MNAR Sensitivity Analyses

Florian Stijven

2024-09-02

Table of Contents

# 1 Introduction

In this report, we assess the optimal estimated treatment regimes under a set of MNAR scenarios with a *global shift*: imputed values are shifted downwards or upwards by a constant amount. We focus on the CESD outcome at 6 months.

Because the CESD scale is confined to the interval, we have to take these bounds into account when implementing global shift scenarios. In the main analyses - and for the MNAR scenarios as well - the CESD outcome is transformed to the real line before imputing, using the following transformation:

where is the transformed value.

Afterwards, the imputed values are back-transformed to their original scale. This ensures that (practically) all imputed values are in the interval. In practice, the imputed values are in the interval. Imputed values in and are converted to and , respectively.

To ensure that imputations in the MNAR scenarios still respect the interval, the global shift is implemented on the scale. Consequently, it is difficult to appreciate the size of the global shift. Therefore, for each shift on the scale, we also compute the average difference between the imputed values under the MAR and the MNAR scenario. Specifically, this is done by computing

where the expectations are with respect to the patients with a missing CESD outcome at 6 months, and and are the imputed values under MNAR and MAR, respectively.

In this report, we consider MNAR scenarios from an extremely large negative to an extremely large positive global shift. However, it should be kept in mind that higher scores for CESD correspond to worse depressive symptoms. If we believe that patients for which the treatment does not work are more likely to drop out of the study, then negative shifts in the CESD are unrealistic.

In this report, all regimes are estimated through Q-learning. Further, all analysis choices are the same as for the main analysis under MAR unless mentioned otherwise. We further only present results with the circular mean as aggregation method. We also consider both the original Browne data and the Browne data where we artificially induced strong interaction effects; these latter data are further referred to as the “updated” Browne data.

# 2 Estimated Values

In this section, we look at the estimated values of the aggregated regimes across the MNAR scenarios. We follow the same approach as in the main analysis to obtain the aggregated regimes and the corresponding estimated values. Figure 2.1 presents these estimated values along with 95% CIs (+/- 1.96 times the pooled standard error). In addition to the regimes estimated by Q-learning, this plot also contains the estimated values of the trivial regimes.

* For both the original and the updated Browne data, the difference between the one-size-fits-all regimes increases with the global shift:
  + A large positive shift increases the difference.
  + A large negative shift decreases the difference.
* The difference between the aggregated regime and the best one-size-fits-all regime is relatively insensitive to a global shift.
  + For the original Browne data, the estimated difference in values of the aggregated and the best one-size-fits-all regime depends on the global shift only for relatively large positive shifts. In addition, the confidence interval for the value of the aggregated regime always contains the estimated value of the best one-size-fits-all regime.
  + For the updated Browne data, the estimated difference in values of the aggregated and the best one-size-fits-all regime is practically independent of the global shift. However, the length of the confidence intervals depends on the global shift. Indeed, the lower limit of the confidence interval for the aggregated regime approaches the estimated value of the best one-size-fits-all regime as the global shift increases.

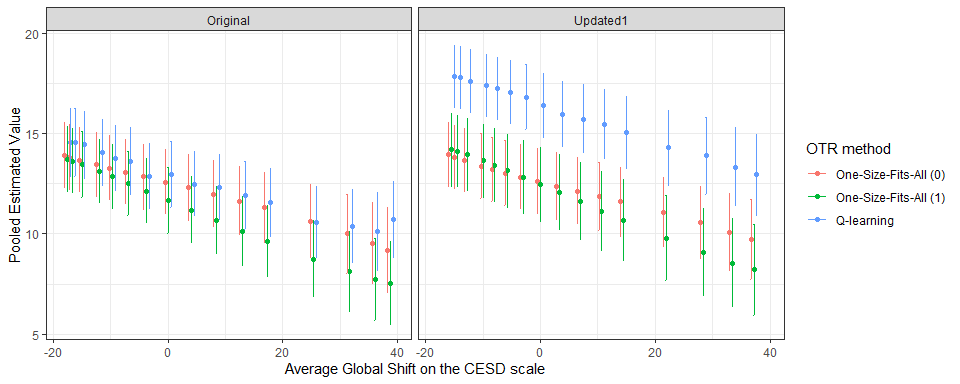


Figure 2.1: Pooled estimates for the value of the aggregated regimes together with 95% confidence intervals under various MNAR scenarios. The pooled estimates and confidence interval are obtained by applying Rubin’s rules. The left-hand panel shows the results with the original Browne data and the right-hand panel shows the results with the updated Browne data.

