Empirical Bayes estimation of the selected treatment mean for two-stage drop-the-loser trials: a meta-analytic approach

Jack Bowden,[a]*[†] Werner Brannath[b] and Ekkehard Glimm[c]

Point estimation for the selected treatment in a two-stage drop-the-loser trial is not straightforward because a substantial bias can be induced in the standard maximum likelihood estimate (MLE) through the first stage selection process. Research has generally focused on alternative estimation strategies that apply a bias correction to the MLE; however, such estimators can have a large mean squared error. Carreras and Brannath (*Stat. Med.* 32:1677-90) have recently proposed using a special form of shrinkage estimation in this context. Given certain assumptions, their estimator is shown to dominate the MLE in terms of mean squared error loss, which provides a very powerful argument for its use in practice. In this paper, we suggest the use of a more general form of shrinkage estimation in drop-the-loser trials that has parallels with model fitting in the area of meta-analysis. Several estimators are identified and are shown to perform favourably to Carreras and Brannath's original estimator and the MLE. However, they necessitate either explicit estimation of an additional parameter measuring the heterogeneity between treatment effects or a quite unnatural prior distribution for the treatment effects that can only be specified after the first stage data has been observed. Shrinkage methods are a powerful tool for accurately quantifying treatment effects in multi-arm clinical trials, and further research is needed to understand how to maximise their utility. © 2013 The Authors. Statistics in Medicine published by John Wiley & Sons, Ltd.

Keywords: drop-the-loser trials; empirical Bayes estimation; meta-analysis; temporal coherency

1. Introduction

Two-stage drop-the-loser designs provide a framework for picking the most effective treatment out of a larger group of candidates and then testing it against a standard therapy in a confirmatory analysis. Although this design is an efficient way to discover effective treatments, the selection mechanism acts to inflate the type I error of the final test statistic [1, 2] and can also induce a substantial bias into the standard maximum likelihood estimate (MLE). With regard to the former, current regulatory authority guidance (e.g. [3]) is unequivocal that the final analysis must control the type I error rate. With regard to the latter, whilst acknowledging that estimation bias is a serious issue affecting the validity of adaptive trials and that bias should be 'minimised', there is a distinct lack of guidance and consensus on how this should be achieved. Research has generally focused on estimators that apply a bias correction to the MLE. One such class of estimators, referred to as uniform minimum variance conditionally unbiased estimators (UMVCUEs), totally removes the MLE's bias [4–7]. Others have proposed iterative or likelihood-based methods that can substantially reduce the bias of the MLE, without being unbiased [7–9]. Unfortunately, methods that explicitly target bias correction generally lead to an estimator with a mean squared error (MSE) larger than that of the MLE.

[a]*MRC Biostatistics Unit Hub for Trials Methodology Research, Cambridge, U.K.*
[b]*Competence Center for Clinical Trials Bremen, Faculty 3, University of Bremen, Bremen, Germany*
[c]*Novartis Pharma AG, CH-4002 Basel, Switzerland*
Correspondence to: Jack Bowden, MRC Biostatistics Unit, IPH, Robinson Way, Cambridge CB2 0SR, U.K.
[†]*E-mail: jack.bowden@mrc-bsu.ac.uk*

Carreras and Brannath [10] have recently proposed the use of shrinkage estimation [11] within the context of a two-stage drop-the-loser trial. In their method, the stage 1 data on all treatments are used to define a shrinkage estimate for the selected treatment at stage 1. This is then combined with the stage 2 estimate for the selected treatment via a weighted average. Under the assumption that the true treatment effects are independent and follow a common normal distribution (with any mean and variance), their estimator is shown to dominate the MLE in terms of MSE, thus providing a very powerful argument for its use in practice. In this paper, we propose an alternative shrinkage estimation strategy for drop-the-loser designs. Our approach is, in some ways, a simpler estimation procedure to that of [10] because it uses all of the available data within a single, standard, shrinkage equation. However, this apparent simplicity does impose some additional complications, which are discussed at length herein.

