Biometrics, 2025, 81(1), ujaf026
https://doi.org/10.1093/biomtc/ujaf026
Biometric Practice

OXFORD

Optimal treatment regime estimation in practice: challenges and choices in a randomized clinical trial for depression

Florian Stijven [1,*], Trung Dung Tran[2], Ellen Driessen[3,4], Ariel Alonso Abad [1], Geert Molenberghs[1,5], Geert Verbeke[1,5], Iven Van Mechelen[2]

[1]I-BioStat, KU Leuven, B-3000 Leuven, Belgium, [2]Faculty of Psychology and Educational Sciences, KU Leuven, B-3000 Leuven, Belgium, [3]Behavioral Science Institute, Radboud Universiteit, 6500 HE Nijmegen, The Netherlands, [4]Depression Expertise Center, Pro Persona Mental Health Care, 6525 DX Nijmegen, The Netherlands, [5]I-BioStat, Universiteit Hasselt, B-3500 Hasselt, Belgium

[*]Corresponding author: Florian Stijven, I-BioStat, KU Leuven, B-3000 Leuven, Belgium (florian.stijven@kuleuven.be).

ABSTRACT

An important aspect of precision medicine is the tailoring of treatments to specific patient types. Nowadays, various methods are available to estimate for this purpose so-called optimal treatment regimes, that is, decision rules for treatment assignment that map patterns of pretreatment characteristics to treatment alternatives and that maximize the expected patient benefit. However, the application of these methods to real-life data has been limited and comes with nonstandard statistical issues. In search of best practices, we reanalyzed data from a randomized clinical trial for the treatment of dysthymic disorder. While the original objective of this trial was to detect a marginally best treatment alternative, we wanted to estimate an optimal treatment regime using 2 prominent estimation methods: Q-learning and value search estimation. An important obstacle in the dataset under study was the occurrence of missing values. This was handled with multiple imputation, a thoughtful implementation of which, however, implied several challenges. Other challenges were implied by the concrete implementation of value search estimation. In this paper, all the choices we have made in the analysis to handle the aforementioned issues are detailed together with a motivation and a description of possible alternatives. Accordingly, this paper may serve as a guide to apply optimal treatment regime estimation in data-analytic practice.

KEYWORDS: multiple imputation; optimal treatment regimes; Q-learning; randomized clinical trial; value search estimation.

1 INTRODUCTION

Given data from a randomized clinical trial (RCT) or an observational study in which patients have been assigned to one out of several treatment alternatives, one may wonder which treatment option has the best mean outcome. Beyond that question, one may wonder which treatment alternative works best for which types of patients, as defined in terms of pretreatment characteristics. Indeed, tailoring treatments to specific patient types is an important challenge within precision medicine. Such a tailoring can be implemented in so-called treatment regimes, that is, decision rules for treatment assignment that map patterns of pretreatment characteristics to treatment alternatives (Chakraborty and Moodie, 2013; Kosorok and Moodie, 2015; Tsiatis et al., 2020). Examples of families of treatment regimes include tree-based and linear ones. In practice, a major challenge is the estimation of treatment regimes that are optimal in some sense, such as yielding an optimal mean outcome when applied to a target population of patients (Tran et al., 2023). Optimal treatment regime (OTR) estimation may yield both trivial, or "one-size-fits-all," regimes that assign all patients to the same treatment alternative and nontrivial regimes for which this is not the case.

In the remainder of the present paper, we will focus on the estimation of optimal linear treatment regimes based on data from an RCT. We focus on linear regimes because of the mainstream nature of linear models in the statistical literature. Assume that, for each patient, data $(X_1, \cdots, X_p, \cdots, X_P; A; Y)$ are available, with $X_1, \cdots, X_p, \cdots, X_P$ denoting P pretreatment characteristics, A the treatment alternative to which the patient was assigned in the RCT (with, for simplicity, A taking the values 0 and 1), and Y a continuous outcome variable (without loss of generality, higher values of Y are assumed to be better). Linear treatment regimes then take the following form:

$$I(\eta_0 + \eta_1 X_1 + \cdots + \eta_P X_P > 0), \tag{1}$$

where $I(\cdot)$ is the indicator function. Such regimes are relatively easy to interpret and can be extended by including interaction and nonlinear terms.

There are many OTR estimation methods. Li et al. (2023) classify them into 3 groups: (i) methods based on estimating the entire data-generating process, (ii) regression-based methods, and (iii) value-search methods. We will use Q-learning (QL)—short for quality-learning— and value search estimation (VSE) as key representatives of the latter 2 groups. We will not use methods based on the entire data-generating process as these may be equivalent to regression-based methods in single-stage dynamic OTR estimation (Li et al., 2023, p. 16).

Received: April 24, 2024; Revised: February 18, 2025; Accepted: March 4, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of The International Biometric Society. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

QL comes down to estimating a regression model such as

$$E(Y|A, \boldsymbol{X}) = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_P X_P + A(\eta_0 + \eta_1 X_1 + \cdots + \eta_P X_P), \quad (2)$$

where $\boldsymbol{X} = (X_1, \ldots, X_p)'$. Subsequently, a treatment regime can be derived from the estimated model and, specifically, from the interaction term(s) with treatment assignment.

From its part, VSE *directly* targets an estimate of the value of the criterion function at the basis of treatment regime optimality; in our case, this is an estimate of the regime's expected potential outcome when applied to the target patient population. A suitable estimator for this purpose is the Augmented Inverse Probability Weighted Estimator (AIPWE) as described by Tsiatis et al. (2020, pp. 61-63). This estimator relies on 2 models: (i) a propensity model for the probabilities of treatment assignment and (ii) a regression model for the outcome variable. Specifically, the AIPWE weighs observations consistent with the regime under study by the inverse probability of being consistent with it (estimated on the basis of the propensity model) and includes a correction term involving the predicted mean outcome given the covariates (estimated on the basis of the regression model). This estimator is doubly robust in that it is consistent if the propensity model, the outcome model, or both are correct. For RCTs, where the propensity model is known, the AIPWE is thus always consistent. However, the AIPWE is more efficient when known propensity scores are replaced by corresponding estimates (eg, empirical proportions under simple randomization) (Tsiatis, 2006, p. 206). Given an AIPWE, the value search estimator maximizes its value across the search space defined by the family of treatment regimes of interest. For linear treatment regimes, this implies a maximization of the AIPWE across all possible values of the vector $\boldsymbol{\eta} = (\eta_0, \ldots, \eta_P)'$ in (1). Note that multiplying $\boldsymbol{\eta}$ by a positive real value leaves the treatment regime unchanged. This lack of identifiability is solved by normalizing $\boldsymbol{\eta}$ to unit norm, after which the search space corresponds to a unit hypersphere. Maximizing the AIPWE across this hypersphere is nontrivial, as the AIPWE value is a nonsmooth function of $\boldsymbol{\eta}$ (Tsiatis et al., 2020). Zhang et al. (2012) proposed a genetic algorithm (GA) to perform this maximization.

While quite a bit of theoretical work has been done on methods for OTR estimation (eg, Chakraborty and Murphy, 2014; Laber et al., 2014), applications of these methods to real-life data are relatively few in number. In this paper, we reanalyze data from an RCT for the treatment of dysthymic disorder (a milder but long-lasting form of depression) and apply methods for linear OTR estimation. In particular, we discuss the challenges that emerged in this analysis, with details on the choices we have made to deal with them. We also motivate these choices, discuss possible alternatives, and examine throughout whether—and if yes, why—these choices influence the final results. Through all this, we also aim to identify general best practices. As such, the paper may serve as a guide for OTR estimation in applied data analysis.

The remainder of this paper is structured as follows: In Section 2, we present the dataset focused on in this paper. Next, we highlight the challenges that emerged in the analysis of these data and how we plan to address them (Section 3). Section 4 presents the results of our analyses along with a discussion. We end the paper with a few general concluding remarks.

