OPEN ACCESS

University of BRISTOL

Hazewinkel, D., Bowden, J., Wade, K. H., Palmer, T. M., Wiles, N. J., & Tilling, K. M. (2021). Sensitivity to missing not at random dropout in clinical trials: use and interpretation of the Trimmed Means Estimator. Unpublished. https://doi.org/10.1101/2021.03.05.21252334

University of Bristol - Explore Bristol Research
General rights

Sensitivity to missing not at random dropout in clinical trials: use and interpretation of the Trimmed Means Estimator

Audinga-Dea Hazewinkel[1,2], Jack Bowden[2,4], Kaitlin H. Wade[1,2],
Tom Palmer[1,2], Nicola Wiles[3], and Kate Tilling[1,2]

[1]*Population Health Sciences, Bristol Medical School, University of Bristol, Bristol,
U.K.*
[2]*Medical Research Council Integrative Epidemiology Unit, Bristol Medical School,
University of Bristol, Bristol, U.K.*
[3]*Centre for Academic Mental Health, Population Health Sciences, Bristol Medical
School, University of Bristol, Bristol, U.K.*
[4]*Exeter Diabetes Group (ExCEED), College of Medicine and Health, University of
Exeter, Exeter,U.K.*

Abstract

Outcome values in randomized controlled trials (RCTs) may be missing not at random (MNAR), if patients with extreme outcome values are more likely to drop out (e.g., due to perceived ineffectiveness of treatment, or adverse effects). In such scenarios, estimates from complete case analysis (CCA) and multiple imputation (MI) will be biased. The trimmed means (TM) estimator operates by setting missing values to the most extreme value, and then "trimming" away equal fractions of both treatment groups, estimating the treatment effect using the remaining data. The TM estimator relies on two assumptions, which we term the "strong MNAR" and "location shift" assumptions. In this article, we derive formulae for the bias resulting from the violation of these assumptions for normally distributed outcomes. We propose an adjusted estimator, which relaxes the location shift assumption and detail how our bias formulae can be used to establish the direction of bias of CCA, MI and TM estimates under a range of plausible data scenarios, to inform sensitivity analyses. The TM approach is illustrated with simulations and in a sensitivity analysis of the CoBalT RCT of cognitive behavioural therapy (CBT) in 469 individuals with 46 months follow-up. Results were consistent with a beneficial CBT treatment effect. The MI estimates are closer to the null than the CCA estimate, whereas the TM estimate was further from the null. We propose using the TM estimator as a

sensitivity analysis for data where it is suspected that extreme outcome values are missing.

1 Introduction

Randomized controlled trials (RCT) are considered the gold standard for assessing causality, because the randomization process ensures that unmeasured characteristics are well-balanced across groups. However, RCTs remain vulnerable to other sources of bias, including those which arise from missing data due to dropout. We focus here on the case where an RCT has missing values solely in the outcome, for example due to drop out.

The impact of missing data depends on the missingness mechanism and the analysis model.[1] Three distinct missingness mechanisms can be distinguished: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). With MCAR, missingness is unrelated to any measured or unmeasured characteristics and the observed sample is a representative subset of the unobserved full data.[3] MAR means the missingness can be explained by observed data and, with MNAR, missingness is a function of the unobserved data (in our case, the outcome) itself.[2]

There are three main ways of dealing with incomplete data: a complete case analysis (CCA), inverse probability weighting (IPW) and multiple imputation (MI). A CCA is the analysis model intended to be applied to the trial data at its outset, restricted only to individuals with observed outcomes. Assuming correct model specification, CCA is unbiased if the outcome is MCAR or MAR[3]. With IPW, the same analysis model is fitted to the complete cases, but each individual is now weighted by the inverse of its probability of being observed. With MI, the observed data are used to repeatedly predict - "impute" - the missing values. The analysis model is applied to multiple imputed datasets and the estimates pooled using Rubin's rules.[1]. Broadly, IPW and MI will be valid if the data are MAR and the weighting and imputation models are correctly specified, respectively. When data are MNAR, however, the treatment effect estimate will be biased in general for any of the aforementioned approaches.[1,3] In practice, observed data cannot be used to determine whether data are MAR or MNAR, and, consequently, if estimates from CCA, MI or IPW analyses are

likely to be biased.

Permutt & Li[4] suggested a trimmed means (TM) estimator for RCTs where patients who drop out have more extreme (unobserved) outcome values than those who do not drop out (e.g., lower value dropout, when comparator group patients, not experiencing a benefit of treatment, leave the study early). Observations are ordered within each treatment group, with the missing observations assigned the lowest rank. Equal proportions of data are trimmed away from the lower end of both treatment group distributions and the TM treatment effect estimate is obtained from a linear regression of the exposure and outcome variables, using the remaining "trimmed" data.

The TM estimator will give an unbiased estimate of the true treatment effect, given two main assumptions - the location shift assumption and the strong MNAR assumption. The first specifies identical distributions of the treatment groups, with the only difference being a mean shift. The second restricts all dropout to the fraction that is trimmed away.[5] Previous studies have performed a variety of simulations investigating the type 1 error and power of the TM estimator in a range of clinical scenarios[6] and for various MNAR/MAR dropout patterns[5]. However, the biases that arise from violations of the strict TM assumptions and how to correct for these biases have yet to be established.

In this article, we derive formulae for the bias resulting from the violation of the location shift and strong MNAR assumptions for a normally distributed outcome, and illustrate, by means of simulations, the TM estimator bias for range of MNAR/MAR mechanisms. Additionally, we propose an adjustment to the estimator which relaxes the location shift assumption and leaves the estimator reliant only on the strong MNAR assumption and normality. For the purpose of this article, consistent with previous publications,[4,5,6] we primarily consider worst value dropout. The principle, however, is equally applicable to higher value dropout, which may occur when patients leave the study perceiving themselves to be recovered, or where higher values indicate a worse response. We also primarily consider the case of 50% fixed trimming, but provide more general formulae for alternate trimming fractions. The TM approach is illustrated in an application to the CoBalT randomized controlled trial (registration ISRCTN38231611), which compares the effectiveness of cognitive behavioural therapy (CBT) as an adjunct to pharmacotherapy versus usual care in patients with treatment resistant depression. We have developed an R package 'tmsens' for performing a TM regression and conducting a TM estimator sensitivity analysis, available from `https://github.com/dea-hazewinkel/tmsens`.

