Estimation of Optimal Treatment Regimes in the Browne Data

Florian Stijven

2023-08-17

Table of Contents

[1 Comparison of Imputation Models 2](#_Toc143194111)

[1.1 Estimated Values of Estimated Optimal Regimes 2](#_Toc143194112)

[1.2 Distance Between Estimated and One-Size-Fits-All Regimes 4](#_Toc143194113)

[1.3 Coefficients of Determination in Outcome Regression Models 4](#_Toc143194114)

[2 Uncertainty due to Missing Data 5](#_Toc143194115)

[2.1 Estimated Regime Parameters 6](#_Toc143194116)

[2.2 Classification of Subjects 8](#_Toc143194117)

[3 Comparison of Value Search Estimator and Q-learning 9](#_Toc143194118)

[3.1 Estimated Values of Estimated Regimes 9](#_Toc143194119)

[3.2 Distance Between Estimated and One-Size-Fits-All Regimes 10](#_Toc143194120)

[4 Aggregating Estimated Treatment Regimes 11](#_Toc143194121)

[4.1 Aggregation Methods 11](#_Toc143194122)

[4.2 Estimated Value of Aggregated Regimes on Imputed Data Sets 12](#_Toc143194123)

[4.3 Distance of Aggregated Regimes from One-Size-Fits-All 13](#_Toc143194124)

[4.4 Inference on Estimated Value of Aggregated Regime 15](#_Toc143194125)

[4.5 Interpretation of the Aggregated Regimes 16](#_Toc143194126)

[5 Details on the Value Search Estimator 21](#_Toc143194127)

In this technical report, the results of the analysis of the Browne data are reported. This document is structured according to four main points:

* **(Issues with the Genetic Algorithm.)** This issue is discussed and empirically examined in a separate document. The results in that document justify certain choices made for the implementation of the genetic algorithm on which the value search estimator in DynTxRegimes relies.
* **Need for an Appropriate Imputation Model** We compare the results under two different imputation models: (i) imputation per arm separately and (ii) global imputation without interaction terms with the treatment variable. Note that the imputation models are based on the full conditional specification approach that is implemented in SAS. These two imputation models are compared in two metrics:
  + Estimated values of the estimated regimes.
  + Distance between the estimated regimes and a one-size-fits-all regime.
* **Uncertainty due to missing data.** This uncertainty is explored by looking at the estimated regime parameters, and the classification of individual subjects. We also compare the degree of uncertainty between Q-learning and value search estimation.
* **Comparison of value search estimation and Q-learning.** These methods are compared by looking at the estimated values of the estimated regimes.
* **Aggregating Estimated Regimes.** The estimated regimes should be aggregated across the imputations. We examine the following 3 approaches to aggregation across the imputed data sets:
  + Direct application of Rubin’s rules
  + Circular mean
  + Majority vote

At the end of this document, details on the exact implementation of the value search estimator are given. This implementation is used throughout this document.

# 1 Comparison of Imputation Models

In this section, the results under the two imputation models are compared. First, we compare the estimated values of the estimated optimal regimes between both imputation approaches. Second, we quantify how far away the estimated regimes are from a one-size-fits-all regime. We only consider analyses based on the multi-run value search estimator in this section. Finally, we look at the R-squared values of the outcome regression model that are used in both value search estimation and Q-learning.

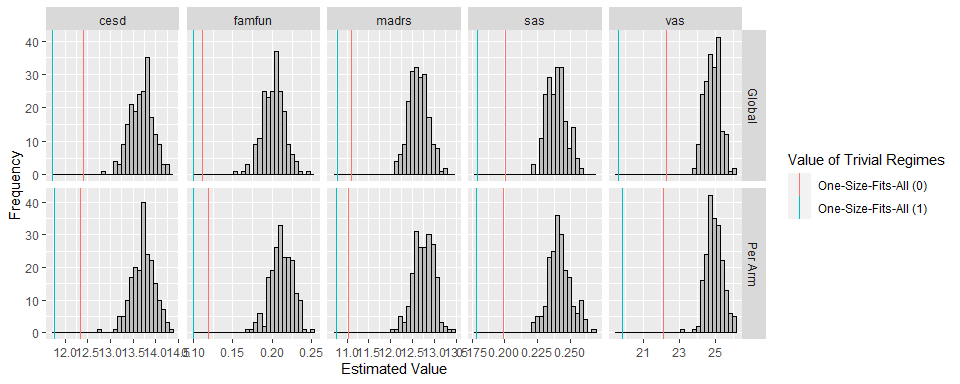
Note that we summarize results regarding the estimated regimes (and derived quantities) in each imputed data set. So, the estimated regimes for the imputed data sets are *not* aggregated first into a pooled estimate of the optimal treatment regime. The conclusions drawn from the results of this section do not necessarily translate to the corresponding pooled optimal treatment regimes.

## 1.1 Estimated Values of Estimated Optimal Regimes

The *estimated* values of the *estimated* optimal treatment regimes are compared between the two imputation models for the change from baseline outcomes. Note that “change from baseline” is actually the outcome at baseline minus the outcome at 6 months. Only for VAS, it is the other way around. This ensures that larger values of the outcome variable are better.

In the next histogram, we summarize the estimated values across the imputed data sets, stratified by imputation model. The variability within the same imputation method is thus merely due to the missing data. Sampling variability is ignored in these plots. We also add vertical lines that correspond to the pooled estimated values of the “one-size-fits-all” treatment regimes. These latter values are estimated by the same AIPW estimator as the one used by the value search estimator. The “pooling” is done by averaging the estimated values across all imputations.

These histograms show very small differences in the distributions of the estimated values for the two imputation models. This is confirmed by looking at the corresponding means in the next table.



Frequency distribution of the estimated values of the estimated regimes across the the imputations. Note that each histogram represents 200 estimated values of 200, possibly different, estimated regimes. The value is estimated by the AIPW estimator which is explained at the end of this document.

The next table shows the average of the estimated values across the imputed data sets for all outcomes. For all outcomes, the average estimated value is larger under imputation per arm than under global imputation. However, the difference is always very small.

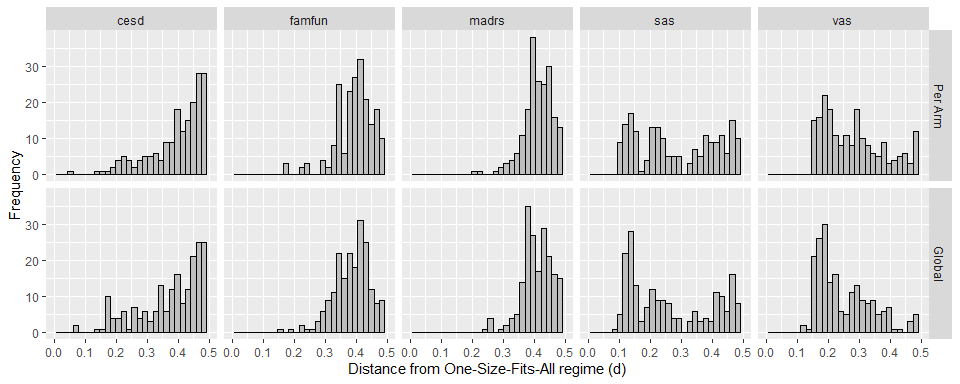
Average estimated value of the estimated regime across the imputed data sets. Note that each value is the average of 200 estimated values of 200, possibly different, estimated regimes. The value is estimated by the AIPW estimator which is explained at the end of this document.

| outcome | Per Arm | Global |
| --- | --- | --- |
| cesd | 13.725 | 13.678 |
| famfun | 0.211 | 0.202 |
| madrs | 12.758 | 12.645 |
| sas | 0.242 | 0.240 |
| vas | 24.961 | 24.841 |

A possible explanation for the small difference between the two imputation models could be that there is little treatment effect heterogeneity to begin with. In this case, both (estimated) imputation models will be very similar. Additionally, there will be less uncertainty in the global imputation model since that imputation model is estimated using observations from all treatment groups combined.

