (7), where suboptimal inter-donation intervals lead to maximal expected loss in the outcome (i.e., ACWL-C2).<sup>8</sup> We only present the results from ACWL-C2, since we observe that adaptive contrast weighted learning using (6) (i.e., ACWL-C1) gives very similar results in simulation studies, as has also been noted in the original publication.<sup>8</sup>
- (iv) D-learning: Direct learning with linear decision functions.<sup>9</sup>

- (v) BART: Bayesian additive regression trees with default prior parameters as specified in the R package `BART`.<sup>10,40</sup>

We consider the training sample size being  $n = 20000$  (similar in size to the number of male/female donors in the INTERVAL trial). We simulate data under 6 settings with different types of covariates and forms of treatment-covariate interaction effects. In each simulation setting, we generate 5 covariates  $X_1, \dots, X_5$  independently. Treatment  $A$  is sampled uniformly from  $\{1, 2, 3\}$ . We assume the outcome  $Y$  is normally distributed with mean  $m(\mathbf{X}) + \Delta(\mathbf{X}, A)$  and variance  $\sigma^2 = 1$ , where  $m(\mathbf{X})$  is the main effect of covariates on the outcome and  $\Delta(\mathbf{X}, A)$  denotes the treatment-covariate interaction effect. We consider  $m(\mathbf{X}) = 1 + 0.5X_4 + 0.3X_5$  and  $\Delta(\mathbf{X}, A) = 0.5\{I(A = 1)\Delta_1(\mathbf{X}) + I(A = 2)\Delta_2(\mathbf{X}) + I(A = 3)\Delta_3(\mathbf{X})\}$ , with  $\Delta_1(\mathbf{X})$ ,  $\Delta_2(\mathbf{X})$ , and  $\Delta_3(\mathbf{X})$  taking on different functional forms. Details on distributions of  $X_1, \dots, X_5$  (we use  $\mathcal{U}\{a, b\}$  to denote the continuous uniform distribution that takes values in the range  $[a, b]$ , and  $\text{Bern}(p)$  to denote the Bernoulli distribution with the success probability being  $p$ ) and expressions for  $\Delta_1(\mathbf{X})$ ,  $\Delta_2(\mathbf{X})$ , and  $\Delta_3(\mathbf{X})$  in settings 1-6 are provided in Table 1. Settings 1 and 2 consider tree-type and linear interaction effects, respectively. True underlying decision boundaries are nonlinear in settings 3 and 4, with setting 3 including a between-covariate interaction. Settings 5 contains discrete covariates (one categorical covariate and one binary covariate). We examine the scenario where the true optimal treatment is the same for all individuals in settings 6 (this scenario mimics the situation in the INTERVAL trial when we target two outcomes separately, i.e., objectives (1) and (2) described in Section 1: in this case, we would expect the donor assignment strategy to be quite “polar” and the non-personalized rule to yield nearly optimal outcomes based on the primary analysis results presented in Di Angelantonio et al.<sup>12</sup>). We consider additional simulation settings with correlated (Appendix B.1) and higher-dimensional (Appendix B.2) covariates in the online supplement.

We evaluate the performance of different methods on a large independent testing dataset of size 10000 using two criteria: (a) the misclassification rate of the estimated optimal ITR compared to the true optimal ITR, and (b) the value function under the estimated ITR, which can be calculated according to (1). Smaller misclassification rates and larger value functions indicate better performance. Each simulation is repeated 100 times and all tuning parameters are selected via 5-fold cross-validation. We report both the mean and the standard deviation of misclassification rates and value functions across 100 replicates. We note that the BART ITR estimation method is different from the other four non-Bayesian methods by nature, and the uncertainty of the value function can be directly quantified under the Bayesian framework (based on posterior samples), as has been discussed in Section 4.4. However, to make different methods comparable, the standard deviation estimates of value functions associated with the BART ITR are calculated as the standard deviation across 100 runs rather than the posterior standard deviation.

Simulation results are presented in Table 1. As expected, when the true underlying decision boundaries are tree-type (setting 1), ACWL that builds on decision trees performs better than  $l_1$ -PLS-HGL,  $l_1$ -PLS-GL, and D-learning. BART leads to slightly smaller misclassification rate than ACWL in this setting. In contrast, when decision boundaries are linear (setting 2),  $l_1$ -PLS-HGL,  $l_1$ -PLS-GL, and D-learning perform similarly well and much better than ACWL. BART is superior to ACWL but slightly worse than the other three methods in this case, possibly due to its “over-parameterization” for linear effects. None of these methods manage to capture the nonlinear structures in settings 3 and 4 properly, and misclassification rates are not as low as in other settings for all methods despite the large training sample size. However, BART outperforms the rest of

the methods to a large extent. This is not surprising given the flexibility of BART. All methods achieve good performance when some covariates are discrete (setting 5). Results for setting 6 implies that when the true optimal treatment is the same for all subjects (“trivial” decision rule that assigns all to the marginally best treatment), all methods perform perfectly with no misclassification. We also conduct simulations for some variations of setting 6 (e.g. increased noise) and find similar results (online supplement Appendix B.3).

Overall, our simulation studies suggest that unless we have *a priori* information on the type of underlying interaction effects (e.g. linear additive or tree-type), the BART multi-arm ITR estimation method should be a good choice in general since it is robust and performs better than or comparable to the other competing methods regardless of the functional form of interaction terms. We note that BART takes longer to run than other non-Bayesian methods, especially in the large sample size case that we examine in this paper (on a Windows-based computing system with 1 core, 3.40 GHz Intel processor, BART takes about 200 seconds per run while other methods take less than 20 seconds per run on our simulated datasets with 20000 observations). Parallel processing with multi-threading can be used to speed up the computation of BART.<sup>43</sup> We also note that in settings with nonlinear decision boundaries, the performance of  $l_1$ -PLS and D-learning may be improved if we use polynomials of higher degrees as basis functions. For D-learning, we would also expect the nonlinear version which estimates the optimal decision function using the component selection and smoothing operator (COSSO) to perform better than the linear version considered in this paper if the underlying decision boundaries are nonlinear. However, our numerical experiments suggest that nonlinear D-learning with COSSO is much more computationally demanding and performs less well than BART when  $n = 20000$ .

R codes to implement these methods on a simulated example dataset are provided at [https://github.com/yux139/ITR\\_multiarm\\_casestudy](https://github.com/yux139/ITR_multiarm_casestudy).

## 6 Application to the INTERVAL trial

To illustrate the use of multi-arm ITR estimation methods on large-scale clinical trials, we apply the five methods compared in Section 5 to the data from male donors in the INTERVAL trial and estimate each blood donor’s optimal inter-donation interval. We also compare the personalized donation strategies with “one-size-fits-all” donor assignment rules that recommend the same inter-donation intervals of 8, 10 or 12 weeks for all male donors. After data processing (exclude donors who had zero attendance over the 2-year trial period), 20574 male blood donors are included in the analysis following the intention-to-treat principle according to donors’ randomized groups. We show in the online supplemental material (Appendix C) that the data cleaning process does not distort the balance of baseline covariates across randomized groups.

### 6.1 Outcomes of interest

We consider two outcomes, namely, the total units of blood collected by the blood service per donor over a 2-year period (the standard practice is to donate 1 unit of blood per session, with a full donation unit containing 470 ml of whole blood<sup>44</sup>), denoted by  $G$ , and the rate of low Hb deferrals per donor attendances during the same period (calculated as the total number of “at session” deferrals for low Hb divided by the total number of attendances in the 2-year trial period), denoted by  $R$ . We note that these two outcomes are not independent: assigning a donor to a more frequent inter-donation interval in principle will lead to an increase in the total units of blood collected. However, this increased frequency may have the opposite effect through increased risks of

**Table 1.** Description of simulation settings 1-6 and simulation results for  $n = 20000$  based on 100 replicates: mean (sd) of misclassification rates and value functions. Methods under comparison include the  $l_1$ -penalized least squares with hierarchical group LASSO variable selection ( $l_1$ -PLS-HGL),  $l_1$ -penalized least squares with group LASSO variable selection ( $l_1$ -PLS-GL), adaptive contrast weighted learning (ACWL), direct learning (D-learning), and Bayesian additive regression trees (BART). The smallest misclassification rates and the largest value functions for each setting are in bold.