2 Methods

2.1 Notation

Consider a clinical trial with n subjects, with a continuous outcome, Y, with patients randomized to receive an active treatment ($R = 1$) or comparator ($R = 0$). Let y_{ij} denote the observed outcome for patient i randomized to treatment group j, where $i = 1, ..., n_j$ and $j = 0, 1$. Let μ_j be the population mean for arm j, with $\mu_j = \mathbb{E}[Y|R = j]$, and its estimate, $\widehat{\mu}_j$, obtained from the corresponding sample mean:

$$\widehat{\mu}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}.$$

.
When there are no missing outcomes in the trial, each $\widehat{\mu}_j$ is an unbiased estimate of μ_j. Let β denote the true treatment effect, given by the difference in treatment group means:

$$\beta = \mu_1 - \mu_0,$$

where β is estimated by $\widehat{\beta}$:

$$\widehat{\beta} = \widehat{\mu}_1 - \widehat{\mu}_0.$$

When the trial contains missing outcomes, we define a missing indicator M, which equals 0 if y_{ij} is observed and 1 if y_{ij} is not observed. Let $\widehat{\mu}_{cj}$ be the mean of all observed values for group j, given by $\widehat{\mu}_{cj} = \frac{1}{n_{mj}} \sum_{i=1}^{n_j} y_{ij}(1 - m_{ij})$, with $n_{mj} = \sum_{i=1}^{n_j}(1 - m_{ij})$ and m_{ij} being subject i's missingness indicator. Then $\widehat{\mu}_{cj}$ estimates the population complete case mean $\mu_{cj} = \mathbb{E}[Y|R = j, M = 0]$ and the complete case estimand of the treatment effect is given by

$$\beta_c = \mu_{c1} - \mu_{c0},$$

with β_c estimated by

$$\widehat{\beta}_c = \widehat{\mu}_{c1} - \widehat{\mu}_{c0} = \frac{1}{n_{m1}} \sum_{i=1}^{n_1} y_{i1}(1 - m_{i1}) - \frac{1}{n_{m0}} \sum_{i=1}^{n_0} y_{i0}(1 - m_{i0}).$$

2.2 Trimmed means estimator

We first define the population trimmed mean, μ_{tj}, in the absence of dropout ($m_{ij} = 0 \; \forall \; i$). For treatment group j, μ_{tj} is given by the expected value of all observations exceeding the quantile $F^{-1}(p)$:

$$\mu_{tj} = \mathbb{E}[Y|R = j, Y > F^{-1}(p)], \tag{1}$$

where p is the proportion of outcomes trimmed away from the lower end of the distribution of each group j (e.g., for $p = 0.25$, the bottom 25% of the distribution would be removed). Then μ_{tj} is estimated by

$$\widehat{\mu}_{tj} = \frac{1}{n_{tj}} \sum_{\substack{y_{ij} > \widehat{F}^{-1}(p)}}^{n_j} y_{ij}, \tag{2}$$

with n_{tj} being the sample size after trimming ($n_{tj} = \lceil n_j(1-p) \rceil$, with $\lceil \ \rceil$ the ceiling function).

Let β_t denote the TM effect estimand, given by the difference in population trimmed means ($\beta_t = \mu_{t1} - \mu_{t0}$), estimated by $\widehat{\beta}_t = \widehat{\mu}_{t1} - \widehat{\mu}_{t0}$. If the outcomes within each treatment group are normally distributed with underlying mean μ_j and variance σ_j^2, then $\mu_{tj} = \mathbb{E}[Y|R = j, Y \geq \sigma_j \Phi^{-1}(p) + \mu_j]$. If the treatment group standard deviations (SDs) are equal ($\sigma_1 = \sigma_0$), $C = \sigma_j \Phi^{-1}(p)$ is common across treatment groups and the population TM effect, β_t, is identical to the population mean difference β, since

$$\begin{aligned}
\beta_t &= \mathbb{E}[Y|R = 1, Y \geq C + \mu_1] - \mathbb{E}[Y|R = 0, Y \geq C + \mu_0] \\
&= \mathbb{E}[Y|R = 1] - \mathbb{E}[Y|R = 0] = \beta.
\end{aligned} \tag{3}$$

This property will also hold in absence of normality, if the outcome distributions for each group are identical in shape and differ only by a mean shift.

We now consider the TM estimator in the presence of dropout. As the estimator assumes worst value dropout, all missing values ($m_{ij} = 1$) are assigned a value smaller than the worst observed outcome:

$$(y_{ij}|m_{ij} = 1) \leq \min(y_{ij}|m_{ij} = 0)$$

The precise value is unimportant as long as the above inequality holds. The trimmed mean is obtained by taking the average of all observations exceeding the quantile of the trimming proportion, p, as in (2), where p must now be equal to or exceed the largest observed dropout proportion across the groups:

$$p > p_{\min} = \max \left(\frac{n_{m1}}{n_1}, \frac{n_{m0}}{n_0} \right).$$

If there is no dropout in the trimmed fraction used to estimate μ_t (the "strong MNAR assumption"), $\widehat{\beta}_t$ is an unbiased estimator for the TM effect β_t. In addition, if the group variances are equal (the "location shift assumption"), β_t is unbiased for β (3). Previously, it has been stressed that the TM estimator estimates an unique estimand - the mean difference of the best X% patients in each treatment group, making it hard to compare with other approaches[5,6]. In

order to identify this TM-specific estimand, only the strong MNAR assumption is required. Under the additional constraint of the location shift assumption, the TM effect becomes an unbiased estimator of the population mean difference as shown in (3).

We now examine the two sources of bias in the TM estimator. Let B_t be the total bias for the TM estimator, representing the deviation of the TM estimate, $\widehat{\beta}_t$, from the population mean difference, β:

$$B_t = \mathbb{E}[\widehat{\beta}_t - \beta] = \mathbb{E}[\widehat{\beta}_t - \beta_t] + (\beta_t - \beta). \tag{4}$$

The first component results from the violation of the strong MNAR assumption, the second from violation of the location shift assumption. Supplementary Figure S1 (Appendix A) illustrates the TM and CCA estimator behaviours across four scenarios with varying dropout patterns and equal and unequal treatment arm SDs. Box 1 summarizes the terminology used throughout this paper.

2.3   The location shift assumption

We now derive expressions for the bias resulting from the violation of the location shift assumption, B_{tLS}, for the case of 50% trimming, with results for a more general trimming fraction, p, derived in Appendix B. We assume that the strong MNAR assumption is satisfied and the outcomes are normally distributed for each treatment arm. Then, $\widehat{\beta}_t$ is an unbiased estimator for β_t and the total bias (4) is given by

$$B_{tLS} = \mathbb{E}[\widehat{\beta}_t - \beta] = \beta_t - \beta.$$

We define the 50% trimmed left-truncated mean (Appendix B):

$$
\begin{aligned}
\mu_{tj} &= \mu_j + \sigma_j \frac{\phi\big(\Phi^{-1}(0.5)\big)}{0.5} \\
&= \mu_j + \sigma_j \sqrt{\frac{2}{\pi}},
\end{aligned}
\tag{5}
$$

Then, the 50% trimmed population mean difference is given by

$$\beta_t = \left(\mu_1 + \sigma_1 \sqrt{\frac{2}{\pi}} \right) - \left(\mu_0 + \sigma_0 \sqrt{\frac{2}{\pi}} \right),$$

with the bias, B_{tLS}, resulting from unequal SDs:

$$B_{tLS} = (\sigma_1 - \sigma_0)\sqrt{\frac{2}{\pi}}. \tag{6}$$

The TM estimator bias, B_{tLS} (6), can alternatively be expressed as a function of the trimmed fraction SDs, rather than of the unobserved full sample SDs (Appendix C):

$$B_{tLS} = (\sigma_{t1} - \sigma_{t0})\sqrt{\frac{2}{\pi - 2}}. \tag{7}$$

From both B_{tLS} bias formulae, (6) and (7), it is apparent that the estimator is unbiased given equal treatment group SDs, with the latter notation (7) making it explicit that equality of trimmed fraction SDs suffices. This property underlies the adjustment to the TM estimator outlined in Section 2.5, which relaxes the location shift assumption, allowing for different treatment group SDs.