## 1.2 Distance Between Estimated and One-Size-Fits-All Regimes

The “distance” between an estimated regime and a one-size-fits-all regime is quantified by where and are the proportions of patients classified to each treatment. Indeed, if then the estimated regime corresponds to a one-size-fits-all regime. In what follows, we compare the distribution of these distances for both imputation models, stratified by the outcomes.



Frequency distribution of the distance from one-size-fits-all regimes for the estimated regimes across the imputations.

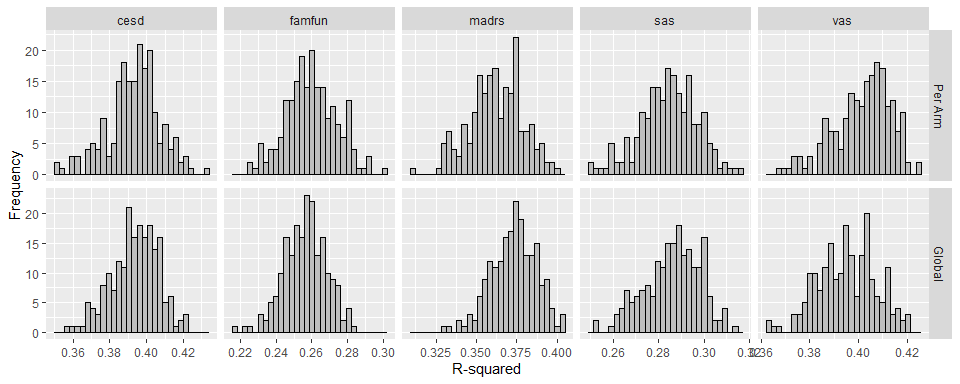
At first sight, there seems to be little difference in the distribution of between both imputation models except for VAS. The estimated regimes for VAS under global imputation are closer to one-size-fits-all regimes than under imputation per arm. The corresponding mean distances are given in the following table. This table again shows that the estimated regimes under global imputation are closer to a one-size-fits-all regime. Although, the difference is very small.

Average distance from one-size-fits-all of the estimated regimes across the imputed data sets. Note that each value is the average of 200 distances of 200, possibly different, estimated regimes.

| outcome | Per Arm | Global |
| --- | --- | --- |
| cesd | 0.402 | 0.378 |
| famfun | 0.395 | 0.386 |
| madrs | 0.414 | 0.412 |
| sas | 0.294 | 0.270 |
| vas | 0.283 | 0.252 |

## 1.3 Coefficients of Determination in Outcome Regression Models

As an additional way to examine the influence of the imputation model, we will look at the R-squared values for the outcome regression models that were used in Q-learning and the value search estimator. In the next histograms, we summarize the R-squared values across the imputed data sets by imputation model and outcome. The following table summarizes the corresponding means.



Frequency distribution of the R-squared values of the outcome regression models in each imputed data set. The exact formulation of this outcome regression model can be found at the end of this document.

Average of the R-squared values for the outcome regression models that were fitted in each imputed data set. The exact formulation of this outcome regression model can be found at the end of this document.

| outcome | Per Arm | Global |
| --- | --- | --- |
| cesd | 0.393 | 0.394 |
| famfun | 0.260 | 0.256 |
| madrs | 0.362 | 0.373 |
| sas | 0.284 | 0.286 |
| vas | 0.402 | 0.395 |

There is very little difference in the R-squared values between the two imputation models. This could also explain the lack of large differences between the two imputation models in terms of the estimated regimes and associated measures.

# 2 Uncertainty due to Missing Data

The histograms for the estimated values of the estimated regimes above show that there is some uncertainty in the estimated value of the estimated optimal treatment regimes across the imputed data sets. This variability arises from the missing data through 2 pathways.

1. The estimator for the value of a given regime differs between imputed data sets. Indeed, the estimated outcome regression model, on which the augmented inverse probability weighted estimator relies, differs between the imputed data sets.
2. The estimated optimal regime can also differ between imputed data sets.

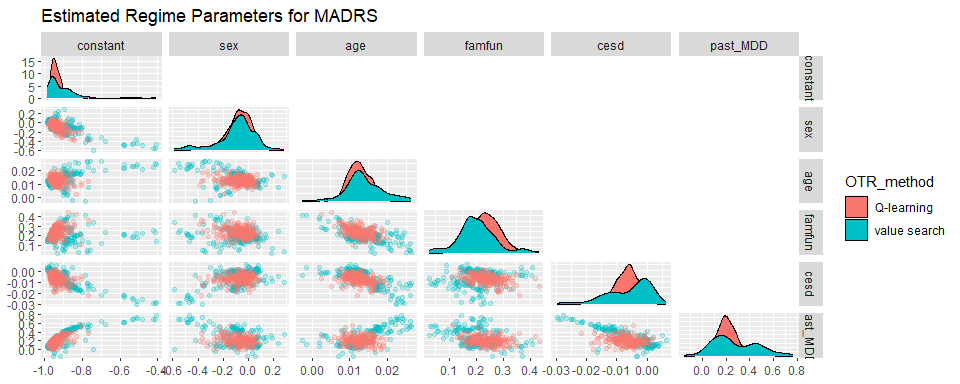
Note that we only show the results for imputation per arm. In this section, two aspects of the uncertainty in the estimated regimes are considered. First, we explore the variability in the estimated regime parameters. Second, we explore the variability in the classification of patients.