Setting	Functional form of interaction	True optimal treatment	Covariates type	Method	Misclassification	Value
1	tree $\Delta_1(\mathbf{X}) = 4 \times I(X_1 > 0.5) - 2$ $\Delta_2(\mathbf{X}) = 2 \times I(X_2 \geq 0.5)I(X_3 < 0.25) - 1$ $\Delta_3(\mathbf{X}) = 0$	1 or 2 or 3	continuous $X_1, \dots, X_5 \sim \mathcal{U}\{-1, 1\}$	$l_1$ -PLS-HGL	0.101 (0.016)	1.226 (0.010)
				$l_1$ -PLS-GL	0.094 (0.015)	1.229 (0.010)
				ACWL	0.028 (0.040)	1.276 (0.023)
				D-learning	0.100 (0.020)	1.226 (0.013)
				BART	<b>0.010 (0.004)</b>	<b>1.286 (0.003)</b>
2	linear $\Delta_1(\mathbf{X}) = 3X_1 - 2X_2$ $\Delta_2(\mathbf{X}) = 5X_3 - X_4 + X_5 - 1$ $\Delta_3(\mathbf{X}) = 0$	1 or 2 or 3	continuous $X_1, \dots, X_5 \sim \mathcal{U}\{-1, 1\}$	$l_1$ -PLS-HGL	0.015 (0.004)	1.736 (0.003)
				$l_1$ -PLS-GL	<b>0.013 (0.004)</b>	<b>1.737 (0.003)</b>
				ACWL	0.171 (0.020)	1.662 (0.016)
				D-learning	0.018 (0.005)	<b>1.737 (0.004)</b>
				BART	0.056 (0.004)	1.730 (0.006)
3	nonlinear $\Delta_1(\mathbf{X}) = 3X_1^2 - \exp(X_2)$ $\Delta_2(\mathbf{X}) = X_3X_4$ $\Delta_3(\mathbf{X}) = 0$	1 or 2 or 3	continuous $X_1, \dots, X_5 \sim \mathcal{U}\{-1, 1\}$	$l_1$ -PLS-HGL	0.566 (0.012)	1.089 (0.008)
				$l_1$ -PLS-GL	0.565 (0.014)	1.088 (0.011)
				ACWL	0.561 (0.016)	1.089 (0.009)
				D-learning	0.572 (0.013)	1.087 (0.008)
				BART	<b>0.192 (0.038)</b>	<b>1.209 (0.010)</b>
4	nonlinear $\Delta_1(\mathbf{X}) = 3X_1^2 - \exp(X_2)$ $\Delta_2(\mathbf{X}) = X_3^3$ $\Delta_3(\mathbf{X}) = 0$	1 or 2 or 3	continuous $X_1, \dots, X_5 \sim \mathcal{U}\{-1, 1\}$	$l_1$ -PLS-HGL	0.350 (0.011)	1.129 (0.004)
				$l_1$ -PLS-GL	0.352 (0.010)	1.128 (0.004)
				ACWL	0.362 (0.011)	1.118 (0.004)
				D-learning	0.359 (0.016)	1.129 (0.004)
				BART	<b>0.163 (0.045)</b>	<b>1.220 (0.012)</b>
5	nonlinear $\Delta_1(\mathbf{X}) = 2\{I(X_1 = 1) + I(X_1 = 2)\}X_2 - 1$ $\Delta_2(\mathbf{X}) = 5I(X_1 = 5)X_3 - 2$ $\Delta_3(\mathbf{X}) = 0$	1 or 2 or 3	continuous + binary + categorical $X_1 \sim \text{discrete uniform}\{1, 5\}$ $X_2 \sim \text{Bern}(0.5)$ $X_3, X_4, X_5 \sim \mathcal{U}\{-1, 1\}$	$l_1$ -PLS-HGL	0.077 (0.019)	1.101 (0.012)
				$l_1$ -PLS-GL	0.078 (0.018)	1.101 (0.011)
				ACWL	0.029 (0.028)	1.129 (0.016)
				D-learning	0.090 (0.032)	1.094 (0.019)
				BART	<b>0.007 (0.005)</b>	<b>1.142 (0.002)</b>
6	tree $\Delta_1(\mathbf{X}) = I(X_1 > 0.5) + 2$ $\Delta_2(\mathbf{X}) = 2 \times I(X_2 \geq 0.5)I(X_3 < 0.25) - 3$ $\Delta_3(\mathbf{X}) = 0$	1 for everyone	continuous $X_1, \dots, X_5 \sim \mathcal{U}\{-1, 1\}$	$l_1$ -PLS-HGL	<b>0.000 (0.000)</b>	<b>2.093 (0.000)</b>
				$l_1$ -PLS-GL	<b>0.000 (0.000)</b>	<b>2.093 (0.000)</b>
				ACWL	<b>0.000 (0.000)</b>	<b>2.093 (0.000)</b>
				D-learning	<b>0.000 (0.000)</b>	<b>2.093 (0.000)</b>
				BART	<b>0.000 (0.000)</b>	<b>2.093 (0.000)</b>

deferrals for low Hb, which may consequently cause existing donors to come back less often and even to leave the donor register permanently.<sup>11,12,45</sup> Potential loss of donors may have a cost impact.<sup>45,46</sup> The current donor loss rate following a deferral for low Hb is 40-50%, and this would incur substantial costs (approximately £2.3 million in the worst-case scenario) for the blood service to recruit sufficient new donors and stabilize the donor base.<sup>47,48</sup> Therefore, when recommending the optimal inter-donation interval to a blood donor, there is a trade-off between the benefit and the risk: neither the optimal ITR solely based on the benefit nor that solely based on the risk may be acceptable, and it is generally not possible to find a strategy that optimizes both (maximizes benefit and minimizes risk) simultaneously. The goal of maximizing the total units of blood collected needs to be considered in conjunction with controlling for the low Hb deferrals. This motivates us to construct a single scalar “utility” outcome which *discounts* the units of blood collected by the increased incidences of low Hb deferrals as follows:

$$U = G - b \times \tilde{R}, \quad (10)$$

where  $G$  is the gain/benefit (total units of blood collected in the 2-year trial period),  $\tilde{R}$  is the risk (number of deferrals for low Hb in the 2-year trial period:  $\tilde{R} = R \times \text{total number of attendances over 2 years}$ ), and  $b$  is

the “trade-off” parameter reflecting the equivalent benefit loss for one unit increase in the risk. In the context of the INTERVAL trial,  $b$  is interpreted as “the equivalent loss in total units of blood collected by the blood service per donor over 2 years for one extra deferral for low Hb per donor attendances during the same period”. We examine different values for  $b$  within a range considered to be reasonable by NHSBT and vary  $b$  from 1 to 5 at an increment of 1 to see how results change with this parameter. A range of  $b$  from 1 to 5 covers the range of the extra costs of deferrals for low Hb considering reduced efficiency of collection, reduced donor retention and increase in recruitment of the many new donors to replace a regular donor who retires from donation.

We assume a larger outcome to be more desirable when we introduce the statistical framework in Section 3.1. This holds for the benefit and the utility outcome, but is not the case for the low Hb deferral rates. We address this issue by considering the maximization of “1 – low Hb deferral rates” instead, which is equivalent to the minimization of the low Hb deferral rates. In addition, the low Hb deferral rates is a proportion and so we use the arcsine square root transformation for variance stabilization.<sup>49</sup> This transformation is monotonically increasing, and thus rank-preserving.

## 6.2 Baseline covariates

Based on findings in the primary trial paper,<sup>12</sup> we include the following 19 variables measured at each donor’s baseline visit:

- Continuous: age, body mass index, Short Form Health Survey version 2 (SF-36v2) physical component score and mental component score, units of whole blood donations in the 2 years before enrollment into the trial, haemoglobin level, white blood cell count, red blood cell count, mean corpuscular haemoglobin, mean corpuscular volume, and platelet count.
- Categorical: ethnicity (Asian, Black, Mixed, White, Other, Unknown), blood group (A+, A–, AB+, AB–, B+, B–, O+, O–), iron prescription (Yes, No, Unknown), smoke ever (Yes, No, Unknown), smoke currently (Yes, No, Unknown), alcohol ever (Yes, No, Unknown), alcohol currently (Yes, No, Unknown), and new or returning donor status (New donor, Returning donor).