BOX 1

List of terms used throughout this paper

Estimands	*Definition*
β	The true treatment effect, given by the difference in population treatment group means, in absence of dropout
β_c	CCA estimand for the treatment effect, subject to the assumption of a homogeneous dropout mechanism across the entirety of the distribution
β_t	TM estimand for the treatment effect, subject to the location shift assumption and the strong MNAR assumption
β_{td}	TM estimand in the presence of dropout, subject to the location shift assumption
β_{at}	Adjusted TM estimand, subject to the strong MNAR assumption, obtained under 50% trimming of normally distributed data
Assumptions	
Strong MNAR	All missing values occur in the fraction that is trimmed away
Location shift	The underlying treatment group distributions differ only by a mean shift (for normally distributed outcomes, this constitutes equal group SDs)
Homogeneous dropout	Dropout is uniformly likely across a range of the distribution, so that within this range, any observation is equally likely to drop out
Bias terms[1]	
B_C	CCA estimator bias on violating the assumption of homogeneous dropout across the entire distribution
$B_{C\max}$	Maximum CCA estimator bias, occurring under dropout restricted to the very lowest or highest values of the distribution
B_t	TM estimator bias on violating the location shift assumption and/or the strong MNAR assumption
B_{tLS}	TM estimator bias on violating the location shift assumption
B_{tSM}	TM estimator bias on violating the strong MNAR assumption, derived under the assumption of a homogeneous dropout mechanism acting on a part of or the entire distribution
$B_{tSM\max}$	Maximum TM estimator bias on violating the strong MNAR assumption, occurring for lower value trimming under highest value dropout and vice versa
$B_{t\max}$	Maximum TM estimator bias on violating the strong MNAR and location shift assumptions ($B_{t\max} = B_{tSM\max} + B_{tLS}$)

1) All bias terms are derived under the assumption of normally distributed outcomes

2.4 The strong MNAR assumption

We now derive expressions for the bias resulting from the violation of the strong MNAR assumption, B_{tSM}, for dropout in the comparator group, with results for dropout in both groups derived in Appendix E. We examine violation due to a specific homogeneous dropout mechanism, where, in a given part of the distribution, any observation is equally likely to be selected for dropout.

We assume that the location shift assumption is satisfied and the outcomes are normally distributed in each treatment arm, prior to dropout. Then, β_t is an unbiased estimator for β and the total bias (4) reduces to

$$B_{tSM} = \mathbb{E}[\widehat{\beta}_t - \beta_t]. \tag{8}$$

We define a new population parameter, β_{td}, which gives the population TM effect for a scenario with a normally distributed treatment group and a comparator group no longer normally distributed due to dropout. Then, $\widehat{\beta}_t$ unbiasedly estimates β_{td}, the TM effect, and the bias, B_{tSM} (8), can be written as

$$B_{tSM} = \beta_{td} - \beta_t, \tag{9}$$

and, with $\beta_t = \mu_{t1} - \mu_{t0}$ and $\beta_{td} = \mu_{t1} - \mu_{td0}$, simplified to

$$B_{tSM} = \mu_{t0} - \mu_{td0}. \tag{10}$$

While μ_{td0} is the trimmed mean of a population with non-normally distributed outcomes, we can express it in terms of a normal distribution, under the assumption of a homogeneous dropout mechanism. Consider a normal distribution with lower bound, $\Phi^{-1}(0)$ and upper bound, $\Phi^{-1}(1)$, and let $f_{u,v}$ denote the fraction of this distribution, so that $f_{0,1} = 1$. Let $f_{0,c}$ denote the fraction affected by homogeneous dropout, with $c \leq 1$, and $\Phi^{-1}(c)$ its upper bound, so that each value in $f_{0,c}$ is equally likely to drop out, with a given probability a:

$$P(M = 0|R = 0) = \begin{cases} a & \text{if } Y|R = 0, Y \leq \Phi^{-1}(c) \\ 0 & \text{if } Y|R = 0, Y > \Phi^{-1}(c). \end{cases}$$

Further, let $f_{0,0.5}$ denote the fraction of the distribution that is trimmed away under 50% trimming, $f_{0.5,1}$ the trimmed fraction used for estimation, and $f_{0.5,c}$ the fraction of the distribution for which the trimmed fraction, $f_{0.5,1}$, is affected by dropout (in violation of the strong MNAR assumption), see Figure 1.
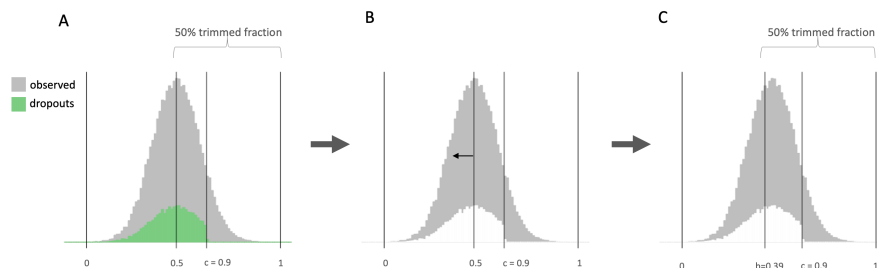
Figure 1: Schematic illustration of the dropout bias mechanism under assumption of homogeneity, for 50% trimming, and 20% dropout ($p_d = 0.2$) spread across 90% of the distribution ($f_{0,c} = 0.9$). A) dropouts (green) and observed patients (grey) for normally distributed outcomes of a given group. The 50% trimmed fraction is given by $f_{0.5,1}$, $f_{0.5,c}$ gives the fraction affected by dropout, $f_{c,1}$ the fraction unaffected by dropout, and $f_{0,0.5}$ the fraction that is trimmed away. B) Under dropout, $f_{0.5,1}$ lacks the observations to make up the 50% trimmed fraction, and the lower boundary, 0.5, shifts downwards in compensation. C) Under dropout, the 50% trimmed fraction is given by $f_{b,1}$. The size of the shift is calculated with (11), giving, for this example, $b = 0.39$.

The 50% trimmed mean is estimated by taking the top half of the distribution under assumption of worse value dropout. When the strong MNAR assumption is satisfied, the trimmed mean is given by $\mu_t = \mu_{0.5,1}$. When the assumption is violated, the fraction $f_{0.5,1}$ no longer contains a sufficient number of observations, and a larger fraction of the distribution, $f_{b,1}$, is taken, with $b < 0.5$ and the trimmed mean given by $\mu_{td} = \mu_{b,1}$. The shift from 0.5 to b is a function of the dropout proportion, p_d, the dropout spread, $f_{0,c}$, the trimming proportion, $f_{0,0.5} = p$, and $f_{0.5,c} = f_{0,c} - p$, with

$$b = f_{0,c} - f_{b,c} = f_{0,c} - \frac{f_{0,c} f_{0.5,c}}{(f_{0,c} - p_d)}. \tag{11}$$

Using this notation, the bias of the TM estimator in (10) can be expressed as

$$B_{tSM} = \mu_{0.5,1} - \mu_{b,1} = -\frac{f_{0.5,c}}{f_{0.5,1}} \sigma_0 (Q_{0.5,c} - Q_{b,c}), \tag{12}$$

with, for example, $Q_{bc} = \left(\phi\left(\Phi^{-1}(c)\right) - \phi\left(\Phi^{-1}(b)\right) \right) / (c - b))$. The full derivation of (12) is given in Appendix D. B_{tSM} (12) is a function of the size of the trimmed fraction ($f_{0.5,1}$), the fraction affected by dropout ($f_{0.5,c}$), the SD of the affected group (σ_0), and the magnitude of the shift from 0.5 to b (11), which is affected by the dropout proportion (p_d). When the strong MNAR assumption is satisfied, no shift occurs, since $b = 0.5$, so that $Q_{0.5,c} = Q_{b,c}$, and B_{tSM} reduces to 0.