2 BROWNE DATA

We reanalyzed data from an RCT by Browne et al. (2002) on the treatment of dysthymia. In this trial, 707 adults with dysthymic disorder from a community-based primary care practice were randomized to treatment with either the antidepressant sertraline alone, interpersonal psychotherapy (IPT) alone, or the combination of sertraline and IPT. The study spanned 2 years, with continuous sertraline treatment throughout, and with IPT being administered only during the initial 6 months post-randomization. Regarding depressive symptoms after 6 months and 2 years, no difference was found between sertraline alone and sertraline plus IPT, whereas both were more effective than IPT alone.

To evaluate treatment effectiveness, patients were assessed at baseline, as well as at 6 months, 1 year, and 2 years after randomization, with regard to 5 measures of depression severity and bad quality of social functioning. Measures of depression severity were the Center for Epidemiologic Studies Depression Scale (CESD), the Montgomery-Åsberg Depression Rating Scale (MADRS), and a visual analogue scale (VAS) for mood. Bad quality of social functioning was assessed by means of the Social Adjustment Scale-self rating (SAS) and the McMaster Family Functioning Assessment Device (FAMFUN). Moreover, many other variables were also measured at baseline, including (biological) sex, age, income, marital status, perceived lack of health quality, and so on.

We focused on a subset of the available data. First, we focused on 2 treatment alternatives only: sertraline and sertraline plus IPT; IPT alone was not considered because, marginally, it was found to be inferior to the 2 other treatments. The search for an optimal regime involving the latter 2 treatments can further be justified from a substantive viewpoint because, although a combined treatment may outperform medication administration alone for certain subgroups of patients, the reverse may also hold for other subgroups given that some patients respond negatively to psychotherapy (see, eg, Strauss et al., 2021; Vybíral et al., 2024). Second, to evaluate treatment effectiveness, we focused on CESD because (i) depression severity was the primary target of interest, (ii) CESD is a reliable and validated self-reported measure of depression severity (Radloff, 1977), and (iii) CESD scores could be unambiguously extracted from the raw data files. We focused on the change in CESD between baseline and 6 months as outcome measure because IPT stopped at 6 months, while sertraline treatment continued throughout the entire study period.

Third, we selected 10 pretreatment characteristics for which it seemed clinically plausible that they are related to (differential) treatment effectiveness. These included sex, age, major depressive disorder (MDD) at present, MDD in the past, perceived lack of health quality, and the baseline measures of the 5 variables that captured depression severity and bad quality of social functioning (CESD, MADRS, VAS, SAS, and FAMFUN). Web Table 1 presents descriptive statistics for these characteristics in each of the 2 trial arms

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

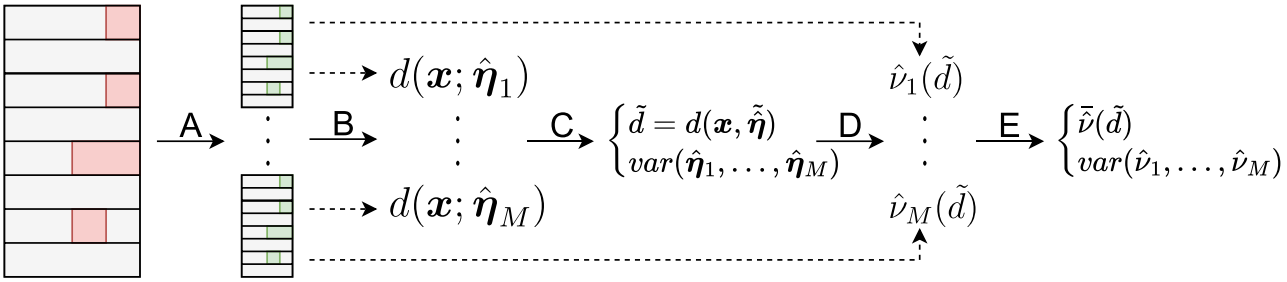


FIGURE 1 Steps in optimal treatment regime estimation in conjunction with multiple imputation. A: Creation of M imputed datasets, B: estimation of optimal treatment regime in each imputed dataset, C: aggregation of the M estimated regimes, D: estimation of the value of the summary regime in the M imputed datasets, E: aggregation of the M estimated values.

under study. We included all 10 variables in the main-effects part of the outcome models that we used in our analyses, both in QL and in the AIPWE of the regimes' expected (potential) outcome. To safeguard the simplicity of the estimated regimes, we included only a subset of the 10 variables in the interaction part of the outcome models and the search space of regimes in which we looked for an optimal member. For this purpose, we first subjected the 5 measures of depression severity and bad quality of social functioning to a principal component analysis (with VARIMAX rotation); this yielded a solution with 2 components: (i) mood, best represented by CESD, and (ii) social functioning, best represented by FAMFUN. From the remaining pretreatment characteristics, we selected 3 variables that did not overlap in meaning with mood or social functioning at baseline: sex (female), age, and MDD in the past (past MDD). We will thus only consider regimes of the following form:

$$I\left(\eta_0 + \eta_1 \texttt{female} + \eta_2 \texttt{age} + \eta_3 \texttt{past MDD} + \eta_4 \texttt{cesd} + \eta_5 \texttt{famfun} > 0\right), \tag{3}$$

where 1 corresponds to sertraline plus IPT and 0 to sertraline alone.

3 CHALLENGES IN THE ANALYSIS OF THE DATA AND HOW WE PLAN TO ADDRESS THEM

3.1 Missing data

The Browne data contain a considerable amount of missing values. First, out of the 476 patients in the sertraline alone and sertraline plus IPT groups, 23 patients dropped from the study before receiving any treatment. Second, among the 453 patients who started treatment, 45 patients had a missing CESD score at 6 months; in addition, at baseline, there are 1 and 4 missing values for the VAS and FAMFUN variables, respectively.

We handled these missing values in 2 ways. First, patients who did not receive any treatment were omitted from further analyses, based on the assumption that they had not yet been informed about their treatment assignment (which implies that the omission in question would not hamper comparative or causal inferences on the treatments under study); to be sure, Browne et al. (2002) do not provide clear information in support or contradiction of this assumption. Second, the remaining missing values were handled through multiple imputation (MI). This is a flexible and efficient approach enabling valid inferences under the assumption of missingness at random (MAR), provided the imputation model is correctly specified. Furthermore, MI can be combined with almost any complete-data estimator, obviating the need for more specialized methods. We can further assess the uncertainty due to missingness through the between-imputation variability. In addition, uncertainty due to within-imputation sampling variability can also be assessed if within-imputation standard errors are available (which is not the case for VSE). Finally, MI easily allows to also impute under various missingness mechanisms that deviate from MAR, thereby providing a relatively straightforward tool to perform sensitivity analyses.

3.2 Challenges in the data analysis

When estimating an OTR in conjunction with handling missing data via MI, the data analysis involves 3 basic steps: (i) MI, (ii) estimation of the optimal regime in each imputed dataset, and (iii) aggregation of the estimated regimes to obtain both a summary estimated regime and the uncertainty around it due to missingness. These 3 steps correspond to A, B, and C in Figure 1. In addition, we also care about the quality of the summary regime in terms of its expected potential outcome when applied to the target patient population; an estimate of this quality, along with the uncertainty around it due to missingness, involves a fourth basic step, corresponding to D + E in Figure 1.

In the analysis of the Browne data, significant challenges arose in each of these 4 steps. Below, we will successively discuss them, along with our proposed solutions.