## 6.3 Evaluation criteria

We calculate proportions of donors assigned to each of the three inter-donation intervals according to the optimal ITR estimated using different methods. We are also interested in the quantity “ITR effect” or “benefit function”.<sup>50,51</sup> The ITR effect,  $\delta$ , associated with the rule  $\mathcal{D}(\mathbf{X})$  is defined as:

$$\delta(\mathcal{D}(\mathbf{X})) = E\{Y|\mathbf{X}, A = \mathcal{D}(\mathbf{X})\} - E\{Y|\mathbf{X}, A \neq \mathcal{D}(\mathbf{X})\}. \quad (11)$$

In the INTERVAL trial, the ITR effect can be empirically estimated as the difference in the average outcome between donors whose assigned inter-donation intervals in the trial are the same as  $\mathcal{D}(\mathbf{X})$  (i.e., average across all donors whose  $a = \mathcal{D}(\mathbf{x})$ ) and those whose assigned inter-donation intervals are different from  $\mathcal{D}(\mathbf{X})$  (i.e., average across all donors whose  $a \neq \mathcal{D}(\mathbf{x})$ ).

When two outcomes are analyzed separately, in addition to ITR effects of the estimated optimal assignment rule  $\widehat{\mathcal{D}}^*(\mathbf{X})$  on the outcome that we aim to optimize, we also calculate the effect of assigning donors according to  $\widehat{\mathcal{D}}^*(\mathbf{X})$  on the other outcome that we do not take into consideration when estimating  $\mathcal{D}^*(\mathbf{X})$ .

**Table 2.** Applications of the  $l_1$ -penalized least squares with hierarchical group LASSO variable selection ( $l_1$ -PLS-HGL),  $l_1$ -penalized least squares with group LASSO variable selection ( $l_1$ -PLS-GL), adaptive contrast weighted learning (ACWL), and direct learning (D-learning) to data from male donors in the INTERVAL trial. Means and standard deviations (in parenthesis) of assignment proportions in % and empirical ITR effects on donation and deferral outcomes across 100 repetitions of 5-fold cross-validation are reported. ITR effects measure the difference in the average outcome between donors whose assigned inter-donation intervals in the trial are optimal (with respect to the method used to estimate the ITR) and those whose assigned inter-donation intervals are non-optimal. A larger ITR effect on donation and a smaller ITR effect on deferral are more desirable. The first four and last four rows correspond to the target being maximizing total units of blood collected by the blood service, and minimizing the low Hb deferral rates, respectively.

Target Outcome	Method	Assignment Percentages			ITR Effects	
		12 weeks	10 weeks	8 weeks	Donation	Deferral
Donation	$l_1$ -PLS-HGL	0.1 (0.0)	0.3 (0.0)	99.6 (0.0)	1.308 (0.004)	0.026 (0.000)
	$l_1$ -PLS-GL	0.0 (0.0)	0.3 (0.3)	99.7 (0.3)	1.311 (0.005)	0.027 (0.000)
	ACWL	0.0 (0.0)	0.0 (0.0)	100.0 (0.0)	1.315 (0.000)	0.027 (0.000)
	D-learning	0.3 (0.1)	0.3 (0.2)	99.4 (0.2)	1.307 (0.006)	0.027 (0.000)
Deferral	$l_1$ -PLS-HGL	94.4 (0.6)	5.5 (0.6)	0.0 (0.0)	-1.188 (0.010)	-0.024 (0.000)
	$l_1$ -PLS-GL	99.7 (0.6)	0.2 (0.5)	0.0 (0.1)	-1.246 (0.006)	-0.024 (0.000)
	ACWL	99.7 (0.6)	0.3 (0.6)	0.0 (0.0)	-1.244 (0.007)	-0.025 (0.000)
	D-learning	95.7 (0.5)	4.1 (0.5)	0.2 (0.1)	-1.200 (0.011)	-0.024 (0.000)

When we consider the combined outcome and aim at maximizing the utility score, ITR effects on the units of blood collected,  $G$ , the low Hb deferral rates,  $R$ , and the utility outcome  $U$  are computed by replacing  $Y$  in (11) with  $G$ ,  $R$ , and  $U$ , respectively. A larger ITR effect on donation and utility, and a smaller ITR effect on deferral are more desirable.

#### 6.4 Non-Bayesian approaches: $l_1$ -PLS, ACWL, and D-learning

We first apply the  $l_1$ -PLS with HGL/GL variable selection, ACWL, and D-learning to data from male donors in the INTERVAL trial. In each of the analyses, we randomly split the data into a training and a validation set with a 4:1 ratio and repeat the procedure 100 times. All tuning parameters are selected via 5-fold cross-validation. For each method, we report the means and the standard deviations of empirical assignment proportions and ITR effects evaluated on the validation data across 100 splits. Note that the standard deviation estimates based on cross-validation capture the randomness of data-splitting (“repeatability”) rather than the uncertainty of the observed data. We provide bootstrap-based standard deviation estimates that reflect the “biological variation” in the online supplemental materials (Appendix D).

**6.4.1 Target two outcomes separately** Table 2 presents the results for analyzing the donation and deferral outcomes separately. As expected from the simulation results, we observe a consistent pattern across different methods in these extreme cases where true optimal decisions should be “almost trivial” and the non-personalized strategy that assigns everyone to the marginally best “treatment” should lead to almost optimal outcomes. As suggested by the “assignment percentages” columns in Table 2, almost all donors (ranging from 99.4% to 100.0%) are assigned to the shortest inter-donation interval (8-week) if the goal is to maximize the total units of blood collected by the blood service, and ITRs estimated using these precision medicine-based methods are indeed very close to “one-size-fits-all” rules. In contrast, if our aim is to minimize the low Hb deferral rates, then the longest inter-donation interval (12-week) should be recommended for most donors

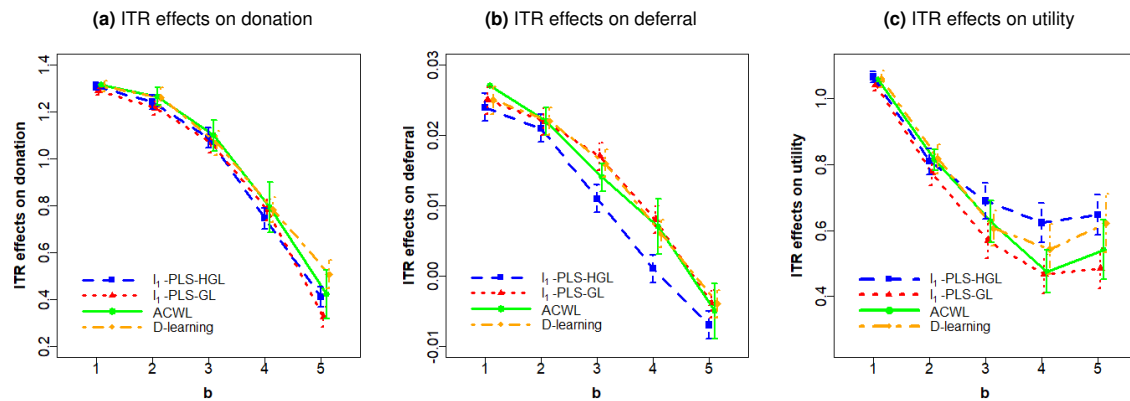
(ranging from 94.4% to 99.7%). These findings are consistent with results from the primary analysis of the INTERVAL trial data.<sup>12</sup>

For comparison, we also calculate the ITR effects on donation and deferral of three non-personalized rules where the same inter-donation interval is recommended for all donors. ITR effects of non-personalized (fixed) rules measure the difference in the average outcome between donors whose assigned inter-donation intervals in the trial are the same as the one specified in the fixed rule and those whose assigned inter-donation intervals are different from that specified in the fixed rule. We still use (11) to calculate empirical ITR effects of fixed rules except that  $\mathcal{D}(\mathbf{X})$  is replaced with fixed rules that do not depend on  $\mathbf{X}$ . ITR effects on donation are -1.248, -0.077, and 1.315 for assigning all donors to the 12-, 10-, and 8-week inter-donation interval, respectively; and ITR effects on deferral are -0.025, -0.002, and 0.027 for assigning all donors to the 12-, 10-, and 8-week inter-donation interval, respectively. This suggests that if the sole interest is in collecting more blood, the non-personalized rule that recommends all donors to donate every 8 weeks leads to the largest increase in the units of blood collected compared to personalized rules estimated using different methods. On the other hand, if we are only concerned with minimizing deferrals for low Hb, then the non-personalized rule that recommends all donors to the 12-week inter-donation interval yields the largest reduction in the rate of low Hb deferrals compared to personalized rules. However, we also observe from Table 2 that by following the optimal rule for maximizing the total units of blood collected by the blood service, the average increase in blood donations is about 1.31 units (616 ml) per donor over 2 years, but at the same time, there is also an increasing number of deferrals for low Hb at about 2.7 per 100 donor attendances on average. This is consistent with our intuition: an “optimal” rule that maximizes clinical benefits also leads to safety concerns (high risks of adverse events), and vice versa. Therefore, it is necessary to strike a balance between the two “competing” outcomes by maximizing the utility score which incorporates the trade-off between the benefit and the risk.