Figure 1 illustrates the strong MNAR bias mechanism for the case of 50% trimming, with 20% dropout ($p_d = 0.2$) spread across 90% ($f_{0,c} = 0.9$) of the distribution. From (11), we obtain $f_{b,c} = \frac{0.9 \times 0.4}{(0.9 - 0.2)} = 0.51$ and $b = 0.9 - 0.51 = 0.39$.

With (12), we calculate $Q_{0.5,c} = 0.59$ and $Q_{b,c} = 0.34$. Specifying $\sigma_0 = 1$, the bias, B_{tSM}, is then given by $\frac{0.4}{0.5}(0.59 - 0.34) = 0.18$. The bias is positive, and the treatment effect is overestimated, resulting from underestimation of the comparator group mean due to dropout in the trimmed fraction.

The bias (12) is defined under the assumption of a homogeneous dropout mechanism. In reality, dropout is unlikely to be homogeneous. In any scenario, however, where the TM approach can be considered an appropriate estimator, there should be reason to believe that the dropout comprises on average lower - worse - outcome values. Then, homogeneous dropout can be considered a worst-case scenario and the bias calculated under this assumption will serve as an upper bound for the bias. This property is illustrated in supplementary Figure S2 (Appendix F).

Unlike the TM estimator, the CCA estimator will be biased under the strong MNAR assumption. We define this bias, B_C, for the case of dropout in the comparator group, with results for dropout in both groups derived in Appendix G. As for the TM estimator, we derive the bias under the assumption of a homogeneous dropout mechanism and a normally distributed outcome, with B_C given by

$$B_C = \beta_c - \beta = (\mu_1 - \mu_{c0}) - (\mu_1 - \mu_0) = -\mu_{c0} + \mu_0 = \mu_0 - \mu_{c0}, \qquad (13)$$

and

$$B_C = -\sigma_0 \left[\left(f_{0,c} - \frac{(f_{0,c} - p_d)}{(1 - p_d)} \right) \frac{\phi(\Phi^{-1}(c))}{c} + \right.$$

$$\left. \left((1 - f_{0,c}) - \frac{(1 - f_{0,c})}{(1 - p_d)} \right) \frac{-\phi(\Phi^{-1}(c))}{1 - c} \right], \qquad (14)$$

with $f_{0,c}$ the dropout spread, c its upper bound, and p_d the dropout proportion. We fully derive (14) in Appendix G.

In the absence of dropout ($p_d = 0$), and for dropout spread homogeneously across the entire distribution ($c = f_{0,c} = 1$), B_C is 0. The CCA estimate will be maximally biased under dropout restricted to the very lowest or highest values of the distribution, the former resulting in overestimation of the complete case mean, μ_{cj}, the latter in its underestimation. We can write the maximum bias for a given treatment group j under high or low value dropout as

$$| B_{C\max} | = \frac{\sigma_j}{(1 - p_{dj})} \phi(\Phi^{-1}(p_{dj})). \qquad (15)$$

The full derivation of (15) is given in Appendix H.

11

The dropout mechanism, under which maximum CCA bias is achieved, can be defined in terms of a threshold selection model. Consider dropout of the p_d highest values, so that all observations exceeding the $(1 - p_d)$'th quantile of the outcome distribution are unobserved. Let us define this threshold as $t_H = \mu_j + \sigma_j \Phi^{-1}(1 - p_{dj}) = \mu_j + \sigma_j \Phi^{-1}(p_{sj})$, so that

$$
p_s = \begin{cases} 0 & \text{if } y \geq t_H \\ 1 & \text{if } y < t_H. \end{cases} \tag{16}
$$

In the broader context of selection models in general, Copas and Jackson[7] defined a bias limit, which is equivalent to the maximum CCA estimator bias that we derive here (15). We extend Copas and Jackson's bias limit to the TM estimator, and define the maximum possible TM bias, which, for the case of lower value trimming, occurs under highest value dropout. For the comparator group, the maximum bias resulting from dropout is then given by

$$
B_{tSM\max} = \frac{-\sigma_0}{1 - p}\left[\phi\big(\Phi^{-1}(p_s)\big) - \phi\big(\Phi^{-1}(p_s - (1 - p))\big) + \phi\big(\Phi^{-1}(1 - p)\big)\right], \tag{17}
$$

with the total maximum bias, $B_{t\max}$, accounting for the potential violation of the location shift assumption, given by

$$
B_{t\max} = B_{tSM\max} + \left(\sigma_1 - \sigma_0\right)\frac{\phi\big(\Phi^{-1}(p)\big)}{1 - p}. \tag{18}
$$

Full derivations of (17, 18) are given in Appendix I, alongside expressions for maximum bias under dropout in the treatment group. Appendix J describes a simple simulation illustrating the maximum CCA and TM estimator biases under dropout assumption violations.

2.5   Adjusted estimator

Here, we define an adjusted estimator which relaxes the location shift assumption. Under the assumption of normally distributed outcomes, the 50% trimmed fractions will have half-normal distributions. The SD of a given fraction can be adjusted by mirroring this half-normal distribution, and rescaling this now complete, if artificial, normal distribution, using the full sample SD of the other group. The latter SD will either be observed, in the absence of dropout, or can alternatively be extrapolated from the observed fraction SD, using the underlying properties of normality. The adjustment can be performed on either treatment groups, but we illustrate the procedure for the comparator group. Let μ_{t0} be the unadjusted 50% comparator trimmed mean, with

$$
\mu_{t0} = \mu_0 + \sigma_0\phi\big(\Phi^{-1}(0.5)\big)/0.5, \tag{19}
$$

12

and μ_{at0} the adjusted comparator trimmed mean, with

$$\mu_{at0} = \frac{\mu_{t0} - \mu_0}{\sigma_0/\sigma_1} + \mu_0. \tag{20}$$

Substituting (19) for μ_{t0} in (20) then gives the comparator group trimmed mean
under the treatment group SD, σ_1. From the adjusted comparator trimmed
mean, μ_{at0} (20), and the unadjusted treatment trimmed mean, μ_{t1} (5), we
obtain the population adjusted TM estimate, under comparator group rescaling,
β_{at0}, as

$$\beta_{at0} = \mu_{t1} - \mu_{at0}.$$

Equivalently, we can obtain the population adjusted TM estimate, under treat-
ment group rescaling, $\beta_{at1} = \mu_{at1} - \mu_{t0}$, with μ_{at1} defined analogously to (20).
As for the unadjusted estimator, violation of the strong MNAR assumption will
bias the estimate, with violation in the comparator group resulting in an overes-
timation of the treatment effect, and the converse true for the treatment group.
We show in simulation (Section 3, Table 1) that this assumption is stronger for
the adjusted estimators than for the unadjusted estimator, with the greatest
bias observed when rescaling the group for which the strong MNAR assumption
is most violated, and recommend that the adjustment be applied to the group
for which the assumption is most plausible. In Appendix K, expressions are de-
rived for the adjusted estimator bias resulting from strong MNAR assumption
violation in either or both treatment groups, under comparator group rescaling
and under treatment group rescaling.