3.2.1 Challenges in the MI

The 2 most popular imputation methods for missing data are imputation under a (multivariate normal) joint model (Schafer, 1997) and fully conditional specification (FCS; van Buuren, 2007). We choose imputation through FCS, as it only requires the specification of an appropriate univariate imputation model for each incomplete variable and therefore allows for more flexibility than a joint model. Furthermore, it has advantages in terms of computational stability. We use FCS as implemented in the MI procedure in SAS 9.4 (van Buuren, 2007).

For all imputation models in this paper, we incorporate baseline measurements of the 10 covariates selected in Section 2, along with measurements at 6, 12, and 24 months following randomization of the 5 variables that captured depression sever-

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

ity and bad quality of social functioning. Specifically, for every univariate model in the FCS procedure, the main effects are included for each of these variables. We use the PROC MI default of normal linear and logistic regression for continuous and categorical variables, respectively. To further capture the full range of uncertainty across imputations, we employ a substantial number of imputations, specifically, 200. These are obtained from 200 independent chains while discarding 100 burn-in iterations in each chain. Convergence of the chains was checked with trace plots.

Imputations based on a normal linear regression model can span the entire real line. The bounded continuous variables in the Browne data, such as the psychological scales, are therefore transformed to the real line during the imputation process. (Besides, one may wish the corresponding transformed distributions to be normal conditional on the other variables in the imputation model because we use normal linear regression for the continuous variables in the FCS procedure; one could check this by means of residual plots.) Let x be a variable in $[a, b]$ (eg, $[0, 60]$ for CESD). This variable is transformed to the real line as follows:

$$\tilde{x} = \log\left(\frac{x^*}{102 - x^*}\right), \tag{4}$$

where $x^* = (x - a) \cdot \frac{100}{b-a} + 1$. The imputed values are backtransformed to the original scale. There remains a small probability that values are imputed outside the bounds, that is, after backtransformation they can lie in $[a - \frac{b-a}{100}, a[$ or in $]b, b + \frac{b-a}{100}]$. Such values are truncated to the boundary values a or b, respectively. Other transformations are possible, but the truncation issue is inevitable because we have to map a closed interval to the open real line. For integer scales, backtransformed imputed values are also rounded to the nearest integer.

We consider 2 imputation models: a per-arm and a global imputation model. In the *per-arm* imputation model, the imputation model is estimated in each treatment arm separately; this is equivalent to a model estimated in the entire dataset that includes all 2-way interactions with treatment. In the *global* imputation model, the univariate models are the same as in the per-arm model, except that data from both treatment groups are combined, and that the treatment variable is included as a predictor. This model, therefore, does not allow for any interactions with treatment.

To make a choice between the 2 imputation models, we should recognize that in MI the imputations should be *proper* (Molenberghs and Kenward, 2007, pp. 145-150). An aspect of proper imputations that is especially relevant here is that the imputation model should be compatible with the modeling choices in the analysis of interest. In the search of OTRs, room should be left for nontrivial optimal regimes. We should, therefore, allow for treatment effect heterogeneity in the imputation model, as otherwise, we might fail to find a nontrivial optimal regime because the treatment effect heterogeneity was weakened in the imputed datasets. The per-arm imputation model allows for treatment effect heterogeneity, while the global imputation model does not. Hence, the per-arm imputation model should be preferred. In Section 4.2.1, we will illustrate the consequences of improper imputations based on a global imputation model.

Finally, MI is valid under MAR, which is an unverifiable assumption. Browne et al. (2002, p. 322) state that "the majority of individuals were unavailable for follow-up due to 'technical' non-medical reasons." While this provides some confidence in MAR, we can never be certain that it holds and, therefore, also conducted a sensitivity analysis for violations of MAR.

3.2.2 *Challenges in the optimal regime estimation*

QL is straightforward for this application, as it only requires the estimation of linear regression models. The estimated regime then follows directly from the estimated regression parameters. In our analyses, a linear regression model for the change in CESD is used, with main effects for the 10 pretreatment characteristics selected in Section 2 and with interaction effects for the covariates in (3). In the mth imputed dataset, the estimated regime parameters are the estimated main effect of treatment, $\hat{\eta}_{0,m}$, and the estimated interaction effects for the covariates in (3), $(\hat{\eta}_{1,m}, \ldots, \hat{\eta}_{5,m})'$.

Unlike QL, VSE is considerably more challenging. To look for a linear regime with maximal value (ie, maximal expected change in CESD) as estimated by the AIPWE, we used the implementation of the GA proposed by Zhang et al. (2012), which relies on the `optimalSeq()` function within the R-package `DynTxRegime` (Holloway et al., 2023). As an outcome model in the AIPWE, we used the same linear regression model as in QL. For the propensity scores, we reverted to the empirical proportions, as discussed in Section 1.

A GA is an optimization method that uses a set of randomized rules to evolve a finite population of trial solutions in such a way that each generation of trial solutions tends to be better than the previous one (Mebane and Sekhon, 2011). While GAs are versatile optimization methods, they are stochastic algorithms; hence, multiple runs with the same starting values but different seeds can lead to different solutions. Importantly, GAs may not always find the global optimum as required in statistical applications.

When using GAs, we must specify many implementation aspects. The suggested choices for several of these aspects vary considerably in the literature on VSE. We will explore the impact of these choices using the Browne data.

3.2.3 *Challenges in the aggregation of estimated regimes*

A next step is the aggregation of the estimated regimes from the different imputed datasets to capture both central tendency and the uncertainty around it. This is not fully trivial as a treatment regime is a decision rule, rather than a parametric model for which established rules of aggregation (Rubin's rules) are available. This becomes especially clear in VSE for linear treatment regimes, where the regimes can be represented by points on a unit hypersphere.

We chose the following approaches for aggregation. First, for QL, we applied Rubin's rules to the interaction part of the outcome models to obtain both central tendency (summary regime) and uncertainty information. The latter can be decomposed in terms of (i) the pooled within-imputation estimation variance and (ii) the between-imputation variability.

Second, we also aggregated unit-vector representations of the regimes estimated via QL and VSE (using the same type of

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

aggregation for both estimation methods so as to allow for better comparisons). Regarding central tendency, because the estimated regimes can be represented by points on a unit hypersphere, a natural candidate summary regime is obtained through the *spherical mean* (Mardia and Jupp, 2000, p. 163):

$$\tilde{\hat{\boldsymbol{\eta}}} = \frac{1}{\left\| \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\eta}}_m \right\|_2} \cdot \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\eta}}_m, \tag{5}$$

where $\| \cdot \|_2$ is the Euclidean norm.

As to uncertainty due to missingness, we propose to depict this via univariate density estimates of the distribution, across imputations, of each slope in $\hat{\boldsymbol{\eta}}$, and via pairwise scatter plots for all pairs of slopes in $\hat{\boldsymbol{\eta}}$ (ie, for all pairs of covariates in (3)). For VSE, these representations based on between-imputation variability reflect not only uncertainty due to missingness but also variability due to the stochastic nature of the GA.

3.2.4 Challenges in the estimation of the quality of the summary estimated regime

To assess the usefulness of the summary regimes, we also estimate their values, that is, the expected decrease in CESD if all patients were to receive treatment according to them. To obtain a single estimate for a summary regime's value, we first estimate the summary regime's value in each imputed dataset (step D in Figure 1) with the AIPWE. We use the same linear regression model as in QL as outcome model, and the empirical proportions instead of the known propensity scores. Second, we pool the estimated values across imputed datasets using Rubin's rules to obtain measures of central tendency and spread. Specifically, let $\hat{v}_1(\tilde{d}), \ldots, \hat{v}_M(\tilde{d})$ denote the estimated values of the summary regime, $\tilde{d}$, in the M imputed datasets, with corresponding estimated variances $V_1, \ldots, V_M$ based on the sandwich technique (Tsiatis et al., 2020, p. 86). The pooled estimated value of $\tilde{d}$ then equals

$$\bar{\hat{v}}(\tilde{d}) = \frac{1}{M} \sum_{m=1}^{M} \hat{v}_m(\tilde{d}), \tag{6}$$

with estimated variance

$$\bar{V} = \bar{V} + \left(1 + \frac{1}{M}\right) B, \tag{7}$$

where $\bar{V} = \frac{1}{M} \sum_{m=1}^{M} V_m$ represents the pooled within-imputation estimation variance and $B = \frac{1}{M-1} \sum_{m=1}^{M} \left(\hat{v}_m(\tilde{d}) - \bar{\hat{v}}(\tilde{d})\right)^2$ the between-imputation variance.