**6.4.2 Target the utility outcome** Table 3 summarizes allocation proportions and ITR effects associated with  $\widehat{\mathcal{D}}^*(\mathbf{X})$  for different values of the trade-off parameter  $b$  in the utility function, and Figure 1 (a)-(c) plot the relationships between estimated ITR effects (on donation, deferral and utility outcomes, respectively) and  $b$  when the target is to maximize the utility score. Again, we observe very similar results using different methods (“similar” in terms of clinical meaningfulness), especially when the value of  $b$  is small. As  $b$  increases, assignment proportions shift from the more frequent to less frequent inter-donation intervals: less donors are allocated to the shortest inter-donation interval, and more donors are allocated to the longest one. Consequently, both the increase in the total units of blood collected (i.e., benefit) and the increase in the low Hb deferral rates (i.e., risk) become smaller. When  $b$  takes the value from 1 to 4, both benefit and risk increase if the target is to maximize corresponding utilities, while in the extreme case where  $b = 5$ , there is an increase in the benefit and a decrease in the risk. As a reference, we also calculate the ITR effects of three non-personalized rules on the utility outcome when the trade-off parameter  $b$  varies from 1 to 5 (Table 4) and compare with those of personalized rules presented in Table 3. Unlike the case when we analyze two outcomes separately where the “one-size-fits-all” rule seems to be sufficient for achieving a desirable outcome, when we target the utility outcome, personalized rules that tailor to each donor’s capacity to donate lead to a higher gain in utility scores compared to non-personalized rules, and the advantage over non-personalized rules becomes more pronounced as  $b$  increases. For example, when  $b = 5$ , ITR effects on the utility score of the best non-personalized rule (assign all donors to the 10-week inter-donation interval) is 0.183, while ITR effects of personalized rules range from 0.485 to 0.648, depending on which method is used to estimate the optimal ITR.

**Table 3.** Applications of the  $l_1$ -penalized least squares with hierarchical group LASSO variable selection ( $l_1$ -PLS-HGL),  $l_1$ -penalized least squares with group LASSO variable selection ( $l_1$ -PLS-GL), adaptive contrast weighted learning (ACWL), and direct learning (D-learning) to data from male donors in the INTERVAL trial assuming the target is to maximize the utility. The trade-off parameter  $b$  in the utility function varies from 1 to 5 at an increment of 1. Means and standard deviations (in parenthesis) of assignment proportions in % and empirical ITR effects on donation, deferral, and utility across 100 repetitions of 5-fold cross-validation are reported. ITR effects measure the difference in the average outcome between donors whose assigned inter-donation intervals in the trial are optimal (with respect to the method used to estimate the ITR) and those whose assigned inter-donation intervals are non-optimal. A larger ITR effect on donation/utility and a smaller ITR effect on deferral are more desirable.

Trade-off Parameter	Method	Assignment Percentages			ITR Effects		
		12 weeks	10 weeks	8 weeks	Donation	Deferral	Utility
$b = 1$	$l_1$ -PLS-HGL	0.9 (0.1)	1.2 (0.1)	97.9 (0.1)	1.309 (0.008)	0.024 (0.001)	1.064 (0.009)
	$l_1$ -PLS-GL	0.3 (0.1)	2.4 (0.8)	97.2 (1.0)	1.289 (0.014)	0.025 (0.001)	1.040 (0.014)
	ACWL	0.0 (0.0)	0.0 (0.1)	100.0 (0.1)	1.314 (0.003)	0.027 (0.000)	1.055 (0.002)
	D-learning	0.7 (0.2)	1.2 (0.5)	98.1 (0.5)	1.309 (0.012)	0.025 (0.001)	1.058 (0.013)
$b = 2$	$l_1$ -PLS-HGL	3.4 (0.1)	4.2 (0.3)	92.4 (0.4)	1.242 (0.016)	0.021 (0.001)	0.809 (0.020)
	$l_1$ -PLS-GL	1.7 (0.4)	7.2 (1.6)	91.1 (2.0)	1.217 (0.027)	0.022 (0.002)	0.774 (0.024)
	ACWL	2.7 (0.8)	3.4 (1.7)	93.9 (1.3)	1.266 (0.019)	0.022 (0.001)	0.814 (0.016)
	D-learning	1.5 (0.4)	5.8 (1.1)	92.7 (1.0)	1.260 (0.022)	0.022 (0.001)	0.816 (0.023)
$b = 3$	$l_1$ -PLS-HGL	8.6 (0.3)	11.9 (0.5)	79.5 (0.6)	1.091 (0.022)	0.011 (0.001)	0.689 (0.028)
	$l_1$ -PLS-GL	4.8 (1.1)	15.0 (3.3)	80.2 (4.4)	1.069 (0.056)	0.017 (0.003)	0.569 (0.041)
	ACWL	9.3 (1.4)	8.2 (2.7)	82.5 (2.2)	1.100 (0.034)	0.014 (0.001)	0.627 (0.032)
	D-learning	3.8 (0.8)	17.5 (1.6)	78.6 (1.3)	1.067 (0.027)	0.016 (0.001)	0.607 (0.027)
$b = 4$	$l_1$ -PLS-HGL	17.0 (0.4)	23.3 (0.5)	59.7 (0.5)	0.745 (0.023)	0.001 (0.001)	0.623 (0.030)
	$l_1$ -PLS-GL	10.5 (2.3)	27.9 (3.2)	61.6 (5.2)	0.782 (0.070)	0.008 (0.004)	0.468 (0.081)
	ACWL	16.8 (1.9)	16.2 (3.5)	67.0 (3.1)	0.793 (0.055)	0.007 (0.002)	0.475 (0.033)
	D-learning	9.9 (1.6)	30.0 (1.7)	60.2 (0.7)	0.783 (0.027)	0.006 (0.001)	0.543 (0.039)
$b = 5$	$l_1$ -PLS-HGL	26.4 (0.4)	33.4 (0.5)	40.3 (0.3)	0.410 (0.022)	-0.007 (0.001)	0.648 (0.031)
	$l_1$ -PLS-GL	18.2 (3.2)	48.4 (6.2)	33.4 (3.4)	0.324 (0.059)	-0.004 (0.002)	0.485 (0.084)
	ACWL	30.1 (2.5)	22.6 (4.2)	47.3 (3.2)	0.422 (0.053)	-0.005 (0.002)	0.541 (0.046)
	D-learning	19.3 (1.6)	37.4 (1.3)	43.3 (0.7)	0.505 (0.031)	-0.004 (0.001)	0.622 (0.045)



**Figure 1.** Plots of the mean and 95% confidence intervals for ITR effects of the optimal ITR estimated using various methods as the trade-off parameter  $b$  in the utility function varies from 1 to 5 at an increment of 1. Optimal ITRs are estimated using data from male donors in the INTERVAL trial assuming the target is to maximize the utility. Methods to estimate the optimal ITR include  $l_1$ -penalized least squares with hierarchical group LASSO variable selection ( $l_1$ -PLS-HGL),  $l_1$ -penalized least squares with group LASSO variable selection ( $l_1$ -PLS-GL), adaptive contrast weighted learning (ACWL), and direct learning (D-learning). ITR effects on the (a) donation, (b) deferral, and (c) utility outcomes are presented. A larger ITR effect on donation/utility and a smaller ITR effect on deferral are more desirable.



**Table 4.** ITR effects of three non-personalized rules on the utility outcome. The trade-off parameter  $b$  in the utility function varies from 1 to 5 at an increment of 1. ITR effects measure the difference in the average outcome between donors whose assigned inter-donation intervals in the trial are the same as the one specified in the non-personalized rule and those whose assigned inter-donation intervals are different from that specified in the non-personalized rule. A larger ITR effect on utility is more desirable.