3   A simulation study

Consider a clinical trial on $n = 1000$ subjects randomized to either a treatment
($j = 1$) or comparator ($j = 0$) arm, with $n_1 = n_0 = 500$. We assume the out-
comes are normally distributed, with a true treatment effect $\beta = 0.5$, and 20%
dropout in the comparator group. Tables 1a and 1b give the mean treatment
effect estimates and bias across $S = 1000$ simulations for the CCA estimator
and the TM estimator under 50% trimming, for four scenarios. In the first
three, dropout is restricted to the lowest 20%, 50% and 75%, respectively, of
the comparator arm outcome distribution, and left completely unrestrained in
the fourth scenario.

In the first two scenarios, the strong MNAR assumption is satisfied and we
observe in Table 1a, for equal treatment SDs ($\sigma_1 = \sigma_0 = 1$), that the TM
estimate, $\widehat{\beta}_t = 0.5$, is unbiased, while the CCA estimate, $\widehat{\beta}_c$, is biased towards
the null. On violating the strong MNAR assumption, this bias decreases, with
$\widehat{\beta}_c$ being unbiased under random dropout ($f_{ac} = 1$). We observe the same

pattern for $\sigma_0 = 1.5$ and $\sigma_1 = 1$ (Table 1b), with the CCA bias now larger due to the higher σ_0, which acts multiplicatively on the bias (14). Under unequal SDs, the location shift assumption is violated, and the TM estimates are biased towards the null. For the first two scenarios, the total bias, B_t, is equal to B_{tLS}, while in scenarios 3 and 4, B_t is given by the sum of B_{tLS} and B_{tSM}, with the latter biasing the TM estimates away from the null. As for the CCA estimator, the increased comparator group SD results in a larger B_{tSM} bias in Table 1b.

Table 1: Estimated treatment effects $(\widehat{\beta})$ and corresponding biases $(\widehat{B})$ for the CCA $(\widehat{\beta}_c)$, TM $(\widehat{\beta}_t)$ and adjusted TM $(\widehat{\beta}_{at})$ estimators, with the latter obtained when performing the adjustment on the comparator group $(\widehat{\beta}_{at0})$ and the treatment group $(\widehat{\beta}_{at1})$. Mean and standard deviation (SD) estimates for $S = 1000$ simulations are reported. The TM estimands are estimated under 50% trimming of normally distributed outcomes $(N = 1000)$, for equal (a) and unequal (b) treatment group SDs, true treatment effect $\beta = 0.5$, and 20% comparator group dropout. Four scenarios are considered: dropouts restricted to 1) the lowest 20% of the distribution $(f_{ac} = 0.2$), 2) the lowest 50%, 3) the lowest 75% and 4) left unrestricted.

(a) $\sigma_0 = \sigma_1 = 1$

Dropout spread		$\widehat{\beta}_c$	$\widehat{\beta}_t$	$\widehat{\beta}_{at1}$	$\widehat{\beta}_{at0}$	$\widehat{B}_{tLS}$	$\widehat{B}_{tSM}$	$\widehat{B}_t$	$\widehat{B}_{at1}$	$\widehat{B}_{at0}$
1) $f_{0,c} = 0.2$	$\widehat{\beta}/\widehat{B}$	0.15	0.50	0.50	0.50	0.00	0.00	0.00	0.00	0.00
	$\widehat{SD}$	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.03	0.03
2) $f_{0,c} = 0.5$	$\widehat{\beta}/\widehat{B}$	0.30	0.50	0.50	0.50	0.00	0.00	0.00	0.00	0.00
	$\widehat{SD}$	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.03	0.03
3) $f_{0,c} = 0.75$	$\widehat{\beta}/\widehat{B}$	0.40	0.56	0.64	0.70	0.00	0.06	0.06	0.14	0.20
	$\widehat{SD}$	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.03	0.03
4) $f_{0,c} = 0.1$	$\widehat{\beta}/\widehat{B}$	0.50	0.69	0.77	0.81	0.00	0.19	0.19	0.27	0.31
	$\widehat{SD}$	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.03	0.03

(b) $\sigma_0 = 1.5$, $\sigma_1 = 1$

Dropout spread		$\widehat{\beta}_c$	$\widehat{\beta}_t$	$\widehat{\beta}_{at1}$	$\widehat{\beta}_{at0}$	$\widehat{B}_{tLS}$	$\widehat{B}_{tSM}$	$\widehat{B}_t$	$\widehat{B}_{at}$	$\widehat{B}_{at0}$
1) $f_{0,c} = 0.2$	$\widehat{\beta}/\widehat{B}$	-0.03	0.10	0.50	0.50	-0.40	0.00	-0.40	0.00	0.00
	$\widehat{SD}$	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
2) $f_{0,c} = 0.5$	$\widehat{\beta}/\widehat{B}$	0.20	0.10	0.50	0.50	-0.40	0.00	-0.40	0.00	0.00
	$\widehat{SD}$	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
3) $f_{0,c} = 0.75$	$\widehat{\beta}/\widehat{B}$	0.34	0.19	0.70	0.82	-0.40	0.09	-0.31	0.20	0.32
	$\widehat{SD}$	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.04	0.04
4) $f_{0,c} = 1$	$\widehat{\beta}/\widehat{B}$	0.50	0.39	0.91	0.96	-0.40	0.29	-0.11	0.41	0.46
	$\widehat{SD}$	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03

The adjusted TM estimates are given by $\widehat{\beta}_{at1}$ and $\widehat{\beta}_{at0}$, for rescaling applied to the treatment and comparator group, respectively. In the first two scenarios, the strong MNAR assumption is satisfied, and both estimators estimate an unbiased treatment effect, $\beta = 0.5$, under unequal SDs (Table 1b). For scenarios 3 and 4, however, both are biased away from the null, with bias components, $\widehat{B}_{at1}$ and $\widehat{B}_{at0}$, bigger than the unadjusted estimator bias, $\widehat{B}_{tSM}$. The greatest bias is observed for $\widehat{B}_{at0}$, a consequence of performing the adjustment on the group for

which the strong MNAR assumption is violated.

Examples of bias calculations for the CCA and (un)adjusted TM estimators
are given in Appendix L. Appendix M describes an additional simulation, for
trimming fractions, p, other than 50%, illustrating the bias trade-off between the
B_{tLS} and B_{tSM} bias components. Appendix N provides R code for obtaining
adjusted and unadjusted TM estimates for a simulated dataset.

4   An application to the CoBalT randomized clinical trial

The CoBalT trial[9,10] was a multicentre trial investigating the effect of CBT as
an adjunct to pharmacotherapy versus usual care (UC) in 469 patients aged
18-75 with treatment-resistant depression. The primary outcome was the BDI-
II score - a self-completed measure of depressive symptoms, with higher values
indicating greater depression. Long-term treatment effect was assessed by a
repeated measures intention-to-treat analysis ($N = 396$), using outcomes at 6
months, 12 months and 3-5 years (average 46 months) and adjusted for BDI-II
at baseline, treatment centre and minimization variables (previously prescribed
antidepressants, presence of a counselor at the general practice, and duration of
current depressive episode at baseline). The repeated measures analysis showed
a beneficial effect of CBT versus UC of -4.7 (95% CI: -6.4,-3.0) with a mean
difference of -3.6 (95% CI: -6.6,-0.6) at 46 months for 136 and 112 patients in
the CBT and UC group, respectively.[10]

4.1   Missing data

Just over half of the initially recruited patients were observed at the final follow-
up, with greater dropout in the UC group than the CBT group (Table 2).
Dropout in the UC group appears to be unrelated to the most recent depression
score, whereas in the CBT group, dropout was more common in those with
higher recent depression scores (Table 3). The CBT group SDs are higher, most
noticeably for patients missing at 46 months, suggesting that the unobserved
full sample CBT SD at 46 months may be higher than the UC group SD. In
Supplementary Figure S3 (Appendix O), histograms of the BDI-II score dis-
tributions are given, across time, for the UC group, the CBT group and their
combined total. We observe that the distributions are moderately right-skewed
and non-normal.