Note that the (within-imputation) value estimator, $\hat{v}_m(\tilde{d})$, suffers from "double dipping": The optimal regime and its value are estimated using the same data. Hence, $\bar{\hat{v}}(\tilde{d})$ may overestimate the true value of the summary regime, $\tilde{d}$. The estimator for an estimated regime's value is further nonregular (Tsiatis et al., 2020, pp. 537-539), which implies that standard asymptotics may not hold or may perform poorly in finite samples (Chakraborty and Murphy, 2014). We return to these issues in the discussion.

4 RESULTS

We first present the primary results of the analysis of the Browne data (Section 4.1). Next, we discuss imputation-related issues (Section 4.2) and issues related to the implementation of VSE (Section 4.3).

In the parts of Sections 4.1 and 4.2 pertaining to VSE, we use an optimal set of implementation choices for the GA that result from Section 4.3. In particular, we use starting values obtained from QL, and we retain the best result out of 5 runs of the GA with a population size of 3000 and 20 runs with a population size of 500.

4.1 Primary results for the Browne dataset

The estimated optimal linear regimes, summarized using the spherical mean, are

$$I(-0.987 - 0.014\texttt{female} + 0.010\texttt{age} + 0.048\texttt{past MDD} + 0.005\texttt{cesd} + 0.150\texttt{famfun} > 0), \tag{8}$$

$$I(-0.996 - 0.005\texttt{female} + 0.003\texttt{age} + 0.039\texttt{past MDD} + 0.022\texttt{cesd} + 0.079\texttt{famfun} > 0), \tag{9}$$

for QL and VSE, respectively. For QL, the estimated optimal regime after aggregation by Rubin's rules (and subsequent unit normalization) is

$$I(-0.988 - 0.001\texttt{female} + 0.010\texttt{age} + 0.045\texttt{past MDD} + 0.005\texttt{cesd} + 0.150\texttt{famfun} > 0). \tag{10}$$

The magnitude and direction of the effects agree moderately between the spherical mean summaries for QL and VSE, and strongly between the 2 types of summaries for QL. In general, the summary regimes imply that patients with a worse condition at baseline (ie, a higher CESD and FAMFUN and a history of MDD) are more likely to be assigned to sertraline plus IPT. This is a reasonable result, since this treatment alternative is more comprehensive than sertraline alone.

Concerning the variability of the estimated optimal regime slopes across the imputed datasets, the diagonal and off-diagonal plots in Figure 2 are, respectively, density estimates and scatter plots of the estimated regime slopes across imputations. Especially for VSE, there is much between-imputation variability in the estimated regime slopes.

As to the estimated value of the summary regimes, the mean values together with 95% confidence intervals (CIs) are represented in the Per Arm rows of the 2 upper left-hand panels of Figure 3.

Since the spherical mean summary regimes for QL and VSE are rather similar, one may expect their estimated values to be similar as well. This is the case, indeed: 12.77 (95% CI: [11.11, 14.43]) and 12.33 (95% CI: [10.73, 13.93]) for QL and VSE, respectively. Furthermore, VSE has to perform in each imputed dataset at least as well as QL in terms of the estimated value because (i) the value search estimator directly maximizes the estimated value of the regime, and (ii) we use starting values from QL. However, this advantage disappears at the level of the summary regimes. Finally, the left-hand panels of Figure 3 suggest that the estimated gains of the summary regimes over the

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

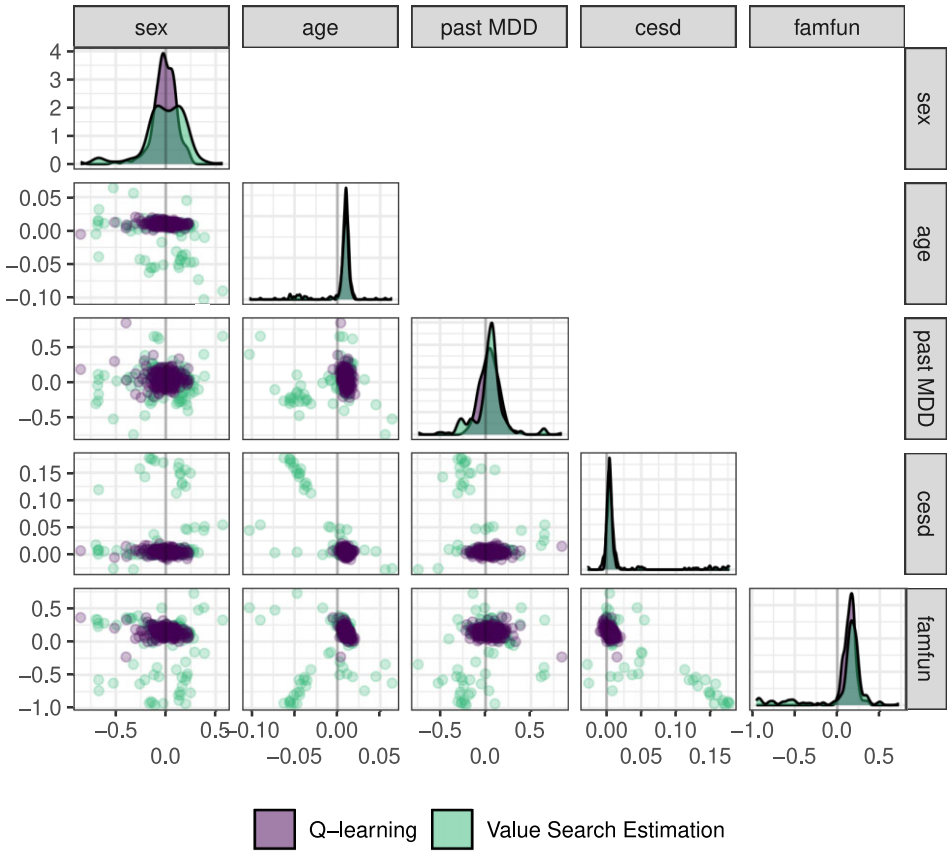


FIGURE 2 Pairwise scatter plots of the optimal regime slopes (denoted by their corresponding covariates) estimated in each imputed dataset. The plots on the diagonal are density estimates of the distributions of the estimated optimal regime slopes across the imputations.

best one-size-fits-all regime are very small. Hence, the advantage of a precision approach is close to negligible.

4.2 Imputation issues

4.2.1 Proper versus improper imputation

We already explained why proper imputation is necessary for a valid analysis. We will now illustrate this with the Browne data, where we compare the proper imputation per arm with the improper global imputation.

The estimated values of the summary regimes after the 2 types of imputation are summarized in the 2 upper left-hand panels of Figure 3. This reveals little difference between the global and per-arm imputation models. This is most likely due to the small amount of treatment effect heterogeneity in the Browne data.

To better demonstrate the importance of proper imputation, we repeat the analysis for a modified version of the Browne data in which we artificially induce a stronger treatment effect heterogeneity. Let Y_1 and Y_0 be the CESD outcome at 6 months and at baseline, respectively. Let X_3 and A further be indicator variables for past MDD and the sertraline plus IPT arm, respectively. The Browne data are then modified as follows:

$$\tilde{Y}_1 = Y_1 - \{0.70 \cdot (Y_0 - 40) + 6 \cdot X_3 - 6\} \cdot A + 0.20 \cdot (Y_0 - 40) + 4 \cdot X_3, \quad (11)$$

where a tilde indicates the modified value. Patients will thus benefit more from sertraline plus IPT (as compared to sertraline alone) if they have a more severe depression at baseline and if they already suffered from MDD in the past (larger Y_0 and X_3).