Non-personalized Rule	ITR Effects on Utility				
	$b = 1$	$b = 2$	$b = 3$	$b = 4$	$b = 5$
Recommend all male donors to donate every 12 weeks	-1.308	-0.828	-0.618	-0.408	-0.199
Recommend all male donors to donate every 10 weeks	-0.025	0.027	0.079	0.131	0.183
Recommend all male donors to donate every 8 weeks	1.055	0.795	0.535	0.275	0.015

**6.4.3 Variable selection by  $l_1$ -PLS-HGL and  $l_1$ -PLS-GL** In addition to estimating the optimal ITR,  $l_1$ -PLS with group LASSO (GL) or hierarchical group LASSO (HGL) variable selection also picks important treatment-covariate interactions when building the prediction model for  $E(Y|X, A)$ . Despite the mismatch issue discussed in Section 4.1, we are interested in investigating which treatment-covariate interactions are estimated as non-zero and regarded as important in the prediction model. For demonstration, we focus on the situation where maximizing the total units of blood collected by the blood service is our primary goal. Table 5 summarizes selection percentages (across 100 repetitions of 5-fold cross-validation) of different treatment-covariate interactions in the prediction model for donation when group LASSO that does not impose strong hierarchy between main effects and interactions, or hierarchical group LASSO that enforces such hierarchy is used for variable selection.

We observe that some interactions between the randomized group (inter-donation interval) and baseline characteristics are selected almost all the time by both variable selection methods, such as blood donations in the 2 years before trial enrollment, baseline haemoglobin level, blood group, etc. We also notice that for some baseline covariates, selection percentages of their interaction with the randomized group differ substantially between the two variable selection approaches, even though the donor assignment proportions and ITR effects estimated using  $l_1$ -PLS-HGL and  $l_1$ -PLS-GL (Table 2) are very similar. A possible explanation for an interaction being selected much more often by HGL than by GL (the case for most continuous covariates) is that the effect of the interaction itself is not strong enough and may be dominated by the main effect, but as has been noted in Lim and Hastie,<sup>33</sup> their proposed HGL method can still discover important interaction terms in this case due to the use of the “glinternet” penalty.

It is worth noting that the high selection rates of interactions between randomized groups and baseline covariates do not imply different recommendations on the inter-donation interval for different donors, and the finding that almost all male donors should donate every 8 weeks to maximize the total units of blood collected (Table 2) does not contradict the observation that many interaction effects are estimated to be non-zero (Table 5). This is because variable selection methods that we use to identify important interactions do not distinguish between quantitative and qualitative interactions, whereas only qualitative interactions can lead to different inter-donation interval recommendations for different subpopulations. In this dataset, it might be the case that even though interactions exist, they are mostly quantitative, as has also been observed in the pre-specified subgroup analysis in the INTERVAL trial.<sup>12</sup>

**Table 5.** Selection percentages of treatment-covariate interactions in the prediction model for the donation outcome across 100 repetitions of 5-fold cross-validation when  $l_1$ -PLS is used to estimate the optimal ITR. Hierarchical group LASSO (HGL) enforces strong hierarchy between main effects and interactions, and group LASSO (GL) does not impose strong hierarchy between main effects and interactions.

Baseline Variables	Variable Type	Variable Selection	
		HGL	GL
Age	continuous	63	21
Body mass index	continuous	33	38
SF-36v2 physical component score	continuous	47	11
SF-36v2 mental component score	continuous	100	20
Blood donations in the 2 years before trial enrollment	continuous	100	100
Haemoglobin level	continuous	100	100
White blood cell count	continuous	100	46
Red blood cell count	continuous	0	11
Mean corpuscular haemoglobin	continuous	100	96
Mean corpuscular volume	continuous	62	11
Platelet count	continuous	100	12
Ethnicity	categorical	47	97
Blood group	categorical	99	99
Iron prescription	categorical	1	74
Smoke ever	categorical	13	39
Smoke currently	categorical	98	99
Alcohol ever	categorical	85	32
Alcohol currently	categorical	100	100
New or returning donor status	categorical	100	32

## 6.5 The Bayesian approach: BART

Following Logan et al.,<sup>10</sup> we first build the conditional mean outcome model using BART, and then we identify the optimal ITR based on posterior predictive distributions of the conditional mean under each randomized group (the resulting optimal ITR is referred to as the “BART ITR”). According to Section 4.4, the BART ITR is the one in which the recommended arm for each individual is given by maximizing the subject-specific MCMC estimate (e.g. posterior mean) of the posterior predictive distribution of  $E(Y|\mathbf{X} = \mathbf{x}, A = a)$ . In the INTERVAL trial, the BART ITR is the rule that assigns each donor to the inter-donation interval that leads to the largest posterior mean of utilities/total units of blood collected, or the inter-donation interval that corresponds to the smallest posterior mean of low Hb deferral rates.

**6.5.1 Target two outcomes separately** Consistent with our findings using frequentist approaches, when we fit the BART model to the INTERVAL data with the target of maximizing the total units of blood collected by the blood service, the BART ITR,  $\widehat{\mathcal{D}}_{\text{BART}}^*(\mathbf{X})$ , assigns all male donors to the most frequent inter-donation interval (8-week) with the posterior mean of the ITR effects on donation being 1.313 (95% credible interval:

[1.237, 1.392]). On the other hand, when the aim is to minimize low Hb deferral rates,  $\widehat{\mathcal{D}}_{\text{BART}}^*(\mathbf{X})$  recommends 91.6% of male donors to donate every 12 weeks, leading to an ITR effect of -0.024 (posterior mean) on the deferral outcome with the 95% credible interval being [-0.027, -0.022].

**6.5.2 Target the utility outcome** We also fit the BART model to the INTERVAL data with the utility score being the target outcome and obtain the posterior predictive distribution of the utility score under each inter-donation interval option. We vary the trade-off parameter  $b$  in the utility function from 1 to 5 at an increment of 1 to see how  $\widehat{\mathcal{D}}_{\text{BART}}^*(\mathbf{X})$  changes with this parameter. For each  $b$ , we compare five donor assignment rules:

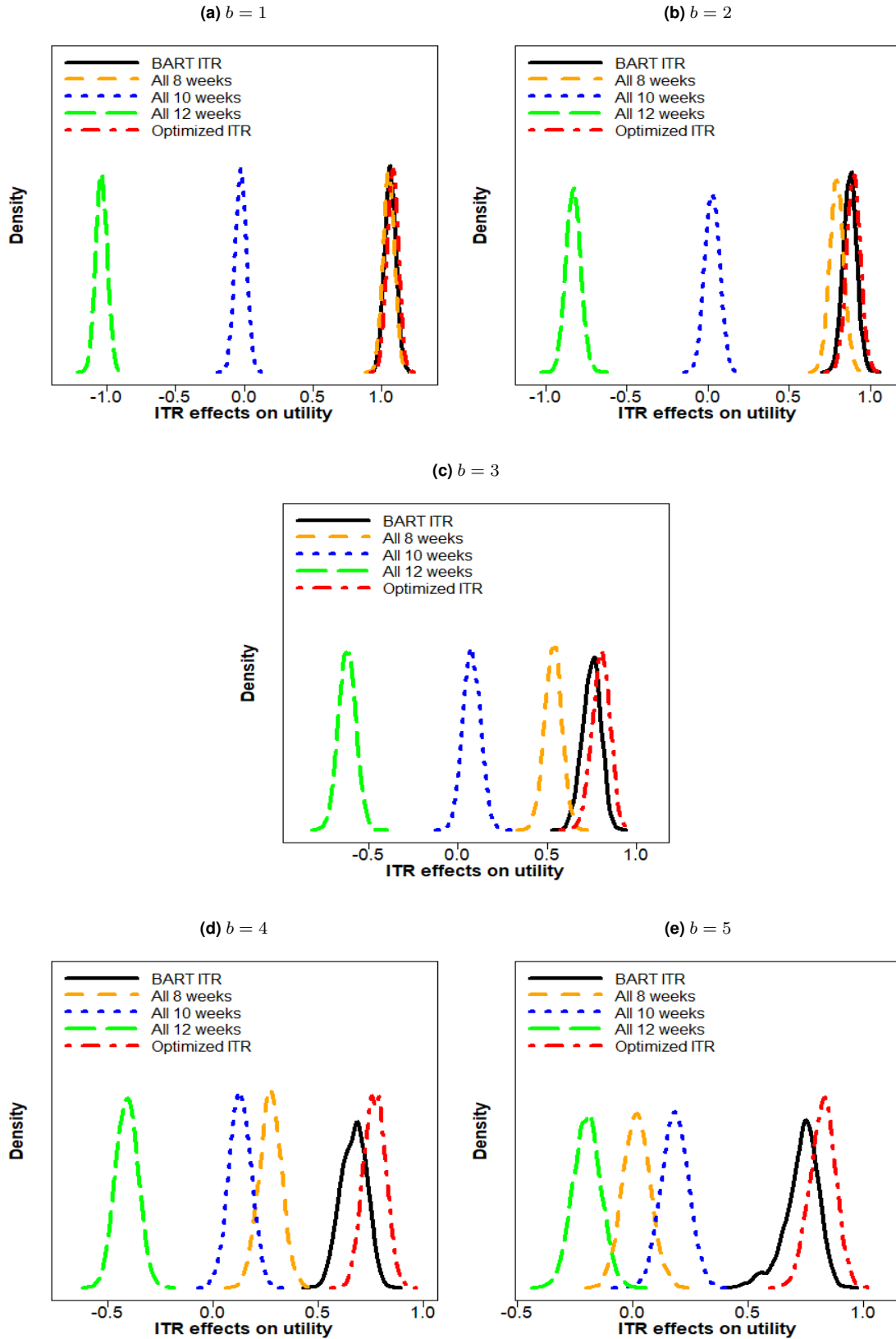
- (i) All donors donate every 8 weeks;
- (ii) All donors donate every 10 weeks;
- (iii) All donors donate every 12 weeks;
- (iv) Donors donate according to the BART ITR (each donor is assigned to the inter-donation interval associated with the maximum posterior mean of the utility score);
- (v) Donors donate according to the optimized ITR based on the BART estimation (an idealized scenario in which for each MCMC draw from the posterior distribution, each donor is assigned to the inter-donation interval associated with the maximum value of the utility score. This is non-achievable in practice, but we use this rule as a reference for the best-case scenario).