In Section 2, we introduce our notation for the two-stage drop-the-loser design. In Section 3, we describe the general principle of shrinkage estimation, Carreras and Brannath's original application of shrinkage estimation to the drop-the-loser trial context, and also present our alternative approach. In Section 4, we introduce several shrinkage estimators that naturally flow from our alternative formulation, and in Section 5, we evaluate the performance of the existing and alternative shrinkage estimators for various two-stage drop-the-loser design scenarios. We conclude in Section 6 with a discussion of the issues raised and point to further avenues of research.

2. The two-stage drop-the-loser design

Let $X_i \sim N\left(\mu_i, \sigma_1^2\right)$, $i = 1, \ldots, k$, be the effect estimates (MLEs) of k experimental treatments $T_1, \ldots, T_k$ at the first stage of a two-stage trial. The common variance term, σ_1^2, is assumed to be known. Assuming that large values indicate the most benefit, the 'best' treatment, T_s, $s \in \{1, \ldots k\}$, is selected as the one with the top-ranking MLE. That is, $X_s = \text{Max}\{X_1, \ldots, X_k\}$. Treatment T_s is taken forward in isolation for testing on an independent population in stage 2. Let $Y_s \sim N\left(\mu_s, \sigma_2^2\right)$ be the estimate for μ_s at stage 2.

Let X_0 and Y_0 represent the normally distributed treatment effect estimates for the control group at stages 1 and 2, with mean μ_0 and variances σ_1^2 and σ_2^2, respectively. At the end of the trial, we are interested in estimating the contrast $\mu_s - \mu_0$. Because the control group always proceeds to the final stage, μ_0 is unbiasedly estimated by its MLE, and we therefore focus our attention on estimation of μ_s only. The MLE of μ_s at stage 2 and its (assumed) variance are given by

$$\hat{\mu}_s = \frac{\sigma_2^2 X_s + \sigma_1^2 Y_s}{\sigma_1^2 + \sigma_2^2}, \quad \text{Var}\left(\hat{\mu}_s\right) = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{1}$$

Because it ignores the selection of X_s, $\hat{\mu}_s$ is positively biased (potentially seriously so), and $\text{Var}\left(\hat{\mu}_s\right)$ is also incorrect. We can express the most efficient unbiased estimate for μ_s as

$$\tilde{\mu}_s = \hat{\mu}_s - \frac{\sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \frac{\phi\left[\frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1^2}\left(\hat{\mu}_s - X_r\right)\right]}{\Phi\left[\frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1^2}\left(\hat{\mu}_s - X_r\right)\right]}, \tag{2}$$

where $\phi(.)$ and $\Phi(.)$ are the standard normal density and distribution functions, respectively, and where X_r is defined as the second best performing treatment at stage 2, that is, $X_r = \text{Max}\{X_1, \ldots, X_k\}/X_s$. This is referred to as the UMVCUE [4, 6].

2.1. Assessing estimators of μ_s

For any estimator of μ_s, μ_s^*, we can express its bias and MSE as

$$\text{Bias}\left(\mu_s^*\right) = \sum_{i=1}^{k} \text{E}\left[\mu_s^* - \mu_i | T_s = T_i\right] P(T_s = T_i),$$

$$\text{MSE}\left(\mu_s^*\right) = \sum_{i=1}^{k} \text{E}\left[\left(\mu_s^* - \mu_i\right)^2 | T_s = T_i\right] P(T_s = T_i). \tag{3}$$

This form (from Posch [12]) makes clear that at the trial outset, μ_s is a random variable. By definition, $\mathrm{Bias}(\tilde{\mu}_s) = 0$, and $\mathrm{MSE}(\tilde{\mu}_s)$ is smaller than any other μ_s^* that is also unbiased. This of course does not mean that $\mathrm{MSE}(\tilde{\mu}_s)$ is smaller than any biased estimator; for example, it is generally true that $\mathrm{MSE}(\tilde{\mu}_s) \gg \mathrm{MSE}(\hat{\mu}_s)$ (see Section 5 for example). Furthermore, it is commonly agreed that MSE (a measure equivalent to its variance + squared bias) provides a far better summary of an estimator's worth than bias alone.