The results for the modified Browne data are summarized in the right-hand panels of Figure 3. Now, imputation per arm, which is proper, leads to summary regimes (obtained via the spherical mean and Rubin's rules) with fairly higher estimated values than improper global imputation. The magnitude of the difference agrees with the limited amount of missing data.

4.2.2 Imputation under MAR versus MNAR

So far, we have assumed MAR for the imputations. However, we cannot exclude missing not at random (MNAR). For instance, some patients may have withdrawn from the study because they were doing very well, or very poorly. Hence, a sensitivity analysis for deviations from MAR is warranted.

The first step in such a sensitivity analysis is to determine under which MNAR scenarios we will reanalyze the data. Given the possible reasons for study withdrawal, we will consider *constant shift* MNAR scenarios. These can be easily implemented in the MI procedure by adding a constant to each imputed value on the transformed CESD scale (using (4)) 6 months post-randomization (Carpenter and Kenward, 2014).

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

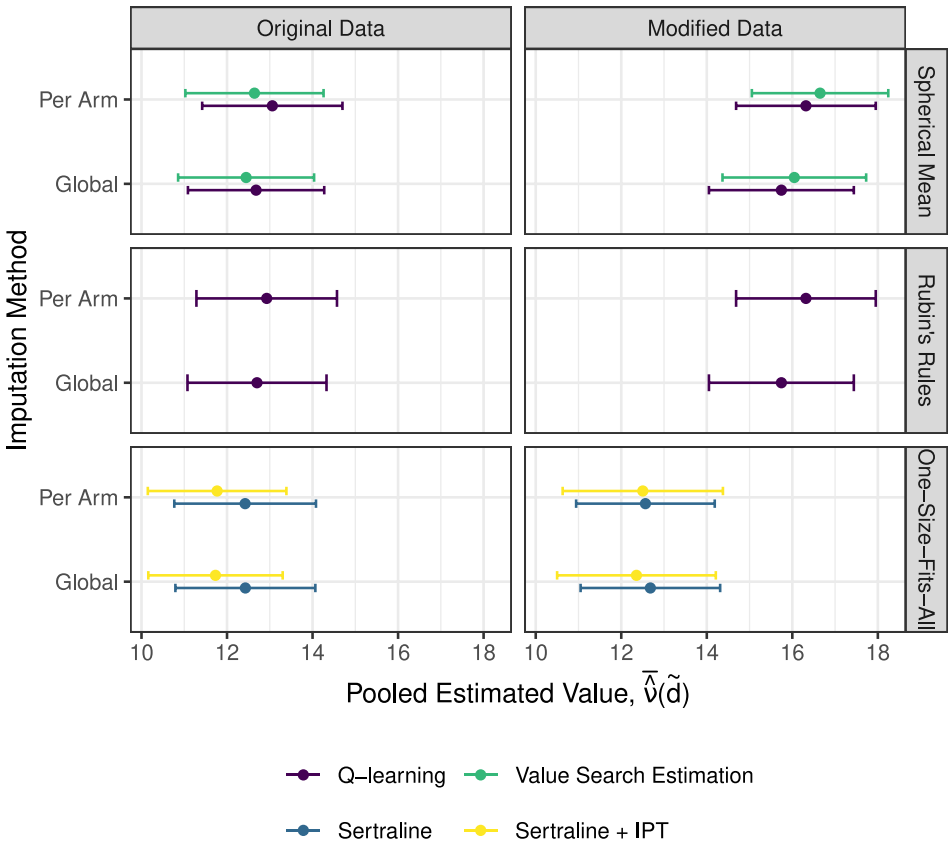


FIGURE 3 Pooled estimates for the value of the summary and one-size-fits-all regimes, together with 95% confidence intervals (with the summary regimes being obtained via the spherical mean or via Rubin's rules) for the original (left-hand panels) and modified (right-hand panels) Browne data, after proper (per-arm) and improper (global) imputation. The modified data (as introduced in Section 4.2.1) mimick the original data (as described in Section 2) but were simulated to contain more signal. The pooled estimates and confidence intervals have been obtained for each regime by applying Rubin's rules to the estimated values of that regime in all imputed datasets. The estimates and confidence intervals in question are also summarized in Web Tables 5 and 6.

While the shift is constant on the transformed CESD scale, it is not constant on the original one. Hence, we also compute the *average* shift on the original scale for each MNAR scenario:

$$\bar{\Delta}_{\text{MNAR}} = E(\text{CESD}_{\text{MNAR}}) - E(\text{CESD}_{\text{MAR}}), \qquad (12)$$

where the expectations are with respect to the patients with a missing CESD outcome at 6 months with $\text{CESD}_{\text{MNAR}}$ and CESD_{MAR} being the imputed values under MNAR and MAR, respectively.

After imputing under MNAR, the same analysis as before is repeated. In this sensitivity analysis, we only consider regimes estimated with QL because QL is computationally much less demanding than VSE and yielded less variable results for the original data.

We first assess the influence of the global shift on the estimated values of the summary regimes. Panel (A) of Figure 4 presents these along with 95% CIs. This plot also contains the estimated values of the one-size-fits-all regimes. A difference in the estimated value between the summary and the best one-size-fits-all regimes only shows up for relatively large positive shifts. Hence, the estimated lack of benefit of the summary regime over the best one-size-fits-all regime is largely insensitive to a global shift MNAR scenario.

In addition to assessing the sensitivity of the estimated values of the summary regimes to MNAR, it may also be useful to assess the regimes' sensitivity regarding their implied classifications of the patients. For this purpose, we computed Cohen's kappa between the classifications implied by the summary regimes under MNAR and by the summary regime under MAR. The results are represented in panel (B) of Figure 4, which shows that the classifications implied by the summary regimes are quite sensitive to deviations from MAR. This would be of little concern, however, if the discordant classifications would only pertain to patients for whom both treatment alternatives are about equally good. The latter would be the case if the importance of a discordant classification depends on the patient-specific (expected) contrast:

$$C_i = E(Y_1|\boldsymbol{X} = \boldsymbol{x}_i) - E(Y_0|\boldsymbol{X} = \boldsymbol{x}_i), \qquad (13)$$

where Y_0 (resp. Y_1) now is the potential outcome under treatment alternative 0 (resp. 1), and $\boldsymbol{x}_i$ are the observed baseline covariates for patient i. We estimate this contrast directly through the estimated regression function on which QL is also based:

$$\hat{C}_i = E(Y|A = 1, \boldsymbol{X} = \boldsymbol{x}_i; \hat{\boldsymbol{\beta}}) - E(Y|A = 0, \boldsymbol{X} = \boldsymbol{x}_i; \hat{\boldsymbol{\beta}}), \qquad (14)$$