Posterior distributions of the ITR effects associated with those five donor assignment rules corresponding to five different trade-off parameters are plotted in Figure 2. In Table 6, we summarize donor allocation proportions based on the BART ITR, and we also report the posterior mean and the 95% equal tail credible interval of the ITR effect for each donor assignment strategy (we note that standard deviation estimates reported in Table 3 for non-Bayesian methods are calculated across 100 repetitions of 5-fold cross-validation, while in Table 6, we quantify the uncertainty of the estimates directly based on posterior samples from BART).

We observe that when  $b$  is small (e.g.  $b=1$  or  $2$ ), the best “one-size-fits-all” strategy is to recommend all donors to donate every 8 weeks. In these cases, both the BART ITR and the optimized ITR are close to the “all 8 weeks” rule in that the ITR effect distribution of the BART ITR and the optimized ITR almost overlaps with that of the “all 8 weeks” rule. When  $b$  takes values from 3 to 5, the BART ITR is better than the best non-personalized assignment rule (“all 8 weeks” for  $b = 3, 4$ , and “all 10 weeks” for  $b = 5$ ) with a high probability, and the advantage of the BART ITR over non-personalized rules becomes more pronounced as  $b$  gets larger. Even though the BART ITR is slightly inferior to the optimized ITR (which is generally not achievable in practice) as expected, the differences are minimal.

## 6.6 Measure of agreement between methods

Results presented in Section 6.4 and 6.5 suggest that donor assignment proportions and empirical ITR effects are fairly similar (at the population level) across different methods for each target outcome. We are also interested in investigating for a given male donor, to what extent the five methods ( $l_1$ -PLS-HGL,  $l_1$ -PLS-GL, ACWL, D-learning, and BART) “agree” on his optimal inter-donation interval (at the individual level). Given the differences in the decision boundary types assumed by each method, we expect that the form of estimated regimes based on different methods may be very different and the set of baseline characteristics included in the optimal decision rule may also vary across methods. However, the optimal inter-donation interval estimated



**Figure 2.** Density plots of ITR effects on utility for five donor assignment rules: recommend all male donors to (i) donate every 8 weeks, (ii) donate every 10 weeks, (iii) donate every 12 weeks, (iv) donate according to the BART ITR, and (v) donate according to the optimized ITR (non-achievable in practice). The trade-off parameter  $b$  in the utility function varies from 1 to 5 at an increment of 1. ITR effects measure the difference in the average outcome between donors whose assigned inter-donation intervals in the trial are optimal (with respect to the method used to estimate the ITR) and those whose assigned inter-donation intervals are non-optimal. A larger ITR effect on utility is more desirable.

**Table 6.** Applications of Bayesian additive regression trees (BART) to data from male donors in the INTERVAL trial assuming the target is to maximize the utility score. The trade-off parameter  $b$  in the utility function varies from 1 to 5 at an increment of 1. Assignment proportions of the BART ITR in % and the posterior mean [95% equal tail credible interval] of the ITR effect on the utility outcome for five donor assignment rules are reported. Assignment rules include: recommend all male donors to (i) donate every 8 weeks, (ii) donate every 10 weeks, (iii) donate every 12 weeks, (iv) donate according to the BART ITR, and (v) donate according to the optimized ITR (non-achievable in practice). ITR effects measure the difference in the average outcome between donors whose assigned inter-donation intervals in the trial are optimal (with respect to the method used to estimate the ITR) and those whose assigned inter-donation intervals are non-optimal. A larger ITR effect on utility is more desirable.

Criteria	Assignment Rule	Trade-off Parameter				
		$b = 1$	$b = 2$	$b = 3$	$b = 4$	$b = 5$
BART ITR Assignment Percentages	12 weeks	0.0	3.6	10.7	17.5	26.7
	10 weeks	3.2	9.4	18.8	25.5	28.4
	8 weeks	96.8	87.1	70.5	57.0	44.9
ITR Effects on Utility Posterior Mean [95% Credible Interval]	All 12 weeks	-1.037 [-1.118,-0.956]	-0.827 [-0.915,-0.739]	-0.618 [-0.712,-0.524]	-0.408 [-0.511,-0.305]	-0.199 [-0.317,-0.082]
	All 10 weeks	-0.025 [-0.106,0.055]	0.027 [-0.060,0.114]	0.079 [-0.017,0.174]	0.131 [0.027,0.232]	0.183 [0.068,0.297]
	All 8 weeks	1.054 [0.973,1.134]	0.794 [0.708,0.883]	0.535 [0.437,0.628]	0.275 [0.169,0.379]	0.016 [-0.101,0.135]
	BART ITR	1.064 [0.983,1.144]	0.876 [0.793,0.959]	0.750 [0.643,0.843]	0.671 [0.553,0.780]	0.732 [0.548,0.852]
	Optimized ITR	1.082 [1.000,1.159]	0.898 [0.814,0.981]	0.802 [0.701,0.893]	0.770 [0.671,0.870]	0.823 [0.710,0.928]

by different approaches for a given donor can still “overlap” despite the possible heterogeneity in the form of estimated regimes.<sup>37</sup>

To evaluate the degree of agreement between different methods in terms of the recommended optimal inter-donation interval for each male donor in the INTERVAL trial, we present two inter-method agreement measures (overall and pairwise). Details and results can be found in the supplemental materials (Appendix E). We conclude that in most cases, the level of agreement is “substantial” or “almost perfect” based on the values of the pairwise B statistics<sup>52</sup> and the guidelines for assessing agreement by Munoz and Bangdiwala.<sup>53</sup>

## 7 Discussion

### 7.1 Concluding remarks

Most statistical methods for estimating the optimal individualized treatment rule (ITR) are restricted to binary treatment comparisons. However, clinical studies with more than two treatment arms are common in practice. In this paper, we review several recent approaches that can be used to estimate the optimal ITR in large-scale clinical studies with more than two treatment options. Methods considered include: the  $l_1$ -penalized least squares with hierarchical group LASSO or group LASSO variable selection, the adaptive contrast weighted learning method, the direct learning method, and a Bayesian approach that builds on Bayesian additive regression trees.

We conduct simulation studies to evaluate the performance of these methods in large-scale clinical trials. Our simulation results suggest that the BART multi-arm ITR estimation method has better or similar performance compared to other methods across different settings (with different types of interaction terms). In addition,

when the sample size is large (as in the INTERVAL trial), all methods perform equally well under scenarios where baseline characteristics interact with treatment arms only quantitatively (and thus the true optimal treatment is the same for all subjects) in the sense that they all accurately identify the true universal decision rules (assign all to the marginally best one).