## 2.1 Estimated Regime Parameters

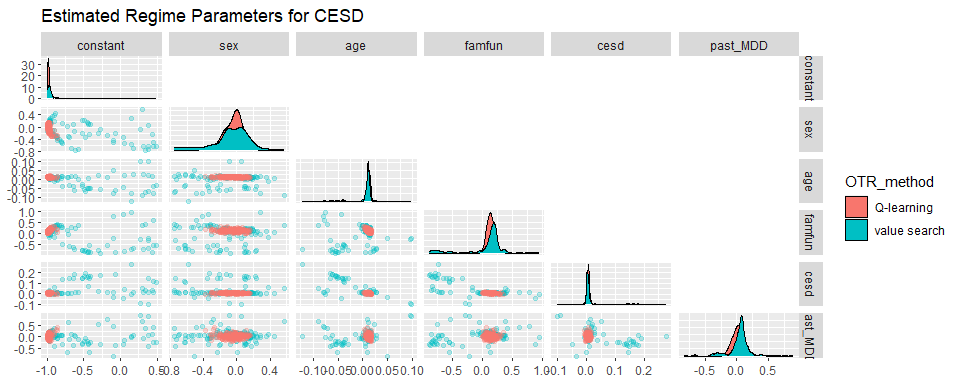
Variability in the estimated regime parameters across the imputed data sets is summarized next. We only look at the results under imputation per arm. We normalize the estimated parameter vectors to have unit norm. This is necessary because regimes are equivalent up to multiplication of all parameters with a constant. Note that treatment 1 is “Sertraline + IPT” and treatment 0 is “Sertraline only”.

The linear treatment regimes contain six parameters of which “constant” is the intercept. The next plots are sets of pairwise scatter plots for all pairs of these six estimated parameters. The diagonal plots are density estimates of the distribution of the corresponding estimates.

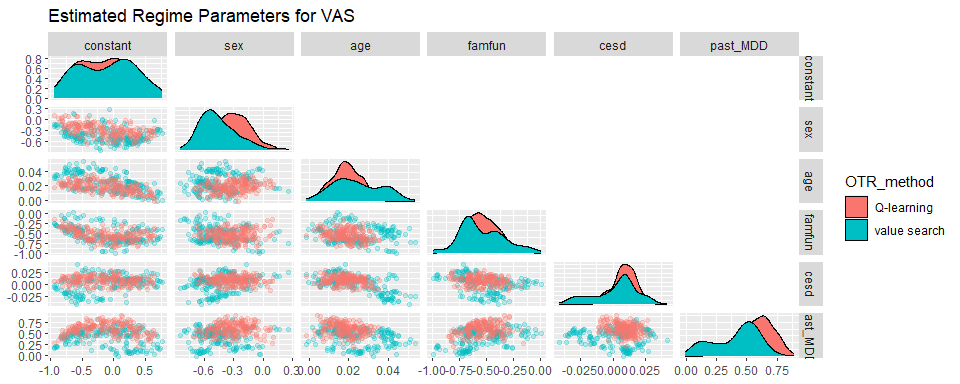
These plots show that there is a considerable amount of uncertainty in the estimated regime parameters due to missing data. In addition, these plots also show that there is considerably more uncertainty in the regime parameter estimates under value search estimation than under Q-learning.



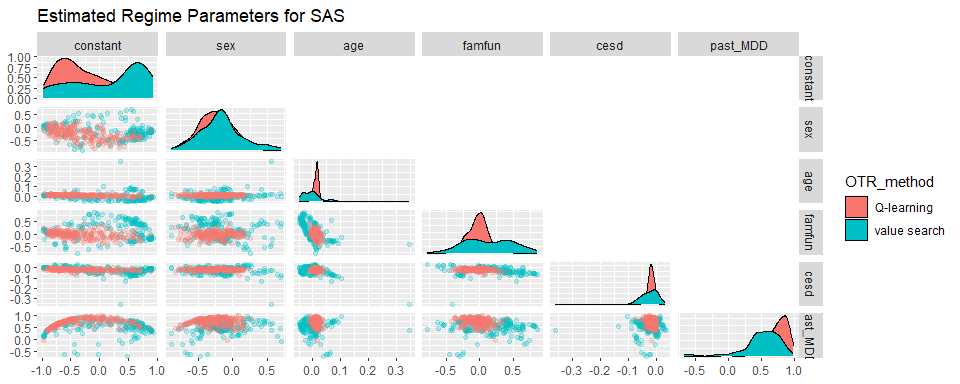
Pairwise scatterplots of the estimated regime parameters across the imputations for change in MADRS as outcome. Note that only the results under imputation per arm are shown here.



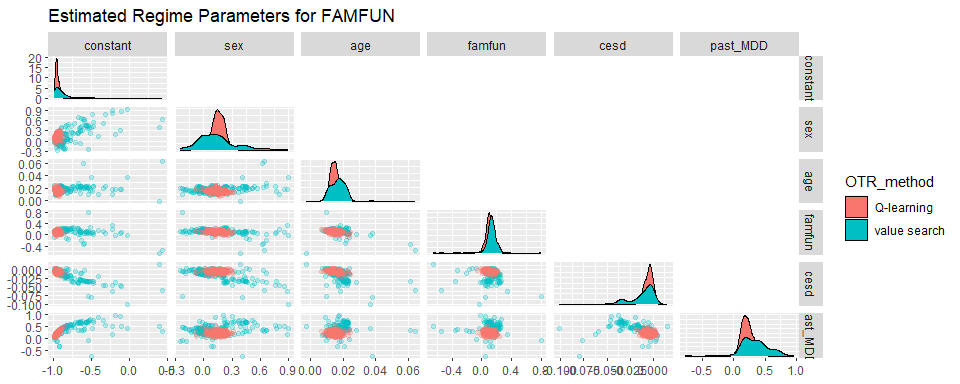
Pairwise scatterplots of the estimated regime parameters across the imputations for change in CESD as outcome. Note that only the results under imputation per arm are shown here.



Pairwise scatterplots of the estimated regime parameters across the imputations for change in VAS as outcome. Note that only the results under imputation per arm are shown here.



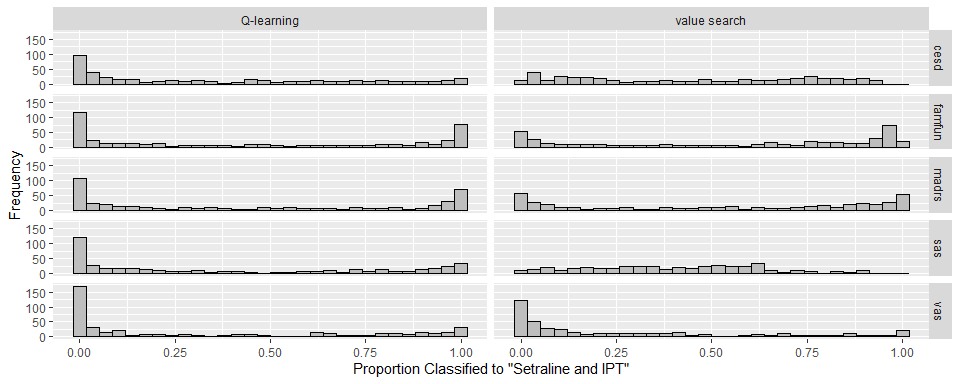
Pairwise scatterplots of the estimated regime parameters across the imputations for change in SAS as outcome. Note that only the results under imputation per arm are shown here.



Pairwise scatterplots of the estimated regime parameters across the imputations for change in FAMFUN as outcome. Note that only the results under imputation per arm are shown here.

## 2.2 Classification of Subjects

Even though the estimated regime parameters differ across imputations, this does not mean that the classifications will differ (considerably). Therefore, in the next plots, we summarize the proportion of times a particular subject is classified to “Sertraline and IPT”. This proportion is computed over the 200 imputed data sets.