# 3 Classification

In this section, we look at how the aggregated regimes classify patients across the MNAR scenarios. First, we describe this classification. Second, why try to explain differences in these classifications.

## 3.1 Descriptive

In Figure 3.1, we plot the proportion of patients classified to *Sertraline* by the aggregated regimes under various MNAR scenarios. This proportion is clearly much more sensitive to MNAR for the original Browne data than for the updated Browne data.

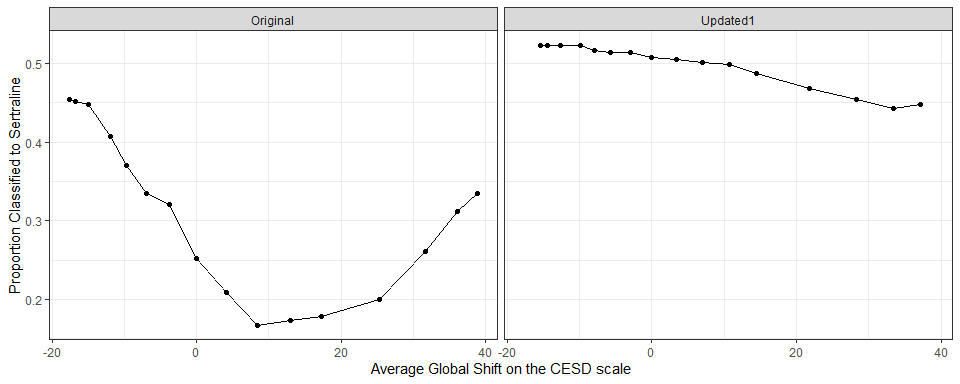


Figure 3.1: The proportion of patients that are classified to Sertraline by the aggregated regime, under various MNAR scenarios. The left-hand panel shows the results with the original Browne data and the right-hand panel shows the results with the updated Browne data.

The marginal classification probabilities shown above do not directly reveal how much the regimes differ in their classifications across MNAR scenarios. Therefore, we also consider the 2x2 tables for comparing two regimes: (i) the regime classifications under MAR and (ii) the regime classifications under a particular MNAR scenario. This is illustrated in the following table.

|  | *Sertraline (MAR)* | *Sertraline + IPT (MAR)* |
| --- | --- | --- |
| ***Sertraline (MNAR)*** |  |  |
| ***Sertraline + IPT (MNAR)*** |  |  |

These cell probabilities are plotted in Figure 3.2 across the MNAR scenarios. Looking at the discordant cell probabilities ( and ) may give us some insights into the differences between the regimes under various MNAR scenarios. Indeed, we see that the classifications are much less sensitive to the MNAR scenarios in the updated Browne data than in the original Browne data. However, it is difficult to judge the relative importance of these differences. We therefore also plot the corresponding Cohen’s kappas in Figure 3.3. This corroborates our conclusions from Figure 3.2:

* In the original Browne data, the patient classifications are very sensitive to the global shift.
* In the updated Browne data, the patient classifications are relatively insensitive to the global shift. Cohen’s kappa remains large, even for extremely large positive shifts.