3. Shrinkage estimation

The standard motivation for using shrinkage methods is to provide simultaneous, accurate estimation for a group of parameters, where accuracy is defined via the combined MSE. For example, if we were interested in jointly estimating the true mean effect of all k treatments $\mu_1, ..., \mu_k$ using only stage 1 data, then it will generally be true that the combined MSE,

$$\sum_{i=1}^{k} \mathrm{E}\left[(\mu_i^* - \mu_i)^2\right], \tag{4}$$

is far smaller when μ_i^* equals $\check{\mu}_i^L$ as opposed to the MLE X_i, where $\check{\mu}_i^L$ is Lindley's estimator [13]:

$$\check{\mu}_i^L = \hat{B}_+ X_i + (1 - \hat{B}_+)\bar{X} \tag{5}$$

and

$$\hat{B}_+ = \max\left\{0, 1 - \hat{C}\right\}, \quad \hat{C} = \frac{(k-3)\sigma_1^2}{\sum_{i=1}^{k}(X_i - \bar{X})^2}, \quad \bar{X} = \sum_{i=1}^{k} X_i / k. \tag{6}$$

Although shrinkage formula (5) was not originally proposed using a Bayesian argument, it can be easily understood and shown to be optimal within a Bayesian framework. Assume that *a priori* $\mu_1, ..., \mu_k$ are themselves independent and identically distributed (i.i.d) $N(\mu, \tau^2)$ random variables and only stage 1 data are available for the k treatments. Given μ_i, the distribution of its MLE X_i is $\hat{\mu}_i|\mu_i \sim N\left(\mu_i, \sigma_1^2\right)$. The *posterior* distribution of μ_i given $\hat{\mu}_i$ is then

$$\mu_i|\hat{\mu}_i \sim N\left(\frac{\tau^2}{\sigma_1^2 + \tau^2}\hat{\mu}_i + \frac{\sigma_1^2}{\sigma_1^2 + \tau^2}\mu, \frac{\sigma_1^2 \tau^2}{\sigma_1^2 + \tau^2}\right). \tag{7}$$

$\check{\mu}_i^L$ can therefore be viewed as an 'Empirical Bayes' estimate for the posterior mean of equation (7), with $\bar{X}$, X_i and $\hat{C}$ substituted for μ, $\hat{\mu}_i$ and $\frac{\sigma_1^2}{\sigma_1^2 + \tau^2}$, respectively. $\bar{X}$ is clearly an unbiased estimate of μ, but it is perhaps less obvious that $\hat{C}$ is an unbiased estimate for $\frac{\sigma_1^2}{\sigma_1^2 + \tau^2}$, regardless of the true value of τ^2. If $1 - \hat{C}$ gives a negative value, it is replaced by 0 in the definition of $\hat{B}_+$. This 'plus rule' has been shown to further reduce the MSE of the resulting estimate $\check{\mu}_i^L$ [14].

3.1. Carreras and Brannath's approach

Hwang [15] explicitly considered estimation of a single mean parameter from a k component system, where all k components have normally distributed estimates with a common variance and the single component is identified by having the largest estimate. This is identical to estimating μ_s using only stage 1 data in a two-stage drop-the-losers trial. He proved that when the treatment means follow a $N(\mu, \tau^2)$ prior distribution (for any μ and $\tau^2 \geqslant 0$), so that their posterior distributions obey equation (7) and $\check{\mu}_s^L$ is defined by equation (5) with $i = s$ (i.e. drop-the-losers selection), then the dominance result $\mathrm{MSE}\left(\check{\mu}_s^L\right) \leqslant \mathrm{MSE}(X_s)$ holds. Within the context of a two-stage drop-the-loser trial, Carreras and Brannath use Hwang's result to show that their estimator for μ_s at stage two

$$\check{\mu}_s^{CB} = t\check{\mu}_s^L + (1 - t)Y_s, \quad \text{for} \quad t = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{8}$$

analogously dominates $\hat{\mu}_s$ from equation (1). Their result relies on the fact that equation (1) is equivalent to replacing $\check{\mu}_s^{CB}$ in equation (8) with X_s. This also occurs naturally when the shrinkage factor C is set to 0.