Frequency distributions of the proportions each subject is classified to “Sertraline and IPT” accross the estimated regimes in the imputed data sets. Each proportion thus correponds to a single patient in the data, and is computed over the 200 imputed data sets. Note that only the results under imputation per arm are shown here.

The above histograms show that the classification for a considerable amount of patients is uncertain as a consequence of the missing data. The proportion of subjects that is consistently classified to the same treatment is small for most outcomes. The uncertainty in the estimated regime parameters thus translates into uncertainty regarding the classification of individual patients.

The larger degree of uncertainty in the regime parameters under value search estimation also translates to the classification of patients. There is less mass around 0 and 1 in the above histograms under value search estimation for all outcomes.

One of the reasons why there is a large amount of variability in the patient classifications between imputations could be that there is little treatment effect heterogeneity to begin with. In this case, the estimated value would be relatively insensitive to changes in the regimes. The objective function would thus be relatively flat in the regime parameters. Consequently, slightly different imputations could tip the classification in the other direction.

Also, it is not unexpected that there are many patients for which the classification is ambiguous. It is very well possible that for many patients, both treatments are roughly equivalent. Still, identifying the subset of patients where one of the treatments is much better than the other remains relevant. Even if this subset is relatively small. Depending on the outcome, there are still patients that are consistently classified into the same treatment.

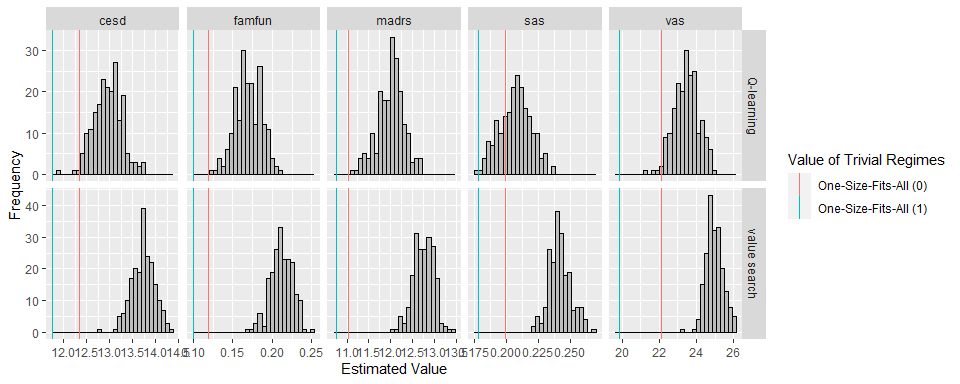
# 3 Comparison of Value Search Estimator and Q-learning

## 3.1 Estimated Values of Estimated Regimes

The estimated values of the estimated regimes are compared between the value search estimator and Q-learning. Note that the latter is just a linear regression model with interactions terms between treatment and (a subset of) the baseline covariates. Note that we only present the results under imputation per arm in this section.

Note that we summarize results regarding the estimated regimes (and derived quantities) in each imputed data set. So, the estimated regimes for the imputed data sets are not aggregated first into a pooled estimate of the optimal treatment regime. The conclusions drawn from the results of this section do not necessarily translate to the corresponding pooled optimal treatment regimes.

The next plots summarize the estimated values across the imputed data sets stratified by outcome and estimation method. Note that the estimator for the value of a given regime is the same AIPW estimator for Q-learning and value search estimation. The details on this AIPW estimator are given at the end of this document.



Frequency distributions of the estimated values for the estimated regimes across the imputations. The vertical lines correspond to the pooled estimates for the value of the one-size-fits-all regimes. The values are estimated by the AIPW estimator which is explained at the end of this document. Note that only the results under imputation per arm are presented here.

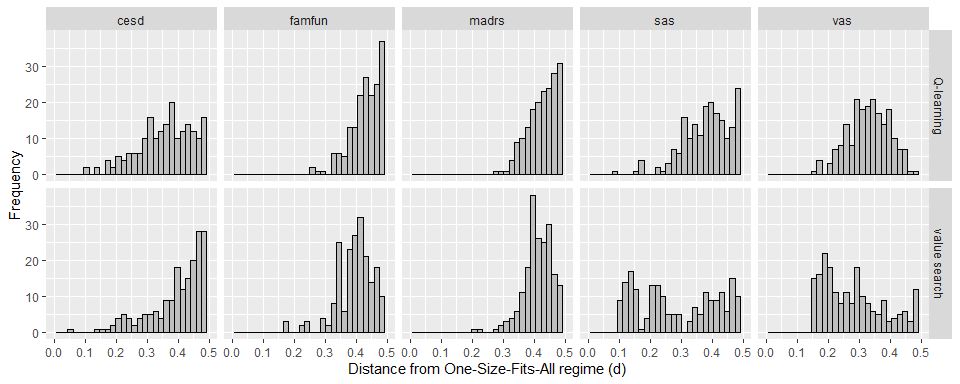
The next table shows the average of the estimated values across the imputed data sets stratified by outcome and estimation method. All these outcomes are a change from baseline where larger values are desired.

Average estimated value of the estimated regimes across the imputed data sets. The value is estimated by the AIPW estimator which is explained at the end of this document. Note that only the results under imputation per arm are presented here.

| outcome | Q-learning | value search |
| --- | --- | --- |
| cesd | 12.961 | 13.725 |
| famfun | 0.170 | 0.211 |
| madrs | 12.007 | 12.758 |
| sas | 0.207 | 0.242 |
| vas | 23.454 | 24.961 |

## 3.2 Distance Between Estimated and One-Size-Fits-All Regimes

We again use to quantify the “distance” between an estimated regime and a one-size-fits-all regime. In what follows, we compare the distribution of these distances for the regimes estimated by the value search estimator and Q-learning, stratified by the outcomes. We only present the results under imputation per arm.



Frequency distribution of the distance from one-size-fits-all regimes for the estimated regimes across the imputations. Note that only the results under imputation per arm are presented here.

There are clear differences in the distribution of between value search estimation and Q-learning. Q-learning leads to higher values of except for CESD. This indicate that the corresponding estimated regimes are further away from one-size-fits-all regimes than for the value search estimator. However, the regimes further away from one-size-fits-all regimes do not lead to higher estimated values.

The corresponding mean distances are given in the following table.