Table 2: Counts (N) and percentages (%) of patients with an observed BDI-II score at baseline, 6 months, 12 months, and 46 months, for the usual care group (UC), the CBT treatment group (CBT), and both groups

	UC		CBT		Total	
	N	%	N	%	N	%
Base	235	100	234	100	469	100
6 mo.	213	90.6	206	88.0	419	89.3
12 mo.	198	84.3	197	84.2	395	84.2
46 mo.	112	47.7	136	58.1	248	52.9

Table 3: Mean BDI-II measurements and SDs at baseline, 6 months, 12 months and 46 months for the usual care group (UC), the CBT treatment group. Patients missing and observed at 46 months are compared for all time points, with mean differences and 95% confidence intervals provided.

	Usual care (UC)			CBT		
	Observed[a]	Missing[a]	Difference[b]	Observed[a]	Missing[a]	Difference[b]
Base	30.9 (10.3)	32.7 (11.3)	-1.8 (-4.6,1.0)	30.4 (9.3)	33.6 (11.7)	-3.2 (-5.9,-0.5)
6 mo.	24.6 (13.1)	24.4 (13.2)	0.2 (-3.3,3.8)	17.4 (13.6)	21.8 (14.9)	-4.5 (-8.5,-0.4)
12 mo.	21.8 (13.3)	21.5 (12.4)	0.2 (-3.4,3.9)	15.8 (12.9)	19.5 (15.8)	-3.8 (-7.9,0.4)
46 mo.	23.4 (13.2)			19.2 (13.8)		

a) Given are mean and standard deviation (SD)

b) Given is the mean difference (observed - missing) with 95% confidence interval

4.2 Analysis details

Using the maximum CCA bias formula (15) from Section 2.4, we would expect the treatment effect to either under- or over-estimate the treatment effect by approximately 10 units of the BDI-II score. This is a wide bound, obtained without taking into account information available from the data and motivating the use of a more precise sensitivity analysis. To this end, we employed the TM approach and two MI models as a sensitivity analysis for the treatment effect at 46 months, which is estimated using linear regression adjusting for treatment centre, BDI-II at baseline and various minimization variables (previously described antidepressants, presence of a counselor at the general practice, duration of current depressive episode at baseline). Adaptive trimming was employed due to the high dropout proportion observed, with 52.3% of the highest values trimmed away in both groups. Confidence intervals were obtained using a permutation-based approach.[4] We considered only the unadjusted TM estimator, as the adjusted estimator (Section 2.5), which adjusts for unequal treatment group SDs, is strongly reliant on normality. The imputation models included auxiliary variables: baseline variables associated with missing BDI-II at any time of follow-up, and various measures of depression and anxiety (BDI-II, PHQ-7, GAD-7, SF-12). Two MI models were considered, with the first using only depression/anxiety information available at baseline (MI-baseline), the second including intermediate outcome measurements at all follow-up times.

To interpret the sensitivity analysis results, we used the bias formulae derived in Sections 2.3 and 2.4 to establish CCA and TM estimator bias directions under different dropout scenarios. Additionally, we calculated the expected TM estimator bias of the CBT treatment effect in the CoBalT data, for three variations of the most plausible dropout scenario. The location shift assumption bias and strong MNAR bias are both functions of the full sample SDs, which remain unobserved. Subject to the specified dropout mechanism, the observed dropout proportions, and under assumption of normally distributed outcomes, these can be inferred from the observed SDs, and used to calculate the relevant bias terms (derivations and R code in Appendix P).

## 4.3 Plausibility of assumptions for CCA, MI and TM estimators

The validity of each method rests on the underlying estimator assumptions about data characteristics, assumptions which cannot (usually) be tested. We used the bias formulae derived in Sections 2.3 and 2.4 to compare TM and CCA estimator behaviour across four plausible dropout scenarios Figure 2).

The CCA estimator will be biased if data are MNAR. Dropout of more depressed patients in the CBT group will result in underestimation of the CBT complete case mean, with the CCA estimator consequently overestimating the beneficial effect of CBT (Figure 2.2).

The TM estimator will be unbiased if the location shift assumption and strong MNAR assumption hold (Figure 2.1). From the CoBalT data, we suspect dropout of patients with more severe depressive symptoms in the CBT group (higher value dropout) and (approximately) homogeneous (MCAR) dropout in the UC group. This will, under higher value trimming, lead to the inclusion of too many high values in the trimmed fraction, and consequently to overestimation of the UC trimmed mean and the treatment effect (Figure 2.2). Intermediate BDI-II measurements at earlier times suggest a comparatively higher SD for the CBT group, implying that the location shift assumption may be violated (Figure 2.3), with the larger spread of values leading to underestimation of the CBT trimmed mean and overestimation of the beneficial effect of CBT. Violation of the location shift assumption, with $SD_{CBT} > SD_{UC}$, combined with violation of the strong MNAR assumption in the UC group, will exacerbate the bias, leading to a substantial overestimation of the CBT treatment effect. This is illustrated in Figure 2.4, where the bias is somewhat mitigated by allowing for the possibility that the strong MNAR assumption may also be violated, if to a much lesser extent, in the CBT group.

MI estimates will be unbiased if the data are MAR and the imputation model is correctly specified. Under higher-value dropout in the CBT group, the MI