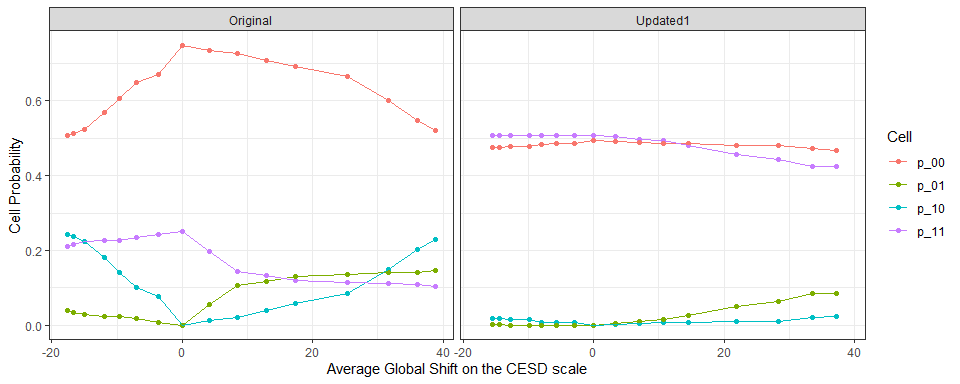


Figure 3.2: Cell probabilities in the 2x2 tables mentioned in the text under various MNAR scenarios. The left-hand panel shows the results with the original Browne data and the right-hand panel shows the results with the updated Browne data.

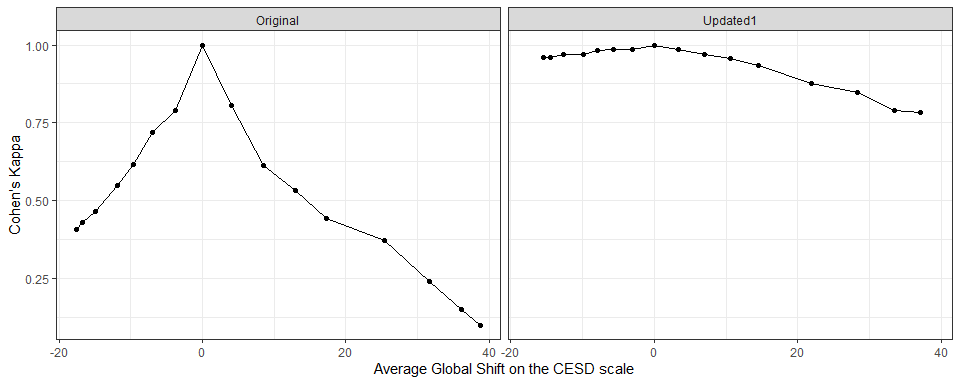


Figure 3.3: Cohen’s kappa values in the 2x2 tables mentioned in the text under various MNAR scenarios. The left-hand panel shows the results with the original Browne data and the right-hand panel shows the results with the updated Browne data.

## 3.2 Explanatory

In the above subsection, we have seen that the classification of patients by the aggregated regimes changes as the global shift changes In this subsection, we try to explain this differential classification using the patient-level treatment contrasts, that is,

In practice, we can only estimate these contrasts. We consider two estimators.

1. Outcome regression-based. This estimator is derived from the outcome regression estimator for the value of a fixed regime. Let be the estimated regression function (that was also used for Q-learning). The contrast is then estimated as
2. AIPWE-based. This estimator is derived from the AIPW estimator for the value of a fixed regime. Let be the estimated propensity score, the treatment group ( for *Sertraline* and for *Sertraline + IPT*), and the the CESD score at 6 months for patient . The contrast is then estimated as

For each patient, we compute the estimated contrast under MAR by first estimating the contrast in each imputed data set. Second, these estimated patient-level contrasts are pooled across imputations with Rubin’s rules. Next, for each MNAR scenario we define two groups according to whether the patient’s classification changes as compared to MAR (i.e., differential classification or not).

Figure 3.4 summarizes the distribution of the absolute values of the estimated contrasts in each differential classification group, stratified by the MNAR scenarios. Note that this plot only includes the results for the original Browne data and the AIPWE-based estimator of the contrast. The corresponding plots for the outcome regression-based estimator are presented in Figure 3.5.

The outcome regression-based contrast estimates clearly explain which patients are differentially classified: patients with small estimated contrasts. These are patients for which there is little difference between both treatment options in terms of expected benefit. However, this is not clear when looking at the AIPWE-based contrast estimates. One possible reason for this discrepancy is that AIPWE-based estimates are much more noisy. Indeed, the range of the x-axis is much wider in Figure 3.4 than in Figure 3.5. A second possible reason is that the regimes were estimated by Q-learning which is directly related to the outcome regression based-estimator of the contrast estimates.