Average distance from one-size-fits-all of the estimated regimes across the imputed data sets. Note that only the results under imputation per arm are presented here.

| outcome | Q-learning | value search |
| --- | --- | --- |
| cesd | 0.361 | 0.402 |
| famfun | 0.435 | 0.395 |
| madrs | 0.432 | 0.414 |
| sas | 0.384 | 0.294 |
| vas | 0.332 | 0.283 |

# 4 Aggregating Estimated Treatment Regimes

## 4.1 Aggregation Methods

In this section we present the results after aggregating the estimated optimal regimes across the imputed data sets. For these aggregated regimes, we make the same comparisons as in the previous sections. Additionally, we compare different aggregation methods:

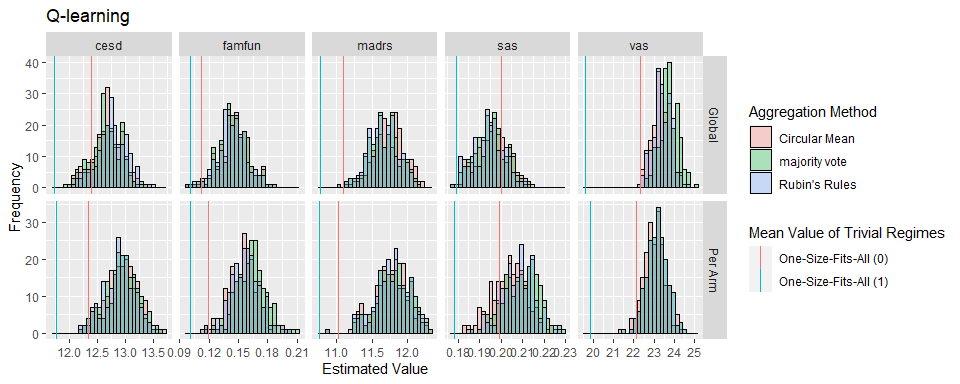
* **Circular Mean**: The parameters of the linear treatment regime are first converted to a unit vector. Next, the circular mean of the estimated unit vectors across the imputed data sets is computed.
  + This aggregation method is intuitively appealing and simple. In addition, the aggregated regime is of the same form as the estimated regimes in each imputed data set.
  + This aggregation method can be applied to the linear regimes estimated by Q-learning and value search estimation.
  + There is no guarantee that the aggregated regime is appropriate. Indeed, the circular mean is a measure for central tendency. Such measures may be less useful in the presence of certain types of uncertainty.
* **Rubin’s Rules**: Rubin’s rules are applied to the estimated outcome regression parameters. This results in a pooled estimate of the outcome regression model. The pooled estimate of the optimal treatment regime is then derived from this pooled estimate of the outcome regression model.
  + The aggregated regime is of the same form as the estimated regimes in each imputed data set.
  + This aggregation method is only applicable for Q-learning.
  + In Q-learning, the regime is estimated indirectly through the outcome regression model. Since Rubin’s rules lead to a single estimate of the outcome regression model, this method can be considered as “the correct” method for aggregating estimated regimes through Q-learning.
* **Majority Vote**: The regimes estimated in each imputed data set are retained. A patient is classified to treatment 1 by the majority vote if it is classified to treatment 1 by more than 50% of the estimated regimes (across the imputed data sets).
  + This is a very general approach to aggregating the estimated regimes. This idea is also applied in bagging (bootstrap aggregating). So, some of the advantages and disadvantages of bagging also apply here.
  + The main advantage of this aggregation approach is that it is very general. Even when the uncertainty is such that central tendency measures are not useful, the majority vote is appropriate.
  + The main disadvantage of this method is that the aggregated regime is not of the same form as the original regimes. Hence, we lose interpretability. Indeed, we go from an easy-to-interpret linear treatment regime to basically a black box regime.

## 4.2 Estimated Value of Aggregated Regimes on Imputed Data Sets

In this subsection we consider the performance of the aggregated regimes in the imputed data sets. This means that we first compute the aggregated regimes through the methods outlined above. Next, we estimate the value of the aggregated regimes in each imputed data set.

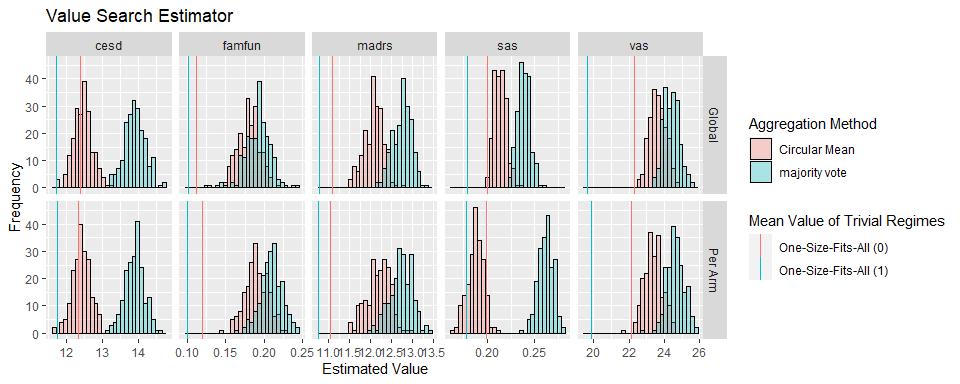
Note that this is different from what we did before in this document. Before, we looked at the estimated value of the regime estimated in the same imputed data set.

In the following histograms, we summarize the estimated values of the aggregated regimes estimated by Q-learning and aggregated by the three methods mentioned above. For Q-learning, there is no difference between the different aggregation methods. For some outcomes, there is a difference between the imputation models.



Frequency distribution of the estimated values (across imputations) for the aggregated regimes that were estimated by Q-learning. Three aggregation methods are considered. The value is estimated by the AIPW estimator which is explained at the end of this document.

In the next histograms, we compare the circular mean and the majority vote as aggregation methods for the value search estimator. There is a very substantial difference between both aggregation methods. Indeed, the majority vote consistently leads to higher estimated values. For SAS as outcome, there is a substantial difference between the imputation models.



Frequency distribution of the estimated values (across imputations) for the aggregated regimes estimated by the value search estimator. Two aggregation methods are considered. The value is estimated by the AIPW estimator which is explained at the end of this document.

## 4.3 Distance of Aggregated Regimes from One-Size-Fits-All

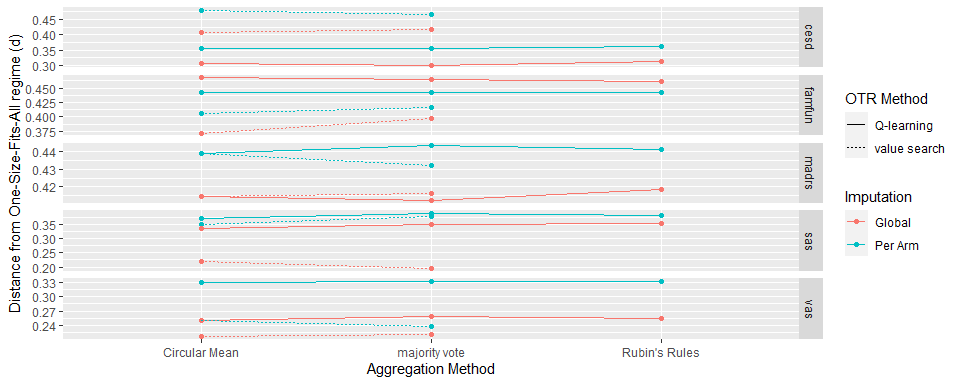
We can compute the distance between the aggregated regimes and a one-size-fits-all regime with as before. Note that there are four patients with a missing value in the baseline FAMFUN score. These patients are excluded in the calculation of . After this exclusion, the value for for a given regime is constant across imputed data sets.