models will impute BDI-II values that are too low in the CBT group, but approximately correct in the UC group, resulting in underestimation of the CBT imputed mean. As with the CCA estimator, this will result in a bias away from the null, with the MI estimators overestimating the benefit of CBT treatment. The inclusion of auxiliary information (baseline and intermediate depression measures) in the MI models could potentially overcome part of the MNAR bias, resulting in estimates that are closer to the truth than CCA. In summary, our bias formulae suggest that under plausible assumptions, TM, MI and CCA will overestimate the treatment effect.
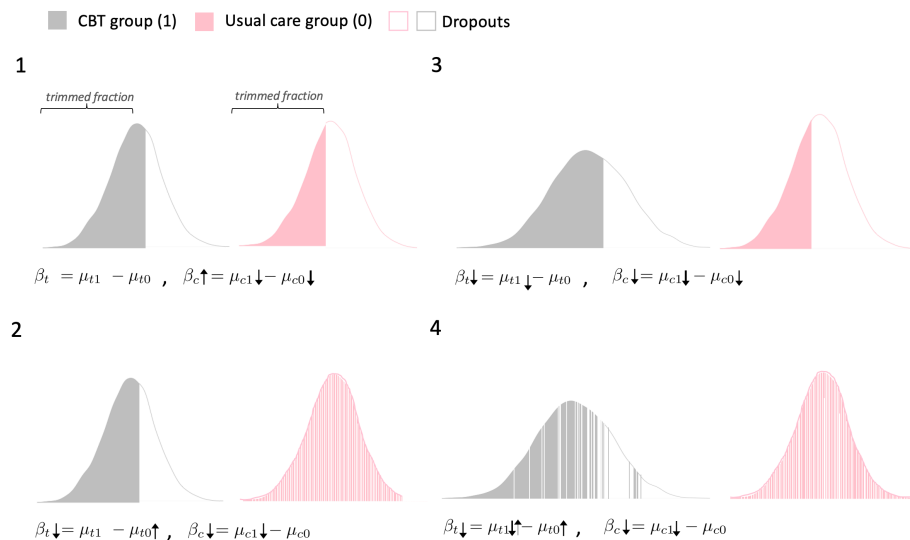
Figure 2: Trimmed means (TM) and complete case analysis (CCA) estimators across four scenarios with varying dropout patterns and treatment arm distributions, for 52.3% dropout in the usual care group (pink), 41.9% dropout in the CBT group (grey) and higher value 52.3% trimming. 1) The location shift assumption and strong MNAR assumption are satisfied, with equal treatment group SDs and strictly higher value dropout. 2) The location shift assumption is satisfied, but the strong MNAR assumption is violated in the usual care group. 3) The location shift assumption is violated, with $\sigma_1 > \sigma_0$, and the strong MNAR assumption is satisfied. 4) The location shift assumption is violated with $\sigma_1 > \sigma_0$, and the strong MNAR assumption is violated in the usual care group and to a lesser degree in the CBT group.

4.4 Results

The CCA treatment effect estimate for CBT of -3.89 (95% CI: -6.95,-0.83) is comparable to the one of the main CoBalT follow-up analysis[10] (Table 4). Slightly smaller effects are estimated by the MI models, comparable to the attenuation observed in the MI sensitivity analysis of the CoBalT follow-up analysis.[10] The TM estimator, in contrast, estimated a considerably bigger beneficial CBT treatment effect of -8.26 (95% CI: -11.20,-5.32). These results are

in line with our expectations regarding the BDI-II score and dropout character-istics, detailed in Section 4.4.

From the CoBalT data, we have reason to suspect higher value dropout in the CBT group and approximately homogeneous dropout in the UC group. Table 5 takes a closer look at the TM estimator bias components calculated under three different assumptions of the dropout mechanism. In scenario A, we assume completely directional, highest value dropout in the CBT group, with all dropout values (41.9%) restricted to the top of the distribution, and entirely homogeneous dropout in the UC group, with the dropout (53.3%) spread equally across the distribution. In scenarios B and C, we make less stringent assumptions, with in scenario B dropout in the top 60% for the CBT group, and in scenario C dropout in the top 60% of the CBT group and top 80% of the UC group. We obtain a total bias, $\widehat{B}_t$ of -17.53, -11.09, and -7.1 for scenarios A, B and C, respectively, and calculate a bias-adjusted estimate, which can be interpreted as an upper bound for the treatment effect estimate, $\widehat{\beta}_t$, with the lower bound given by the TM estimate obtained from the data ($\widehat{\beta}_t = -8.26$). For all three scenarios, the CCA and MI estimates (Table 4) fall within the $\widehat{\beta}_t$ bounds, with the bias-adjusted estimate of scenario C most comparable, at -1.16. Our sensitivity analysis, comparing CCA, TM and MI, is consistent with a beneficial CBT treatment effect, but suggests that it may be more modest than indicated by the CCA estimator.

Table 4: Treatment effect estimate ($\widehat{\beta}$)) standard error (SE), and 95% confidence interval (95% CI), for the trimmed means (TM) estimator, complete case analysis (CCA), multiple imputation model with baseline outcome measurements (MI baseline) and multiple imputation model with intermediate outcome measurements (MI intermediate).

	$\widehat{\beta}$	SE	95% CI
TM	-8.26	1.47	(-11.2, -5.03)
CCA	-3.89	1.53	(-6.92, -0.87)
MI (baseline)	-3.33	1.39	(-6.07, -0.58)
MI (intermediate)	-2.56	1.37	(-5.65, -0.08)

Table 5: Trimmed mean estimator treatment effect bounds and bias components under different dropout assumptions. A) CBT group dropout restricted to the top 41.9% of the distribution and homogeneous UC group dropout; B) CBT group dropout in the top 60% of the distribution and homogenous UC group dropout; C) CBT group dropout in the top 60% of the distribution and UC group dropout in the top 80%. Shown are the bias components for violation of the location shift assumption ($\widehat{B}_{LS}$), the strong MNAR assumption in the UC group ($\widehat{B}_{SM0}$), the strong MNAR assumption in the CBT group ($\widehat{B}_{SM1}$), the total bias ($\widehat{B}_t$), the bias adjusted estimate ($\widehat{\beta}_{tBA}$), and the maximum bounds of the TM estimate, $\widehat{\beta}_t$.

	Group	Spread	$\widehat{B}_{LS}$	$\widehat{B}_{SM0}$	$\widehat{B}_{SM1}$	$\widehat{B}_t$	$\widehat{\beta}_{tBA}$	TM estimate bounds
A	UC CBT	1 0.419	-7.17	-10.36	0	-17.53	9.27	[-8.26, 9.27]
B	UC CBT	1 0.6	-1.14	-10.36	0.41	-11.09	2.83	[-8.26, 2.83]
C	UC CBT	0.8 0.6	-2.01	-5.5	0.41	-7.1	-1.16	[-8.26, -1.16]

5   Discussion

Our work extends existing TM literature by deriving formulae that can be used to calculate the resulting bias, establish bias direction, and also, under certain additional assumptions, calculate a limit for the maximum expected bias. In Section 4.3, we show how these formulae can be used to aid interpretation of sensitivity analyses. In addition to this, we describe an adjustment to the estimator that relaxes the location shift assumption.

Missing data is a common feature in RCTs and may result in biased inference, with any analysis relying on unverifiable assumptions about the relationship between observed and missing data. To increase confidence in the primary results, their robustness should be assessed by performing sensitivity analyses under a range of plausible alternative assumptions.[1,2,11,12] While conducting such analyses in the presence of missing data is recommend practice, only a minority of affected trials report performing any kind of sensitivity analysis. Six review articles describing the handling of missing data in RCTs found that only 22% of the 649 trials reviewed reported some kind of sensitivity analysis.[13-18] The majority retained the missingness assumptions of the original analysis, with only a small subset relaxing the original assumptions, and almost none considering a MNAR mechanism.[13]

Applying sensitivity analyses can be limited by the complexity of the methods, and a lack of transparency about the assumptions underlying each method. Two well-known approaches for dealing with incomplete data, which can be adapted to sensitivity analyses, are selection models (SM) and pattern mixture models (PMM). Selection models specify the relationship between outcome and the probability of being missing, while PMM define the outcome distribution for dropout across a range of missingness patterns.[2] An example is the delta-adjusted PMM, which assumes that patients who drop out have a mean outcome that differs by a fixed amount, *delta*, to the ones who remain in study.[12] This parameter, *delta*, is user-specified and typically informed by expert knowledge, in an attempt to characterize the unknown relationship between observed and unobserved data. However, choosing plausible values and distributions for sensitivity analysis parameters such as *delta* is far from straightforward.[19,20] A simpler alternative is to specify some extreme scenarios, and, under these, re-estimate the main analysis results. An example of such an extreme scenario analysis is the combination approach of best-worst and worst-best case sensitivity analyses, in which beneficial outcomes are assigned to dropouts in one group and harmful ones to dropouts in the other group, and vice versa, yielding two estimates under opposing assumptions[21,22]. The TM estimator, operating under the assumption of strict directional dropout and requiring only the specification of the size of trimming fraction, can be considered part of this family of easily implemented extreme-case scenario estimators. Such simple bias-sensitivity analyses will often be sufficient to assess the robustness of inferences to potential bias sources[11], but in the event of ambiguity, can be followed up with a more nuanced and fine-tuned sensitivity analysis (e.g. from the family of SM or PMM models).