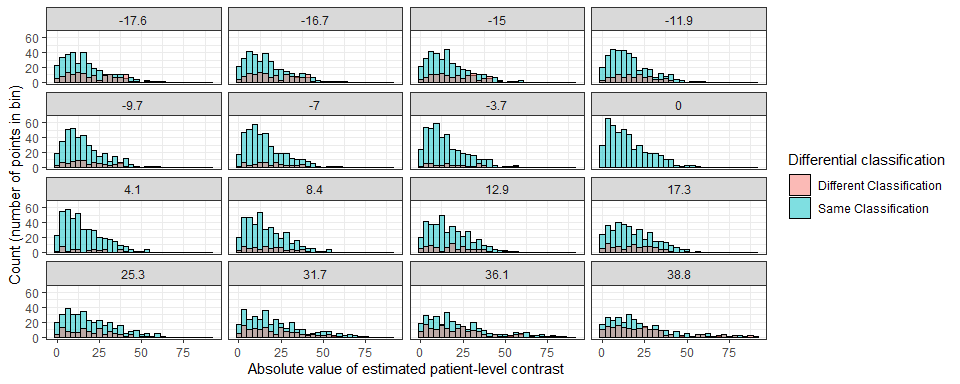


Figure 3.4: Distribution of the AIPWE-based estimates of the patient-level contrasts in the original Browne data. Each subplot shows the histogram for these estimated contrasts, colored by whether the patient was differentially classified as compared to the MAR scenario. The subplot titles indicate the average shift in the corrsponding MNAR scenario.

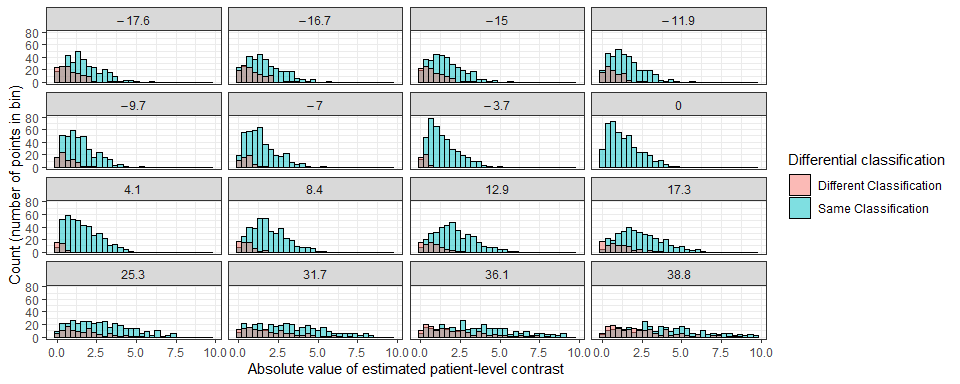


Figure 3.5: Distribution of the outcome regression-based estimates of the patient-level contrasts in the original Browne data. Each subplot shows the histogram for these estimated contrasts, colored by whether the patient was differentially classified as compared to the MAR scenario. The subplot titles indicate the average shift in the corrsponding MNAR scenario.

# 4 Conclusions

In the original Browne data, the aggregated regime is sensitive to MNAR scenarios with a global shift in terms of the classification of patients. However, in all these MNAR scenarios, the estimated values of the aggregated regimes remain close to the corresponding estimated values of the best one-size-fits-all regime. The conclusions of the analysis under MAR thus remain valid under the global shift MNAR scenarios: there is no evidence of a useful non-trivial treatment regime.

In the updated Browne data, the aggregated regime is relatively insensitive to the MNAR scenarios with a global shift. Both the classification of patients and the aggregated regime’s value (relative to the best one-size-fits-all regime) are not severely changed under these MNAR scenarios. The conclusions of the analysis under MAR thus remain valid under the global shift MNAR scenarios: there is evidence of a useful non-trivial treatment regime and most patients are classified to the same treatment under MNAR.