For completeness, we give the values for all aggregated regimes in the following table.

Distance from one-size-fits-all for all aggregated regimes. Regimes estimated by the value search estimator cannot be aggregated with Rubin’s rules; an NA is given in the corresponding cells.

| outcome | imputation | OTR\_method | Circular Mean | Rubin’s Rules | majority vote |
| --- | --- | --- | --- | --- | --- |
| cesd | Per Arm | Q-learning | 0.356 | 0.363 | 0.354 |
| cesd | Per Arm | value search | 0.481 | NA | 0.465 |
| cesd | Global | Q-learning | 0.307 | 0.312 | 0.301 |
| cesd | Global | value search | 0.408 | NA | 0.419 |
| famfun | Per Arm | Q-learning | 0.443 | 0.443 | 0.443 |
| famfun | Per Arm | value search | 0.405 | NA | 0.416 |
| famfun | Global | Q-learning | 0.468 | 0.461 | 0.465 |
| famfun | Global | value search | 0.372 | NA | 0.396 |
| madrs | Per Arm | Q-learning | 0.439 | 0.441 | 0.443 |
| madrs | Per Arm | value search | 0.439 | NA | 0.432 |
| madrs | Global | Q-learning | 0.414 | 0.419 | 0.412 |
| madrs | Global | value search | 0.414 | NA | 0.416 |
| sas | Per Arm | Q-learning | 0.372 | 0.383 | 0.388 |
| sas | Per Arm | value search | 0.350 | NA | 0.376 |
| sas | Global | Q-learning | 0.336 | 0.352 | 0.350 |
| sas | Global | value search | 0.220 | NA | 0.194 |
| vas | Per Arm | Q-learning | 0.330 | 0.332 | 0.332 |
| vas | Per Arm | value search | 0.249 | NA | 0.238 |
| vas | Global | Q-learning | 0.249 | 0.254 | 0.258 |
| vas | Global | value search | 0.216 | NA | 0.220 |

The information in the above table is more concisely summarized in the following plot. is practically constant between the different aggregation methods for Q-learning (solid lines). In contrast, varies considerably in some settings between the aggregation methods for the value search estimator. This is consistent with the results of the previous subsection.

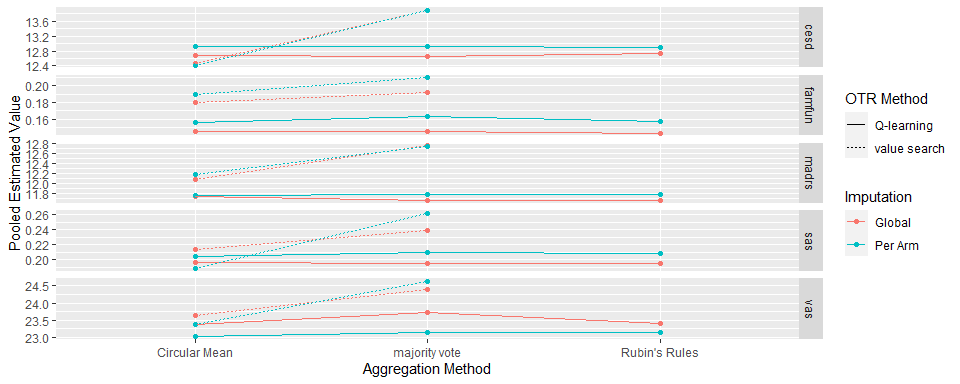


Distance from one-size-fits-all for all aggregated regimes. This plot visualizes the information from the previous table.

## 4.4 Inference on Estimated Value of Aggregated Regime

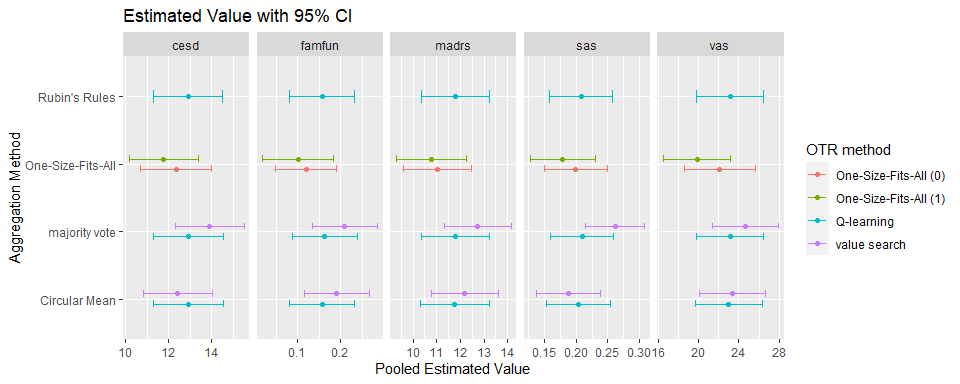
Finally, we pool the estimated values of the aggregated regimes using Rubin’s rules. The within imputation standard error for the AIPW estimator is based on the sandwich technique. Note that inferential procedures for the AIPW estimator are valid for a *fixed* treatment regime. However, we use these inferential procedures for inference for a treatment regime that has been estimated with the same data.

In the next plot, we summarize the pooled estimates for the value of the aggregated regimes. As before, there is little difference between the aggregation methods for Q-learning and a large difference for value search estimation. Indeed, for value search estimation, the majority vote performs considerably better than the circular mean. In terms of the regime’s value, the circular mean is not the optimal aggregation method in this setting. However, the circular mean is better than the majority vote in terms of simplicity and interpretability of the aggregated regime.



Pooled estimates for the value of the aggregated regimes. The pooled estimateds are obtained by applying Rubin’s rules. Estimates from the same set of imputed data sets are conected by lines to better visualize the influence of the aggregation methods.

The next plot presents the information for imputation per arm from the previous plot in a different manner. In addition, the error bars represent the 95% CI, i.e., +/- 1.96 times the pooled standard error.



Pooled estimates for the value of the aggregated regimes together with 95% confidence intervals. The pooled estimates and confidence interval are obtained by applying Rubin’s rules. Note that we only present the results under imputation per arm in this figure.

## 4.5 Interpretation of the Aggregated Regimes

In this subsection, the aggregated estimated regime parameters are interpreted. For Q-learning, we consider Rubin’s rules for aggregation. For value search estimation, we consider the circular mean for aggregation. We also restrict our attention to imputation per arm.

Note that all linear treatment regimes that we are considering are of the following form,

where is the indicator function. Note that treatment 1 is “Sertraline + IPT” and treatment 0 is “Sertraline only”. We further present 3 alternative representations of the same regimes:

1. Aggregated regime parameters as they are. These parameter estimated are normed.
2. Aggregated regime parameters for the covariates divided by the standard deviation. This ensures that all covariates have unit variance and are “on the same scale”. These parameter estimates is not normed.
3. Aggregated regime parameters after centering the continuous covariates. After centering, the parameter estimated are again normed to facilitate comparisons.