3.2. An alternative formulation

Carreras and Brannath's method for estimating μ_s following a two-stage drop-the-loser trial can itself be viewed as a two-stage approach. That is, only the stage 1 data are used in the standard shrinkage estimator $\check{\mu}_s^L$, and then, stage 2 data on the selected treatment, Y_s, are added separately in equation (8) afterwards. Although $\check{\mu}_s^{CB}$ has the nice dominance property over the MLE, it is useful to consider whether it can itself be improved upon. For example, why not use all of the stages 1 and 2 data to define a single shrinkage estimator? Although this sounds straightforward, it does impose some extra complications. Despite being only truly concerned with estimation of the top-ranking treatment's mean, μ_s, Carreras and Brannath's method is defined to find shrinkage estimates given k parameter estimates with a *common* variance. This means that we are assuming the posterior distribution for μ_i implied by (7) (ignoring the stage 2 data), which enables the use of $\check{\mu}_s^L$ from (5). However, if we use all of the data (including the stage 2 data Y_s), we may assume the following set-up:

$$\mu_i \sim N\left(\mu, \tau^2\right),$$
$$\hat{\mu}_i | \mu_i \sim N(\mu_i, W_i), \quad \text{where}$$
$$\hat{\mu}_i = X_i, W_i = \sigma_1^2 \quad \text{if } i \neq s \quad \text{or} \quad \hat{\mu}_i = \frac{\sigma_2^2 X_i + \sigma_1^2 Y_i}{\sigma_1^2 + \sigma_2^2}, W_i = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad \text{if } i = s.$$

Further assuming that the μ_is are stochastically independent and the $\hat{\mu}_i$s are conditionally independent given μ_i, then

$$\mu_i | \hat{\mu}_i \sim N\left(\frac{\tau^2}{W_i + \tau^2}\hat{\mu}_i + \frac{W_i}{W_i + \tau^2}\mu, \frac{W_i \tau^2}{W_i + \tau^2}\right) \tag{9}$$

becomes the distribution with which to construct shrinkage estimates for the μ_is. The fact that W_i is not constant makes specification of single appropriate shrinkage factor C and estimate for the grand mean μ far less straightforward. In Section 4, we will discuss estimation for this new setting, pointing out its connection with model fitting in the area of meta-analysis.

4. Estimation for the new target

4.1. Direct estimation

Under the framework of equation (9), if the parameters μ and τ^2 were known, the best estimate for μ_s is given by

$$\frac{\tau^2}{W_s + \tau^2}\hat{\mu}_s + \frac{W_s}{W_s + \tau^2}\mu \tag{10}$$

Furthermore, with infinite data, it is clear that directly replacing μ and τ^2 in this expression with consistent estimates, $\hat{\mu}$ and $\hat{\tau}^2$, would give equation (10). If one assumes that τ^2 is known, then the variance of the posterior distribution in (9) also becomes completely known. It is then possible to show that the MLE for μ is

$$\hat{\mu}_{\tau^2} = \frac{\left(\sigma_1^2 \sigma_2^2 + \tau^2 \sigma_2^2\right) X_s + \left(\sigma_1^4 + \tau^2 \sigma_1^2\right) Y_s + (k-1)q\bar{X}_{-s}}{kq + \sigma_1^4}, \tag{11}$$

where $q = \sigma_1^2 \sigma_2^2 + \tau^2\left(\sigma_1^2 + \sigma_2^2\right)$ and $\bar{X}_{-s} = 1/(k-1)\sum_{i \neq s} X_i$. τ^2 could then be estimated by maximising the posterior likelihood of (9) with respect to τ^2 at $\mu = \hat{\mu}_{\tau^2}$. See the Appendix for further details. We will refer to the estimate for μ_s obtained by directly plugging in the previous estimates to equation (10) as $\hat{\mu}_s^{MPL}$. Although this approach is fairly crude, it will be interesting to observe the performance of $\hat{\mu}_s^{MPL}$ compared to several 'legitimate' shrinkage estimators that are now introduced.