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

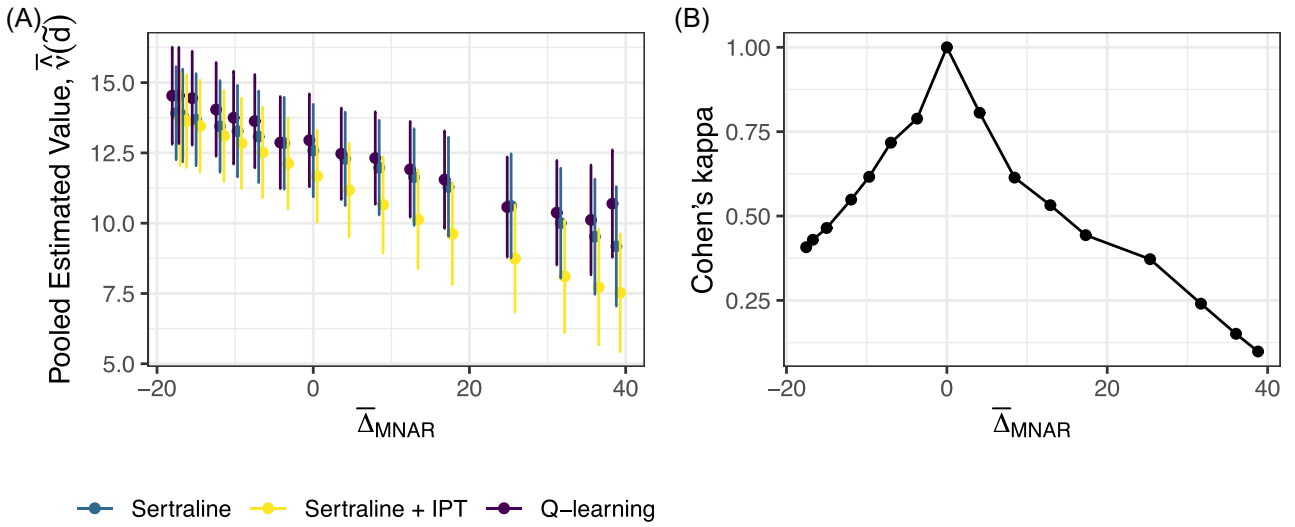


FIGURE 4 Results under various global shift missing not at random (MNAR) scenarios. The regimes have been estimated through Q-learning, normalized to unit norm, and aggregated in terms of their spherical mean. These results are based on 20 imputations for each MNAR scenario. (A) Pooled estimates for the value of the summary regimes together with 95% confidence intervals obtained by applying Rubin's rules. (B) Cohen's kappa for comparing the classifications of the summary regimes under MNAR with those of the summary regime under missingness at random (MAR). In the computation of Cohen's kappa, we used the same set of patients as in the other analyses, except the 5 patients with a missing value in the baseline covariates. $\bar{\Delta}_{\text{MNAR}}$: Average global shift on the CESD scale.

where $\hat{\boldsymbol{\beta}}$ is the estimated regression parameter vector. For each MNAR scenario and corresponding summary regime, we divide the patients into 2 groups: (i) patients with the same classification as under MAR and (ii) patients with a different classification. In Figure 5, we plot the absolute values of the estimated contrasts, $|\hat{C}_i|$, in histograms colored by group and stratified according to the MNAR scenario. This figure clearly shows that mainly patients with small estimated contrasts tend to be differentially classified.

4.3 Issues related to the implementation of VSE

Here, we will investigate several implementation aspects of the GA that are used for VSE. We will do so while using change scores for all 5 measures of depression severity and bad quality of social functioning in order to arrive at more general insights. Successively, we will discuss (i) the choice of starting values for the GA and (ii) the number of independent runs from a given set of starting values and the population size (a critical tuning parameter that represents the number of trial solutions in each generation).

4.3.1 Starting values

Zhang et al. (2012, sup. materials) discuss the choice of several implementation aspects of the GA; as part of this, they put the starting values to 0. The same is done by Tsiatis et al. (2020, dedicated website).

In Web Figures 1 and 3, we compare the estimated values for a fixed type of optimization approach (selecting the best result from 5 runs with a population size of 3000 and 20 runs with a population size of 500) with 3 different sets of starting values:

(1) starting values obtained from QL, normalized to unit norm;
(2) random starting values, independently sampled from a uniform distribution in $(-1, 1)$;
(3) all starting values set to 0.

Web Table 3 first shows that starting values determined by QL generally perform best. Second, the differences in quality of the estimated rules between the starting values depend on the outcome variable. Third, Web Figure 1 shows that there are some outcome-imputation combinations with little to no difference between the starting values. Still, all this does not detract from the rule of thumb to always go for starting values obtained from QL.

4.3.2 Population size and number of independent runs

A population size of 3000 is used by Zhang et al. (2012, sup. materials). In the code of Zhang and Zhang (2018), the population size varies from 700 to 3000, depending on the number of parameters in the regime. In the accompanying website for the Tsiatis et al. (2020) volume, usage of the GA is demonstrated with a population size of 1000. In all 3 cases, a single run from each starting configuration is used.

In Figure 6, we compare the performance of 7 optimization strategies in which the number of independent runs equals 1, 5, or 20, and the population size takes values of 500, 1000, and 3000. For computational reasons, the number of 20 runs was combined with a population size of 500 only. Starting values were obtained from QL, as these appeared to perform best in Section 4.3.1. In the multi-run scenarios, only the run with the highest estimated value is retained. In Web Appendix B, the

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

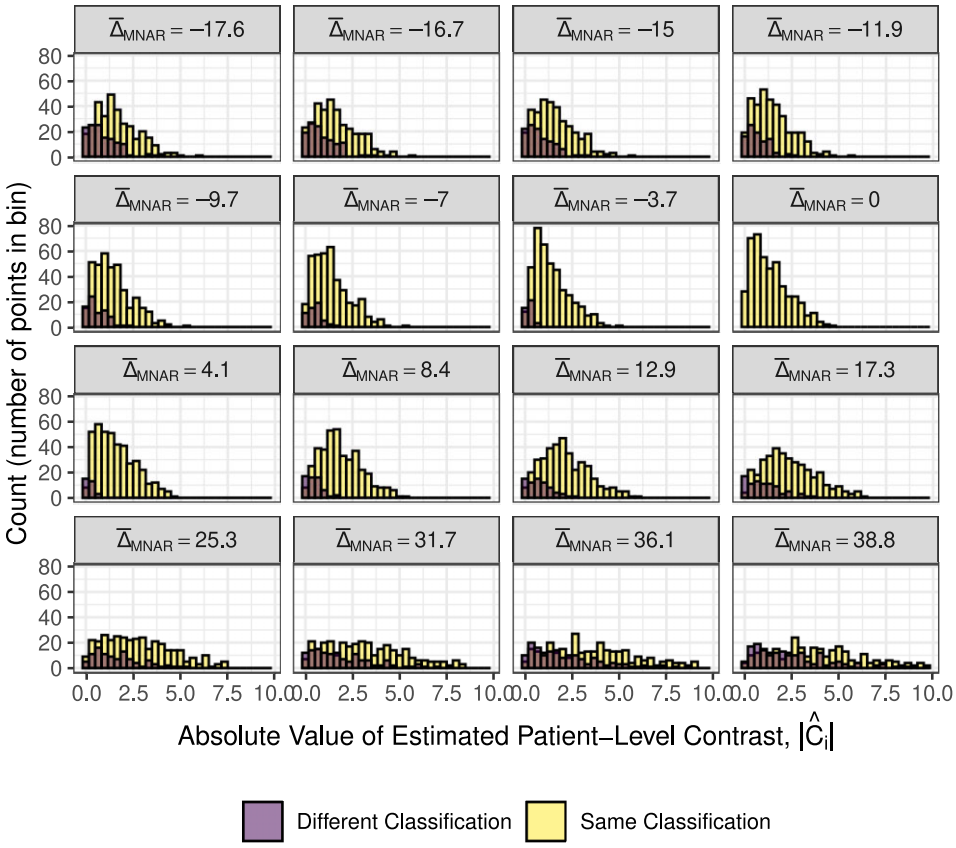


FIGURE 5 Distribution of the contrast estimates visualized in histograms. Each subplot shows the histogram for the estimated contrasts, colored according to whether the patient was differently classified as compared to the missingness at random (MAR) scenario. The subplot titles indicate the average shift in the corresponding MNAR scenario: $\bar{\Delta}_{MNAR}$. In each subplot, the 2 histograms have been superimposed.

setup of these optimization approaches is described in more detail.