Linear regime parameter estimates for the aggregate regimes (without additional modificiations). The respective parameter vectors have unit norm.

| outcome | OTR\_method | (Intercept) | sexFemale | age | famfun | cesd | past\_MDDYes |
| --- | --- | --- | --- | --- | --- | --- | --- |
| cesd | value search | -0.991 | -0.088 | 0.004 | 0.079 | 0.022 | 0.061 |
| cesd | Q-learning | -0.990 | -0.032 | 0.012 | 0.136 | 0.005 | 0.023 |
| famfun | value search | -0.903 | 0.167 | 0.019 | 0.132 | -0.015 | 0.373 |
| famfun | Q-learning | -0.961 | 0.153 | 0.015 | 0.106 | -0.004 | 0.204 |
| madrs | value search | -0.929 | -0.099 | 0.014 | 0.209 | -0.006 | 0.287 |
| madrs | Q-learning | -0.947 | -0.074 | 0.012 | 0.241 | -0.007 | 0.201 |
| sas | value search | 0.284 | -0.286 | -0.003 | 0.225 | -0.044 | 0.886 |
| sas | Q-learning | -0.400 | -0.291 | 0.014 | -0.033 | -0.020 | 0.868 |
| vas | value search | -0.141 | -0.549 | 0.029 | -0.657 | 0.002 | 0.496 |
| vas | Q-learning | -0.167 | -0.351 | 0.021 | -0.615 | 0.010 | 0.685 |

To further facilitate the interpretation of the estimated regime parameters, we multiply the estimated regime parameters by the corresponding variable’s standard deviation. This allows us to better compare the importance of each covariate between estimated regimes. Of course, the resulting parameter vector will no longer have unit norm.

“Standardized” linear regime parameter estimates. Note that the respective parameter vectors no longer have unit norm because of this centering.

| outcome | OTR\_method | sexFemale | age | famfun | cesd | past\_MDDYes |
| --- | --- | --- | --- | --- | --- | --- |
| cesd | value search | -0.042 | 0.047 | 0.048 | 0.269 | 0.031 |
| cesd | Q-learning | -0.015 | 0.136 | 0.083 | 0.062 | 0.011 |
| famfun | value search | 0.079 | 0.224 | 0.080 | -0.177 | 0.186 |
| famfun | Q-learning | 0.072 | 0.174 | 0.065 | -0.054 | 0.102 |
| madrs | value search | -0.047 | 0.169 | 0.128 | -0.077 | 0.144 |
| madrs | Q-learning | -0.035 | 0.145 | 0.147 | -0.079 | 0.101 |
| sas | value search | -0.135 | -0.031 | 0.137 | -0.525 | 0.443 |
| sas | Q-learning | -0.138 | 0.169 | -0.020 | -0.240 | 0.434 |
| vas | value search | -0.259 | 0.340 | -0.401 | 0.027 | 0.248 |
| vas | Q-learning | -0.166 | 0.245 | -0.376 | 0.121 | 0.343 |

Finally, we also compute the coefficients after centering age, FAMFUN, and CESD. The “centered” intercept indicates to which treatment a male without past MDD and average age, FAMFUN and CESD is classified. The latter person is further termed the “reference person”. We first rewrite the linear treatment regime as follows,

where a tilde indicates the deviation from the mean, and the bar indicates the mean. Next, the factors in the above linear equation are reorganized as follows,

The centered intercept now is . If this centered intercept is positive, then a male without past MDD and average age, FAMFUN and CESD is classified to “Sertraline + IPT”. To facilitate comparisons between the centered regimes, we convert the centered regime parameters to unit norm in the next table.

Linear regime parameter estimates where the continuous covariates have been centered. Note that the respective parameter vectors have been normalized after centering. This allows for a better comparison of the coefficients.

| outcome | OTR\_method | centered\_intercept | sexFemale | age | famfun | cesd | past\_MDDYes |
| --- | --- | --- | --- | --- | --- | --- | --- |
| cesd | value search | -0.037 | -0.652 | 0.030 | 0.583 | 0.167 | 0.453 |
| cesd | Q-learning | -0.378 | -0.211 | 0.076 | 0.885 | 0.034 | 0.149 |
| famfun | value search | -0.494 | 0.338 | 0.039 | 0.266 | -0.030 | 0.754 |
| famfun | Q-learning | -0.658 | 0.416 | 0.040 | 0.289 | -0.012 | 0.556 |
| madrs | value search | -0.135 | -0.265 | 0.039 | 0.562 | -0.017 | 0.771 |
| madrs | Q-learning | -0.252 | -0.221 | 0.037 | 0.722 | -0.020 | 0.603 |
| sas | value search | -0.509 | -0.257 | -0.002 | 0.202 | -0.039 | 0.795 |
| sas | Q-learning | -0.445 | -0.285 | 0.014 | -0.032 | -0.020 | 0.848 |
| vas | value search | -0.347 | -0.520 | 0.028 | -0.622 | 0.002 | 0.470 |
| vas | Q-learning | -0.382 | -0.329 | 0.020 | -0.577 | 0.009 | 0.642 |

We summarize our findings based on the above results.

* The general magnitude and direction of the effects agree moderately between value search estimation and Q-learning.
* Although the direction of the estimated coefficients generally (but not always) agrees across outcomes, there are considerable differences in magnitude. Hence, the importance of a covariate in an estimated regime depends on the outcome variable.
  + We should interpret the estimated coefficients with care and think about standardization and/or centering. For example, the estimated coefficients are small for past MDD with CESD as outcome, even after standardization. The corresponding coefficients for the other outcomes are much larger. However, after centering, the coefficients are large for past MDD with CESD as outcome.
* Past MDD is the only covariate where the coefficient is consistently positive. Hence, for all estimated regimes displayed above, patients with past MDD tend to be classified more to “Sertraline + IPT”.
  + The coefficients for the other covariates are not consistently positive or negative. However, we show next that the allocations of most of these regimes still agree moderately.

We can also approach the differences between the estimated regimes from a different perspective. Next, we look at the misclassification probabilities between the classifications of the 10 estimated regimes we already examined above. These probabilities summarize in a single quantity how similar or different the estimated regimes are. These are given in the next table. We gain the following insights:

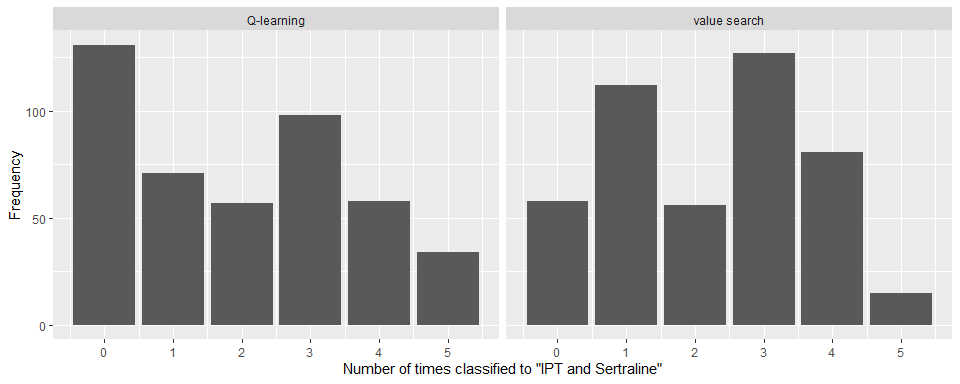
* Regimes for the same outcome, but estimated by a different method, result in a similar classification. This can be seen from the misclassification probabilities smaller than 0.5. Only the misclassification probability between value search estimation and Q-learning for CESD is moderately high: 28.7% of the patients are classified differently.
* Most of the misclassification probabilities are smaller than 0.5. In fact, all misclassification probabilities are smaller than 0.5 if we exclude CESD. This means that the estimated optimal regimes for different outcomes tend to classify patients to the same treatments. Even for different outcome variables, the misclassification probabilities can be quite low, e.g., between 0.149 and 0.212 for allocations of regimes estimated for MADRS and FAMFUN as outcome. Note that this similarity is not apparent from the estimated regime parameters.