While the TM estimator is simpler to implement than SM and PMM models, requiring no explicit specification of a missingness model or numerical sensitivity parameters, it is theoretically similar. The TM estimator can be thought of as a threshold SM, with the threshold value informed by the size of the trimming fraction, and with the probability of missing set to zero in the trimmed fraction and left unspecified in the part of the distribution that is trimmed away. A similar parallel can be drawn between TM and the delta-adjusted PMM, with the *delta* mean difference between observed and missing values implicit in the former and a direct result of the choice of trimming fraction. Under the strong MNAR assumption, this difference should be equal to or bigger than the difference between the mean of the trimmed fraction and the mean of the fraction that is trimmed away.

The TM estimator will give an unbiased estimate of the true treatment effect under the strong MNAR assumption and the location shift assumption. The TM estimator is only unbiased for the particular case of MNAR data where the trimmed fraction remains unaffected by dropout, and will be biased, even under the null, when outcomes are MAR. In practice, missing data will frequently

be characterized by more than a single missingness mechanism (e.g., missing outcomes may be a mix of MAR and MNAR). Ocampo et al.[5] sought to resolve this by combining the TM estimator with multiple imputation. This approach, however, relies on the assumption that missing values determined by MAR and MNAR can be distinguished using auxiliary information and its success will be heavily affected by the reliability and interpretation of the available information.

As assumptions underlying sensitivity analyses cannot be verified, it is important to know how different underlying data structures affect a given estimator. While previous studies have described the TM assumptions[4-6], the biases arising from the violation of these strict assumptions have not previously been characterized. By quantifying the bias resulting from violation of these assumptions and establishing bias direction under a range of plausible data scenarios, we show how conclusions can be drawn from a sensitivity analysis, moving beyond a comparison of estimates. For example, in the CoBalT analysis (Section 4.4), the TM estimator, under very strict assumptions, estimated a beneficial effect of CBT treatment of approximately -8. However, when considering more plausible data generating scenarios and relaxing these assumptions, we obtained a bias-corrected estimate, which was much smaller.

In order to identify relevant sensitivity analyses, a framework, such as the one employed for the CoBalT analysis in Section 4.3, should be followed. Here, our initial bias results suggest that all estimators will be biased away from the null and overestimate the CBT treatment effect. The logical next step would involve identifying sensitivity analyses that might be biased in the opposite direction for a plausible data scenario. The TM estimator will be particularly useful in cases where it can be expected to be biased in the opposite direction to the CCA analysis. For example, if we had seen evidence of higher value dropout in both groups (i.e., previous depression being higher for those with missing outcome data in both UC and CBT groups) and evidence of similar SDs in the two arms, then we would expect the CCA estimator to underestimate the treatment effect given greater dropout in the comparator group, and overestimate the effect given higher dropout in the treatment group, while the TM estimator would remain unbiased for both.

Commonly applied sensitivity analyses employ complex models, for which bias quantification methods are not readily available. In contrast, the TM estimator is a simple approach, for which, as we show, analytic expressions for the bias due to missingness are available. In this paper, we show how the TM estimator can be used as a sensitivity analysis, by using bias formulae to interpret results in context of information available from the data, as we show in our CoBalT analysis. We recommend performing a TM sensitivity analysis when directional dropout is plausible, using the bias formulae to establish bias direction and interpret estimates.

Declarations

Data Statement

Data from the CoBalT trial is available on request from the authors

Funding statement

Declaration of interest statement

The authors declare no conflicts of interest.

Acknowledgments

References

1. Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med.* 2013;86(3):343-358.

2. Little R, Rubin D. Statistical analysis with missing data. New York: Wiley; 1987. 278 p.

3. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epidemiol.* 2019;48(4):1294-1304.

4. Permutt T, Li F. Trimmed means for symptom trials with dropouts. *Pharm Stat*. 2017;16(1):20-28.

5. Ocampo A, Schmidli H, Quarg P, Callegari F, Pagano M. Identifying treatment effects using trimmed means when data are missing not at random. arXiv:1908.01044 [stat.ME]. 2019.

6. Wang MD, Liu J, Molenberghs G, Mallinckrodt C. An evaluation of the trimmed mean approach in clinical trials with dropout. *Pharm Stat*. 2018;17(3):278-289.

7. Copas J, Jackson D. A bound for publication bias based on the fraction of unpublished studies. *Biometrics*. 2004;60(1):146-153.

8. O'Neill B. Some useful moment results in sampling problems. *American Statistician*. 2014;68(4):282-296.

9. Wiles N, Thomas L, Abel A, et al. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. *Lancet*. 2013;381(9864):375-384.

10. Wiles NJ, Thomas L, Turner N, et al. Long-term effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: follow-up of the CoBalT randomised controlled trial. *Lancet Psychiatry*. 2016;3(2):137-144.

11. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969-1985.

12. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Stat Med*. 2020;39(21):2815-2842.

13. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs; a review of the top medical journals. *BMC Med Res Methodol*. 2014;14:118. Published 2014 Nov 19. doi:10.1186/1471-2288-14-118

14. Hussain JA, Bland M, Langan D, Johnson MJ, Currow DC, White IR. Quality of missing data reporting and handling in palliative care trials demonstrates that further development of the CONSORT statement is required: a systematic review. *J Clin Epidemiol*. 2017;88:81-91.

15. Rombach I, Knight R, Peckham N, Stokes JR, Cook JA. Current practice in analysing and reporting binary outcome data-a review of randomised controlled trial reports. *BMC Med*. 2020;18(1):147. doi:10.1186/s12916-020-01598-7.

16. Rombach I, Rivero-Arias O, Gray AM, Jenkinson C, Burke Ó. The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: a review of the current literature. *Qual Life Res.* 2016;25(7):1613-1623.

17. Thabane L, Mbuagbaw L, Zhang S, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol.* 2013;13:92. Published 2013 Jul 16. doi:10.1186/1471-2288-13-92

18. Wood A, White I, Thompson S. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials.* 2004;1(4):368–376.

19. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. Clin Trials. 2007;4(2):125-139.

20. Mason AJ, Gomes M, Grieve R, Ulug P, Powell JT, Carpenter J. Development of a practical approach to expert elicitation for randomised controlled trials with missing health outcomes: Application to the IMPROVE trial. *Clin Trials.* 2017;14(4):357-367.

21. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC Med Res Methodol.* 2017;17(1):162. Published 2017 Dec 6. doi:10.1186/s12874-017-0442-1.

22. Jakobsen JC, Wetterslev J, Winkel P, Lange T, Gluud C. Thresholds for statistical and clinical significance in systematic reviews with meta-analytic methods. *BMC Med Res Methodol.* 2014;14:120. Published 2014 Nov 21. doi:10.1186/1471-2288-14-120.