We further illustrate the use of these methods by applying them to the data collected from male donors in the INTERVAL trial. Results are fairly consistent across different approaches in terms of clinical meaningfulness. When the target is to maximize the total units of blood collected by the blood service, or to minimize the low Hb deferral rates (i.e., the benefit outcome and the risk outcome are not considered in conjunction with each other), all methods detect almost no qualitative heterogeneity of the “inter-donation interval” effects. More specifically, if we focus on maximizing the total units of blood collected, then the optimal ITR tends to choose shorter inter-donation intervals that are associated with increased benefits at the cost of higher risks of deferrals for low Hb and almost all donors are assigned to the highest frequency of donation. On the other hand, if our aim is to minimize the low Hb deferral rates, then the optimal ITR picks longer inter-donation intervals and almost everyone is assigned to the lowest frequency of donation. These results are not surprising and support the trial’s primary findings that interactions between baseline characteristics and the inter-donation interval are not qualitative, but rather quantitative, and almost all donors are able to give blood more frequently than the current standard practice.<sup>12</sup> Maximizing the benefit and minimizing the risk are two competing goals. While the optimal decision for each goal is obvious, the two decisions may be very different. To deal with this, we create a utility score that balances two outcomes and derive the optimal ITR with the goal of maximizing the utility score. Investigation of the utility outcome suggests some heterogeneity in the optimal inter-donation interval across donors with different baseline characteristics, and such heterogeneity becomes larger as the trade-off parameter  $b$  in the utility function (the equivalent benefit loss for one unit increase in risk) gets larger.

## 7.2 Generalizability of estimated ITRs

One issue that warrants highlighting is the generalizability of estimated ITRs to a broader population. In general, the entry criteria for clinical trials are restrictive, and trial participants may not be representative of the more general population. Therefore, failure to detect qualitative interactions between baseline characteristics and the treatment assignment among trial participants does not preclude the existence of such qualitative interactions over the entire eligible population, and we should be cautious about generalizing ITRs estimated using trial data to a broader population.<sup>32,54</sup> However, this may not be a major problem for the INTERVAL trial. As suggested by Moore et al.,<sup>17</sup> participants in the INTERVAL trial were broadly representative of the national donor population of England, and it is likely that estimated ITRs based on INTERVAL data are generalizable to the general donor population.

We note that the INTERVAL data cannot be used to validate the estimated optimal donation strategies. A confirmatory follow-up trial comparing estimated rules with the current clinical practice should be conducted before applying estimated rules to future donors.

We also comment that the generalizability and validation issues discussed in this section regarding the estimated optimal ITRs are relevant not only to the INTERVAL trial but also more generally to other trials, as have been discussed in literatures.<sup>27,32,54,55</sup>

### 7.3 Model extensions and future work

We can extend this work in several directions. We treat the inter-donation interval as a nominal variable with 3 categories. However, there is a natural ordering of the 12-, 10-, and 8-week inter-donation interval and these three options can be considered as three levels of an ordinal “treatment”. We would expect to have some information loss by treating an ordinal variable as nominal. Thus, it would be of great interest to explore ways to efficiently incorporate information on the ordinality into decision-making and investigate how much we gain by doing this.

We note that the optimal donor assignment strategy is highly dependent on  $b$ . We examine different values for  $b$  within a reasonable range and estimate the optimal ITR under each  $b$ . Ideally,  $b$  should be specified based on clinicians’ domain knowledge. However, there is a lack of accurate information on  $b$  in most cases because the right balance between the benefit and the risk is usually not obvious and can be different for different subgroups.<sup>56</sup> Kosorok and Laber<sup>2</sup> summarized other recent work on addressing multiple outcomes in precision medicine. For example, an alternative approach to handle the trade-off between the risk and the benefit is through the constrained optimization framework, i.e., consider maximizing the benefit under the constraint of controlling the average risk under a pre-specified and clinically meaningful threshold.<sup>57</sup> Similar to the utility-based approach, the choice of the threshold value is important for making the right decision.

When deriving the optimal ITR, we only use baseline measurements that are routinely collected at the regular donation session. In the future, we may include additional blood-based biomarkers (e.g. those related to iron stores such as ferritin and transferrin) to estimate the optimal donation strategy. Extra costs would be incurred collecting such data, and failure to account for such additional costs can lead to suboptimal decisions from the cost-effectiveness perspective. We will investigate how much additional information on biomarkers adds to the reinforcement of decision-making (compared to the donation strategy estimated solely based on routinely collected data).

Dynamic treatment regimes (DTR) refer to sequential decision rules that adapt over time to the changing status of each subject to maximize the expected long-term target outcomes.<sup>21,58–60</sup> In the INTERVAL trial, donors were only randomized once at the baseline visit (and then fixed at the initial randomized group), unlike in the sequential multiple assignment randomized trial (SMART) where “treatment assignments” at later decision points are based on the response to the previous randomization and updated individual characteristics.<sup>61</sup> In addition, we do not have the data on each donor’s Hb levels and iron stores at subsequent donation sessions, even though they may affect the optimal gap time before the next donation. Therefore, based on the data currently available from the INTERVAL trial, we only consider the estimation of single-stage ITRs using baseline measurements in this paper. It would be useful to incorporate dynamic donor stratification and estimate the optimal personalized donation strategy that reflects both heterogeneity across donors and heterogeneity over time within each donor when other data sources become available.

### Acknowledgments

The authors would like to thank the reviewers for their helpful comments and suggestions, which led to an improved paper. In addition, we thank members of the NIHR Blood and Transplant Research Unit in Donor Health and Genomics including Dr Steven Bell, Dr Emanuele Di Angelantonio, Dr Stephen Kaptoge, Dr Lois Kim, Andrew Browne, and Tom Bolton for discussions and feedback on an earlier version of this manuscript.

A complete list of the investigators and contributors to the INTERVAL trial is provided in Di Angelantonio et al.<sup>12</sup> The academic coordinating centre would like to thank blood donor centre staffs and blood donors for participating in the INTERVAL trial.

## Declaration of conflicting interests

The authors declare that there is no conflict of interest with respect to the research, authorship, and publication of this article.

## Funding

This work was supported by the UK Medical Research Council programme MC\_UU\_00002/2 and the Cambridge International Scholarship. Participants in the INTERVAL trial were recruited with the active collaboration of NHS Blood and Transplant England ([www.nhsbt.nhs.uk](http://www.nhsbt.nhs.uk)), which has supported field work and other elements of the trial. The academic coordinating centre for INTERVAL was supported by core funding from: NIHR Blood and Transplant Research Unit in Donor Health and Genomics (NIHR BTRU-2014-10024), UK Medical Research Council (MR/L003120/1), British Heart Foundation (SP/09/002; RG/13/13/30194; RG/18/13/33946) and the NIHR [Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust]. This work was also supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. The views expressed in this paper are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

## Supplemental material

Supplemental material is available online for this article.

## References

1. Lipkovich I, Dmitrienko A and BD'Agostino Sr R. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Stat Med* 2017; 36(1): 136–196.
2. Kosorok MR and Laber EB. Precision medicine. *Annu Rev Stat Appl* 2019; 6(1): 263–286.
3. Baron G, Perrodeau E, Boutron I et al. Reporting of analyses from randomized controlled trials with multiple arms: a systematic review. *BMC Med* 2013; 11(1): Article 84.
4. Lou Z, Shao J and Yu M. Optimal treatment assignment to maximize expected outcome with multiple treatments. *Biometrics* 2018; 74(2): 506–516.
5. Zhou X, Wang Y and Zeng D. Outcome-weighted learning for personalized medicine with multiple treatment options. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. pp. 565–574.
6. Lee Y, Lin Y and Wahba G. Multicategory support vector machines. *J Am Stat Assoc* 2004; 99(465): 67–81.