4.2. Shrinkage estimation: Carter and Rolph's standard prior approach

Carter and Rolph [16] investigate shrinkage methods for jointly estimating the parameters of a k component system with unequal variances, as set up in equation (9). We will use it specifically to yield an estimate for μ_s in the context of a two-stage drop-the-loser trial of the form

$$\hat{B}_{+s}\hat{\mu}_s + \left(1 - \hat{B}_{+s}\right)\hat{\mu}, \quad \text{where} \quad \hat{B}_{+s} = \max\left\{0, 1 - \hat{C}_s\right\}, \tag{12}$$

so as to approximate equation (10). Although not immediately obvious, it can be shown that their suggested approach is equivalent the following procedure. First define a new weight $V_i = (W_i + \tau^2)$, and find the value of τ^2 that solves $Q(\tau^2) = k - 1$, where

$$Q(\tau^2) = \sum_{i=1}^{k} V_i^{-1}\left(\hat{\mu}_i - \hat{\mu}(\tau^2)\right)^2, \quad \text{for} \quad \hat{\mu}(\tau^2) = \frac{\sum_{i=1}^{k} V_i^{-1}\hat{\mu}_i}{\sum_{i=1}^{k} V_i^{-1}}. \tag{13}$$

If $Q(\tau^2) = k - 1$ for a $\tau^2 < 0$, then the estimate is truncated to 0. $Q(\tau^2)$ is known in the area of medical meta-analysis as the 'generalised' Q statistic [17, 18], and this estimation method for μ is known as the Paule–Mandel method of moments algorithm [19, 20]. Let the estimate for τ^2 derived in this manner be equal to $\hat{\tau}^2_{PM}$. The shrinkage factor for μ_s is then approximated by

$$\hat{C}_s = \frac{(k-3)W_s}{\left(\hat{\tau}^2_{PM} + \overline{W}\right)Q\left(\hat{\tau}^2_{PM}\right) + (k-3)(W_s - \overline{W})}, \quad \text{where} \quad \overline{W} = \sum_{i=1}^{k} W_i/k. \tag{14}$$

The form of $\hat{C}_s$ may appear complicated, but it has the nice property that if the W_is are all equal, then it reduces to the original $\hat{C}$ given in equation (6). In our context, the W_is can only approach equality as $\sigma_2^2 \to \infty$ or as the stage 2 sample size tends to zero. $\hat{C}_s$ can be inserted into equation (12) along with the MLE $\hat{\mu}_s$ and grand mean $\hat{\mu}$ given by $\hat{\mu}\left(\hat{\tau}^2_{PM}\right)$, to yield a new shrinkage estimator for μ_s. We will refer to this estimator as $\check{\mu}_s^{\tau^2}$ – the 'τ^2' denoting that this parameter is additionally and explicitly estimated.

4.3. Shrinkage estimation: Carter and Rolph's proportional prior approach

Carter and Rolph [16] also propose an alternative method for applying shrinkage estimation within the unequal variance context, which usefully avoids an iterative estimation of τ^2. It relies on the assumption of a different prior distribution for the treatment parameters, namely $\mu_i \sim N(\mu, W_i\tau^2)$. This asserts that the prior uncertainty around each treatment's mean is directly proportional to the variance of its estimate, $\hat{\mu}_i$. By replacing τ^2 with $W_i\tau^2$ in (9), it is clear that this implies the posterior distribution:

$$\mu_i|\hat{\mu}_i \sim N\left(\frac{\tau^2}{1+\tau^2}\hat{\mu}_i + \frac{1}{1+\tau^2}\mu, \frac{W_i\tau^2}{(1+\tau^2)}\right), \tag{15}$$

the mean of which becomes an alternative target to estimate. Because this mean does not depend on W_i, it suffices to calculate a single shrinkage factor $\hat{C}$ using all of the data. This can then, in conjunction with an estimate for μ, be used to estimate μ_s via equation (12). Turning first to estimation of μ: Under the proportional prior, the unconditional distribution of the estimates is $\hat{\mu}_i \sim N\left(\mu, \sigma_i^2(1 + \tau^2)\right)$. Therefore, given weights $V_i^* = \sigma_i^2(1 + \tau^2)$, the inverse variance weighted average

$$\hat{\mu}(\tau^2) = \frac{\sum_{i=1}^{k} V_i^{-1*}\hat{\mu}_i}{\sum_{i=1}^{k} V_i^{-1*}} = \hat{\mu}(0) \quad \forall \tau^2.$$