Misclassification probability for each pair of aggregated regimes. Note that we only present the results under imputation per arm. The Q-learning and value search regimes are aggregated by Rubin’s rules and the circular mean, respectively.

|  | cesd\_Q-learning | cesd\_value search | famfun\_Q-learning | famfun\_value search | madrs\_Q-learning | madrs\_value search | sas\_Q-learning | sas\_value search | vas\_Q-learning | vas\_value search |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cesd\_Q-learning | 0.000 | 0.287 | 0.267 | 0.410 | 0.274 | 0.341 | 0.461 | 0.508 | 0.405 | 0.381 |
| cesd\_value search | 0.287 | 0.000 | 0.492 | 0.577 | 0.477 | 0.486 | 0.561 | 0.648 | 0.443 | 0.494 |
| famfun\_Q-learning | 0.267 | 0.492 | 0.000 | 0.178 | 0.207 | 0.194 | 0.301 | 0.361 | 0.383 | 0.403 |
| famfun\_value search | 0.410 | 0.577 | 0.178 | 0.000 | 0.212 | 0.149 | 0.283 | 0.321 | 0.428 | 0.452 |
| madrs\_Q-learning | 0.274 | 0.477 | 0.207 | 0.212 | 0.000 | 0.125 | 0.258 | 0.274 | 0.408 | 0.379 |
| madrs\_value search | 0.341 | 0.486 | 0.194 | 0.149 | 0.125 | 0.000 | 0.232 | 0.287 | 0.403 | 0.410 |
| sas\_Q-learning | 0.461 | 0.561 | 0.301 | 0.283 | 0.258 | 0.232 | 0.000 | 0.149 | 0.256 | 0.263 |
| sas\_value search | 0.508 | 0.648 | 0.361 | 0.321 | 0.274 | 0.287 | 0.149 | 0.000 | 0.374 | 0.341 |
| vas\_Q-learning | 0.405 | 0.443 | 0.383 | 0.428 | 0.408 | 0.403 | 0.256 | 0.374 | 0.000 | 0.114 |
| vas\_value search | 0.381 | 0.494 | 0.403 | 0.452 | 0.379 | 0.410 | 0.263 | 0.341 | 0.114 | 0.000 |

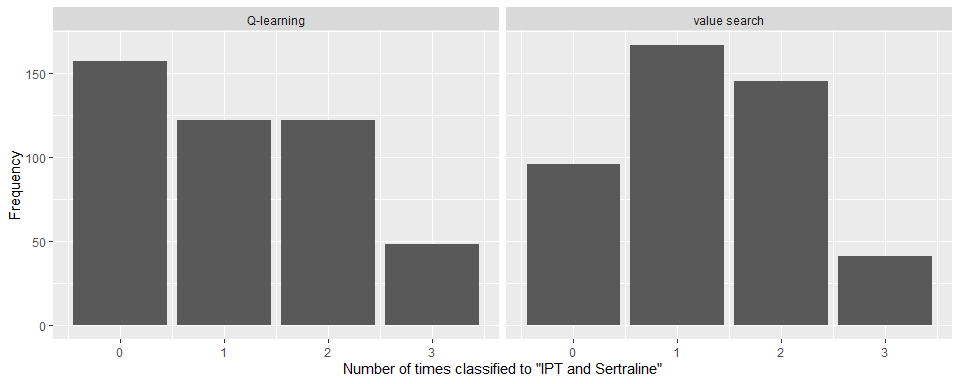
The misclassification probability only looks at pairs of estimated regimes. It is furthermore also interesting to examine how well the classifications agree when considering the estimated regimes for all 5 outcomes simultaneously. In the next plot, we summarize how many times a patient is classified to “IPT and Sertraline” for the 5 aggregated regimes (one for each outcome). We do this for Q-learning and value search estimation separately.

Ideally, the classifications for individual patients agree for all estimated regimes, no matter for which outcome the regime was estimated. Indeed, the optimal treatment for depression should be relatively insensitive to the scale that is used to quantify depressive symptoms. Clearly, the figure below shows that this is not the case here. Especially for value search estimation, few patients are classified to the same treatment for all 5 outcomes.



Frequency distribution of the number of times each patient is classified to “IPT and Sertraline” when considering the five aggregated regimes, one for each outcome variable. Note that we only present the results under imputation per arm. The Q-learning and value search regimes are aggregated by Rubin’s rules and the circular mean, respectively.

The above paragraph may be unfair because only 3 out of the 5 outcome variables are measuring mood. We therefore repeat the same figure for CESD, MADRS, and VAS only.



Frequency distribution of the number of times each patient is classified to “IPT and Sertraline” when considering only the aggregated regimes for outcomes that measure mood (CESD, MADRS, and VAS). Note that we only present the results under imputation per arm. The Q-learning and value search regimes are aggregated by Rubin’s rules and the circular mean, respectively.

# 5 Details on the Value Search Estimator

The value search estimator estimates the optimal treatment regime within a restricted class of treatment regimes . This is done by maximizing an estimator for the value of a fixed regime, . The value search estimator is thus defined as,

In this document, the AIPW estimator used as . This estimator for the value of a fixed treatment regime requires the specification of a propensity score model and an outcome regression model:

* **Propensity score model**. Since treatment assignment was randomized, we estimate the propensity scores by the empirical proportions. This corresponds to a logistic regression model with only an intercept.
* **Outcome regression model**. A linear regression model with sex, past\_MDD, current\_MDD, phealth, age, madrs, sas, famfun, cesd, vas as main effects and interactions terms between treatment and sex, age, famfun, cesd, and past\_MDD.

The restricted class of treatment regimes considered in the value search estimator are all linear treatment regimes with the following covariates: sex, age, famfun, cesd, and past\_MDD.

The actual maximization is a very difficult problem because the objective function is not smooth. The maximization is carried out by a genetic algorithm. The genetic algorithm is run independently 20 times with a population size of 500 and 5 times with a population size of 3000. This approach is justified in a separate document. Multiple runs of this algorithm are required because the algorithm often ends up in local optima. By running this algorithm multiple times with different population sizes, the chances of finding the global optimum are increased. Although, there is no guarantee that the global optimum will found by this approach.