7. Qian M and Murphy SA. Performance guarantees for individualized treatment rules. *Ann Stat* 2011; 39(2): 1180–1210.
8. Tao Y and Wang L. Adaptive contrast weighted learning for multi-stage multi-treatment decision-making. *Biometrics* 2017; 73(1): 145–155.
9. Qi Z and Liu Y. D-learning to estimate optimal individual treatment rules. *Electron J Stat* 2018; 12(2): 3601–3638.
10. Logan BR, Sparapani R, McCulloch RE et al. Decision making and uncertainty quantification for individualized treatments using bayesian additive regression trees. *Stat Method Med Res* 2019; 28(4): 1079–1093.
11. Moore C, Sambrook J, Walker M et al. The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial. *Trials* 2014; 15: Article 363.
12. Di Angelantonio E, Thompson SG, Kaptoge S et al. Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45000 donors. *Lancet* 2017; 390(10110): 2360–2371.
13. Holmes CC and Watson JA. Machine learning for randomised controlled trials : identifying treatment effect heterogeneity with strict control of type I error. *bioRxiv* 2018; URL <https://www.biorxiv.org/content/early/2018/06/02/330795>. <https://www.biorxiv.org/content/early/2018/06/02/330795.full.pdf>.
14. Karp JK and King KE. International variation in volunteer whole blood donor eligibility criteria. *Transfusion* 2010; 50(2): 507–513.
15. Goldman M, Magnussen K, Gorlin J et al. International forum regarding practices related to donor haemoglobin and iron. *Vox Sang* 2016; 111(4): 449–455.
16. Vuk T, Magnussen K, De Kort W et al. International forum: An investigation of iron status in blood donors. *Blood Transfus* 2017; 15(1): 20–41.
17. Moore C, Bolton T, Walker M et al. Recruitment and representativeness of blood donors in the INTERVAL randomised trial assessing varying inter-donation intervals. *Trials* 2016; 17(1): Article 458.
18. Robins JM. Causal inference from complex longitudinal data. In Berkane M (ed.) *Latent Variable Modeling and Applications to Causality*. New York, NY: Springer New York. ISBN 978-1-4612-1842-5, pp. 69–117.
19. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat - Theory Methods* 1994; 23(8): 2379–2412.
20. Watkins CJ and Dayan P. Q-learning. *Machine Learning* 1992; 8(3): 279–292.
21. Murphy SA. Optimal dynamic treatment regimes. *J Royal Stat Soc B* 2003; 65(2): 331–355.
22. Robins JM. Optimal structural nested models for optimal sequential decisions. In Lin DY and Heagerty PJ (eds.) *Proceedings of the Second Seattle Symposium in Biostatistics*. New York, NY: Springer New York, 2004. pp. 189–326.
23. Henderson R, Ansell P and Alshibani D. Regret-regression for optimal dynamic treatment regimes. *Biometrics* 2010; 66(4): 1192–1201. DOI:10.1111/j.1541-0420.2009.01368.x. URL <http://dx.doi.org/10.1111/j.1541-0420.2009.01368.x>.
24. Wallace MP and Moodie EEM. Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics* 2015; 71(3): 636–644.
25. Wallace M, Moodie E and Stephens D. Dynamic treatment regimen estimation via regression-based techniques: Introducing r package dtrreg. *J Stat Softw* 2017; 80(2): 1–20.
26. Zhao Y, Zeng D, Rush AJ et al. Estimating individualized treatment rules using outcome weighted learning. *J Am Stat Assoc* 2012; 107(499): 1106–1118.

27. Zhou X, Mayer-Hamblett N, Khan U et al. Residual weighted learning for estimating individualized treatment rules. *J Am Stat Assoc* 2017; 112(517): 169–187.
28. Liu Y, Wang Y, Kosorok MR et al. Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Stat Med* 2018; 37(26): 3776–3788.
29. Lou Z, Shao J and Yu M. Optimal treatment assignment to maximize expected outcome with multiple treatments. *Biometrics* 2018; 74(2): 506–516.
30. Scholkopf B and Smola AJ. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2001. ISBN 978-026219475-4.
31. Yuan M and Lin Y. Model selection and estimation in regression with grouped variables. *J Royal Stat Soc B* 2006; 68(1): 49–67.
32. Gunter L, Zhu J and Murphy S. Variable selection for qualitative interactions. *Stat Methodol* 2011; 8(1): 42 – 55.
33. Lim M and Hastie T. Learning interactions via hierarchical group-lasso regularization. *J Comput Graph Stat* 2015; 24(3): 627–654.
34. Gail M and Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* 1985; 41(2): 361–372.
35. Imai K and Ratkovic M. Estimating treatment effect heterogeneity in randomized program evaluation. *Ann Appl Stat* 2013; 7(1): 443–470.
36. Tian L, 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.
37. Zhang B and Zhang M. Variable selection for estimating the optimal treatment regimes in the presence of a large number of covariates. *Ann Appl Stat* 2018; 12(4): 2335–2358.
38. Zhang B, Tsiatis AA, Laber EB et al. A robust method for estimating optimal treatment regimes. *Biometrics* 2012; 68(4): 1010–1018.
39. Bishop CM. *Pattern Recognition and Machine Learning*. Springer, 2007.
40. Chipman HA, George EI and McCulloch RE. BART: Bayesian additive regression trees. *Ann Appl Stat* 2010; 4(1): 266–298.
41. Tan YV and Roy J. Bayesian additive regression trees and the general bart model. *Stat Med* 2019; 38(25): 5048–5069.
42. Hill J, Linero A and Murray J. Bayesian Additive Regression Trees: A Review and Look Forward. *Annu Rev Stat Appl* 2019; 7(1).
43. Sparapani R, Spanbauer C and McCulloch R. Nonparametric machine learning and efficient computation with bayesian additive regression trees: the BART R package. *J Stat Softw* 2019; In press.
44. JPAC. Guidelines for the blood transfusion services in the UK, 8th edition, 2019. URL <https://www.transfusionguidelines.org/red-book>.
45. Grieve R, Willis S, De Corte K et al. *Options for possible changes to the blood donation service: health economics modelling*. Health Services and Delivery Research, Southampton, UK: NIHR Journals Library, 2018.
46. Willis S, De Corte K, Cairns JA et al. Cost-effectiveness of alternative changes to a national blood collection service. *Transfus Med* 2019; In press.
47. Hillgrove T, Moore V, Doherty K et al. The impact of temporary deferral due to low hemoglobin: future return, time to return, and frequency of subsequent donation. *Transfusion* 2011; 51(3): 539–547.
48. Green J and Davies K. Understanding the implications of the INTERVAL and COM-PARE studies, 2018. URL <https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/7847/>

- interval-and-compare-implications-for-haemoglobin-screening-board-jan-18.pdf.
49. Tukey JW. *Exploratory Data Analysis*. Addison-Wesley, 1977.
  50. Xu Y, Yu M, Zhao Y et al. Regularized outcome weighted subgroup identification for differential treatment effects. *Biometrics* 2015; 71(3): 645–653.
  51. Qiu X, Zeng D and Wang Y. Estimation and evaluation of linear individualized treatment rules to guarantee performance. *Biometrics* 2018; 74(2): 517–528.
  52. Bangdiwala SI. A graphical test for observer agreement. In *Proceedings of the 45th International Statistical Institute Meeting*. Amsterdam, 1985. pp. 307–308.
  53. Munoz SR and Bangdiwala SI. Interpretation of kappa and B statistics measures of agreement. *J Appl Stat* 1997; 24(1): 105–112.
  54. Zhao Y, Zeng D, Tangen CM et al. Robustifying trial-derived optimal treatment rules for a target population. *Electron J Stat* 2019; 13(1): 1717–1743.
  55. Chen G, Zeng D and Kosorok MR. Personalized dose finding using outcome weighted learning. *J Am Stat Assoc* 2016; 111(516): 1509–1521.
  56. Liu W, Zhang Z, Nie L et al. A case study in personalized medicine: Rilpivirine versus efavirenz for treatment-naive HIV patients. *J Am Stat Assoc* 2017; 112(520): 1381–1392.
  57. Wang Y, Fu H and Zeng D. Learning optimal personalized treatment rules in consideration of benefit and risk: With an application to treating type 2 diabetes patients with insulin therapies. *J Am Stat Assoc* 2018; 113(521): 1–13.
  58. Moodie EEM, Richardson TS and Stephens DA. Demystifying optimal dynamic treatment regimes. *Biometrics* 2007; 63(2): 447–455.
  59. Chakraborty B and Moodie EE. *Statistical methods for dynamic treatment regimes: reinforcement learning, causal inference, and personalized medicine*. Statistics for biology and health, New York, NY: Springer, 2013. ISBN 9781461474272.
  60. Zhao Y, Zeng D, Laber EB et al. New statistical learning methods for estimating optimal dynamic treatment regimes. *J Am Stat Assoc* 2015; 110(510): 583–598.
  61. Kosorok MR and Moodie EEM. *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*. SIAM, 2016. ISBN 978-161197418-8.