Figure 6 leads to 3 important insights. First, for all outcomes, multi-run approaches outperform their single-run counterparts. Second, in some imputed datasets, the differences between the optimization approaches are very substantial, whereas in other ones they are small. Third, there is no approach that consistently leads to the highest estimated value, even when looking at each outcome separately.

Web Table 2 further summarizes the proportion of times each approach performs best across imputations by outcome. This table shows that the overall best-performing approach depends on the outcome variable. Specifically, 20 runs with a population size of 500 performs best for change in CESD and MADRS, whereas 5 runs with a population size of 3000 performs best for change in FAMFUN, SAS, and VAS. Therefore, all results reported in the previous sections of this paper are based on applying both approaches (with starting values obtained from QL) and selecting the best result among them.

5 CONCLUDING REMARKS

5.1 Dealing with missing values

In OTR estimation for the Browne data, we faced the ubiquitous problem of missing data. This was addressed by MI, which, while a method of choice, requires care in this context.

First, proper imputation is always required; within the context of OTR estimation, this implies that imputation per arm is needed.

Second, a standard implementation of MI relies on assuming MAR, which may not always be realistic in a clinical context. Hence, a sensitivity analysis to violations of MAR is warranted.

Third, MI implies the need to aggregate results across the imputed datasets. For linear treatment regimes, one can simply use Rubin's rules in case of QL based on parametric regression models. For VSE, however, Rubin's rules cannot be applied. To capture the central tendency in these estimated regimes, represented by unit vectors, nonstandard tools such as the spherical mean are required. Capturing spread is even less straightforward. Furthermore, aggregation is even more challenging for families of regimes other than the linear one (eg, decision trees).

5.2 OTR estimation

QL resides in the comfort zone of statisticians, as it merely relies on a regression methodology. However, from a principled viewpoint, there are good reasons to prefer value search for OTR estimation: It directly targets what one cares about, namely, retrieving a treatment regime with a maximal expected potential outcome. The primary target is different for QL, which is geared toward optimizing the fit between an outcome model and the data. Here, the treatment regime is a byproduct that is derived from the estimated outcome model. Moreover, whereas QL relies on

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

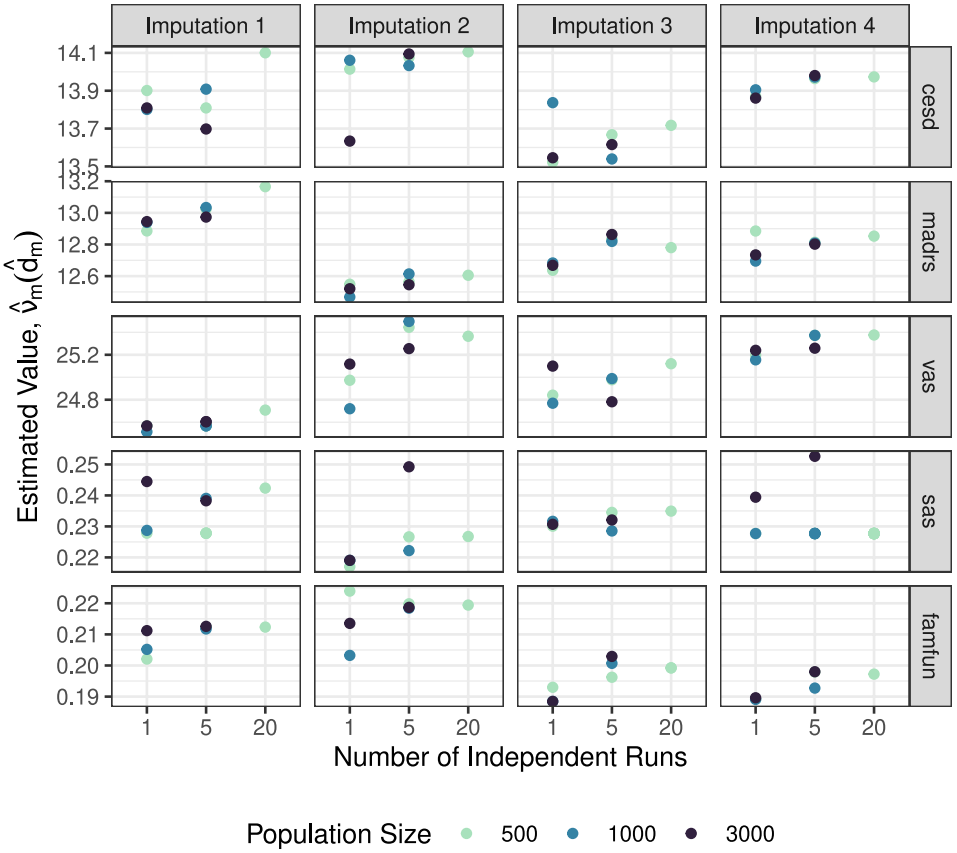


FIGURE 6 Effect of population size and number of runs on the performance of the genetic algorithm. The results are given for (the first) 4 imputed datasets with 5 different change scores as outcome variables. The *y*-axis represents the estimated value of the regime. Both the regime itself and its value are estimated in the *m*th imputed dataset; $\hat{v}_m(\hat{d}_m)$ is, therefore, indexed twice by the imputation number *m*. Starting values are obtained from Q-learning.

a correctly specified regression model, VSE relies on a doubly robust value estimator such as the AIPWE, which is consistent in RCTs even if the outcome model part is misspecified.

Nevertheless, running the GA requires a careful choice of several implementation aspects. Also, the GA is computationally expensive, and when it is run multiple times on each imputed dataset—as we recommend—the run-time on a personal computer quickly becomes prohibitive. Furthermore, the regimes estimated by VSE varied considerably across imputations. This might be due to the AIPWE (as a function of the regime parameters $\boldsymbol{\eta}$) being a nonsmooth and, hence, difficult objective function. Hence, the GA might often return locally optimal solutions. Anyhow, this variability implies considerable inferential uncertainty.

All this means that a large part of the principled advantages of VSE may melt away in empirical practice. Even more so as some of the limitations of QL (viz., its reliance on a correctly specified outcome model) can be overcome by using suitable model diagnostics (Taylor et al., 2015).

5.3 A two-fold limitation

We now address a two-fold limitation of the analyses presented in this paper that was already mentioned at the end of Section 3.2.4. First, the data are used twice: once for estimating OTRs and once for estimating the value of the OTRs. Second, the estimator of the value of an estimated regime is nonregular (Tsiatis et al., 2020).

The double use of the data may lead to overoptimism in the estimation of the value of the estimated regimes (and perhaps even more so when using VSE where the estimated value is maximized directly). In the Browne data, we did not find evidence for nontrivial OTRs that clearly outperform their one-size-fits-all counterparts; hence, correcting for a probable positive bias would not change our substantive conclusions. Yet, in general, such a correction is desirable.

The nonregularity of the estimator of the value of an estimated regime may lead to poor coverage of CIs based on standard asymptotic approximations, such as the sandwich technique. This may call for corrections, too.

As to possible ways out, Sies and Van Mechelen (2019) investigated several resampling methods that may correct for overoptimism of point estimators for the value of estimated regimes. Furthermore, Chakraborty and Murphy (2014) proposed an *m*-out-of-*n* bootstrap method for valid CIs for this value. In addition, Shen et al. (2023) studied the performance of the *m*-out-of-*n* bootstrap for QL in conjunction with MI. However, all these methods are computationally very expensive in combination with MI. Furthermore, simulations by Zhang et al. (2012)

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

suggest that standard asymptotic methods may yield CIs with adequate coverage, casting doubt about a strong need for complex resampling methods in practice. Still, there are also indications that standard asymptotic approximations may perform poorly when there is a sizeable subpopulation of patients for whom treatment alternatives are about equally effective (Tsiatis et al., 2020), which might be the case in the Browne data.

5.4 Best practices

A primary objective of this paper was the identification of best practices. At this point, one could consider such practices at the level of specific recommendations; for instance, one could recommend to do imputation per arm in OTR estimation based on RCT data. However, because an almost infinite number of obstacles and choices may show up in data-analytic practice, we rather want to derive best practices at a meta-level: Think through the different obstacles and choices that show up in an application, examine the alternatives, consider which elements may help in choosing among them, assess how choices influence the results, and make each time a well-motivated choice. In fact, the most important contribution of the present paper can be considered to showcase how these meta-level best practices can be implemented in the specific context of OTR estimation.

One may further wonder to what extent the best practices in question are already being used in applied OTR research. Yet, the overwhelming majority of work on OTR estimation consists of methodological contributions (with or without illustrative applications), with instances of pure applications being rare (as suggested by a quick PubMed search with the search term "optimal treatment regime," which yielded 53 relevant hits, 49 of which pertained to methodological work and only 4 to pure applications). Anyhow, we hope that this paper clarified how in pure applications inevitable obstacles, often insufficiently addressed in the methodological literature, can be dealt with, and that this is challenging but doable.

5.5 Extensions

Extensions of the approach proposed in the current paper to other outcome types such as discrete ones (eg, Moodie et al., 2024) are possible. For instance, we used FCS for MI; this method can handle any type of missing variables provided a correct specification of the conditional distributions. Also, the linear models used in QL and in VSE can be replaced with generalized linear models for non-continuous outcome types. The latter extension has also been implemented in the *DynTxRegime* R package (Holloway et al., 2023). It is uncertain, however, how the GA will behave for non-continuous endpoints.

The approach taken in the present paper can also be extended to multistage dynamic treatment regimes (see, eg, Tsiatis et al., 2020). Indeed, QL and VSE as implemented in *DynTxRegime* can be used to estimate a multistage regime in each imputed data set, leading to one decision function for each stage. These can subsequently be aggregated stage-wise over the imputed data sets following the principles used in the current paper. To be sure, a multistage context with missing data also poses additional challenges at the level of imputations (Shortreed et al., 2014).

In general, the approach taken here can be extended in various directions. For each extension, though, one should not lose sight of the best practices at the meta-level.

ACKNOWLEDGMENTS

We thank Dillon Browne for sharing the Browne data with us.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices, Tables, and Figures referenced in Sections 2 and 4 are available with this paper at the Biometrics website on Oxford Academic. All code is available from the same website and from github.com/florianstijven/Browne-data-OTR/releases/tag/v0.0.1, including a README file and a simulated data set based on the Browne data.

FUNDING

This research was supported by the Research Foundation Flanders (G080219N). F.S. received funding from the Agentschap Innoveren & Ondernemen and Janssen Pharmaceutica (HBC.2022.0145). E.D. received funding from the Dutch Research Council (016.Veni.195.215 6806). The computational resources used in this work were provided by the Flemish Supercomputer Center, funded by the Research Foundation Flanders and the Flemish Government.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

The Browne data that we reanalyzed in this paper are available from Radboud University for use by other researchers. Restrictions apply to the availability of these data, which were used under license in this paper. Requests can be made with the relevant author (ellen.driessen@ru.nl), who will seek approval from the original trial investigator.

REFERENCES

Browne, G., Steiner, M., Roberts, J., Gafni, A., Byrne, C., Dunn, E. et al. (2002). Sertraline and/or interpersonal psychotherapy for patients with dysthymic disorder in primary care: 6-month comparison with longitudinal 2-year follow-up of effectiveness and costs. *Journal of Affective Disorders*, 68, 317–330.

Carpenter, J. R. and Kenward, M. G. (2014). Sensitivity analysis with multiple imputation. In: *Handbook of Missing Data Methodology* (eds. Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A. and Verbeke, G.), 435–469. Boca Raton, FL: CRC Press. Chapter 19.

Chakraborty, B. and Moodie, E. E. (2013). *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*, vol. 76, New York: Springer.

Chakraborty, B. and Murphy, S. A. (2014). Dynamic treatment regimes. *Annual Review of Statistics and its Application*, 1, 447–464.

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

Holloway, S. T., Laber, E. B., Linn, K. A., Zhang, B., Davidian, M. and Tsiatis, A. A. (2023). *DynTxRegime: Methods for Estimating Optimal Dynamic Treatment Regimes*. R package version 4.14.

Kosorok, M. R. and Moodie, E. E. (2015). *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*. Alexandria: SIAM.

Laber, E. B., Lizotte, D. J., Qian, M., Pelham, W. E. and Murphy, S. A. (2014). Dynamic treatment regimes: technical challenges and applications. *Electronic Journal of Statistics*, 8, 1225–1272.

Li, Z., Chen, J., Laber, E., Liu, F. and Baumgartner, R. (2023). Optimal treatment regimes: a review and empirical comparison. *International Statistical Review*, 91, 427–463.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Chichester: Wiley.

Mebane, W. R. and Sekhon, J. S. (2011). Genetic optimization using derivatives: the rgenoud package for R. *Journal of Statistical Software*, 42, 1–26.

Molenberghs, G. and Kenward, M. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley and Sons.

Moodie, E. E. M., Bian, Z., Coulombe, J., Lian, Y., Yang, A. Y. and Shortreed, S. M. (2024). Variable selection in high dimensions for discrete-outcome individualized treatment rules: reducing severity of depression symptoms. *Biostatistics*, 25, 633–647.

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385–401.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall/CRC.

Shen, J., Hubbard, R. A. and Linn, K. A. (2023). Estimation and evaluation of individualized treatment rules following multiple imputation. *Statistics in Medicine*, 42, 4236–4256.

Shortreed, S. M., Laber, E., Scott Stroup, T., Pineau, J. and Murphy, S. A. (2014). A multiple imputation strategy for sequential multiple assignment randomized trials. *Statistics in Medicine*, 33, 4202–4214.

Sies, A. and Van Mechelen, I. (2019). Estimating the quality of optimal treatment regimes. *Statistics in Medicine*, 38, 4925–4938.

Strauss, B., Gawlytta, R., Schleu, A. and Frenzl, D. (2021). Negative effects of psychotherapy: estimating the prevalence in a random national sample. *BJPsych Open*, 7, e186.

Taylor, J. M. G., Cheng, W. and Foster, J. C. (2015). Reader reaction to "A robust method for estimating optimal treatment regimes" by Zhang et al. (2012). *Biometrics*, 71, 267–273.

Tran, T. D., Abad, A. A., Verbeke, G., Molenberghs, G. and Van Mechelen, I. (2023). Reflections on the concept of optimality of single decision point treatment regimes. *Biometrical Journal*, 65, 2200285.

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.

Tsiatis, A. A., Davidian, M., Holloway, S. T. and Laber, E. B. (2020). *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. Boca Raton, FL: CRC Press.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, 16, 219–242.

Vybíral, Z., Ogles, B. M., Řiháček, T., Urbancová, B. and Gocieková, V. (2024). Negative experiences in psychotherapy from clients' perspective: a qualitative meta-analysis. *Psychotherapy Research*, 34, 279–292.

Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, 68, 1010–1018.

Zhang, B. and Zhang, M. (2018). Variable selection for estimating the optimal treatment regimes in the presence of a large number of covariates. *The Annals of Applied Statistics*, 12, 2335–2358.

Downloaded from https://academic.oup.com/biometrics/article/81/1/ujaf026/8087929 by KU Leuven Libraries user on 25 April 2025

Received: April 24, 2024; Revised: February 18, 2025; Accepted: March 4, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of The International Biometric Society. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com