$\hat{\mu}(0)$ is referred to in meta-analysis as the 'fixed-effects' estimate for μ, as opposed to a 'random-effects' estimate, of which $\hat{\mu}\left(\hat{\tau}^2_{PM}\right)$ is an example. Of course, when τ^2 is estimated to be 0, they are equal.

Turning now to estimation of μ_s via $\hat{C}$: It can easily be shown that Carter and Rolph's approach in this context is equivalent to choosing:

$$\hat{C} = \hat{C}(0) = \frac{k-3}{Q(0)}, \quad \text{where} \quad Q(0) = Q(\tau^2 = 0). \tag{16}$$

$\hat{C}(0)$ can be seen as a simple generalisation of the $\hat{C}$ in equation (6) for the case where the W_i terms are not constant. $Q(0)$ is known as Cochrane's heterogeneity statistic in meta-analysis and is closely related to the DerSimonian and Laird estimator for τ^2, τ^2_{DL} [20]. For example, when $Q(0) > k - 1$,

$$\frac{Q(0) - (k-1)}{Q(0)} = \frac{\tau^2_{DL}}{\tau^2_{DL} + \bar{\sigma}^2} = I^2$$

where $\bar{\sigma}^2$ is called the 'typical' within study variance and I^2 is a popular measure of 'inconsistency' (heterogeneity) among studies in a meta-analysis [21]. We will refer to the resulting estimator (which utilises $\hat{C}(0)$ and $\hat{\mu}(0)$) as $\check{\mu}_s^0$.

4.4. Incorporating 'Limiting Translation'

It is well known that shrinkage methods can perform poorly with respect to specific parameter components of a larger system, when the magnitude of the specific parameters are among the most extreme. For this reason, Efron and Morris [22] suggest the use of a 'Limiting Translation' (LT) strategy that constrains one shrinkage estimator to be within a certain distance of the MLE. Efron [14] and Johnson [23] suggest a practical choice for this constraint of one unit of the MLE's standard error. The effect of applying LT to a single extreme parameter component of a larger system, say a μ_i from $\mu_1, ..., \mu_k$, is to dramatically reduce the i'th contribution to the overall MSE of equation (4), at the expense of increasing the total value of equation (4) by a small margin.

Although from equation (3), we can see that, at the trial outset, μ_s is not a single parameter but rather a weighted mixture of all k fixed parameter values $\mu_1, ..., \mu_k$, the specific values of those parameters may mean that μ_s is consistently an outlier. For example, this would certainly be the case if one treatment was far more effective than any other because it would monopolise the value of μ_s. We can apply LT to the shrinkage estimator $\check{\mu}_s^0$ by subtracting $\hat{\mu}_s$ from equation (12), constraining the result to be $\leqslant \sqrt{W_s}$ and noting that the definition of $\hat{C}(0)$ in equation (16) becomes

$$\hat{C}(0) = \min\left\{\frac{(k-3)}{Q(0)}, \frac{\sqrt{W_s}}{|\hat{\mu}(0) - \hat{\mu}_s|}\right\}. \tag{17}$$

This estimator will be referred to as $\check{\mu}_s^0(LT)$. LT versions of all other estimators are clearly possible but are not considered here.

4.5. Some implications of using $\check{\mu}_s^0$

When deriving the form of $\check{\mu}_s^0$, there is no inherent mathematical difficulty in assuming an $N(\mu, W_i\tau^2)$ prior distribution (with varying W_is) for the μ_is because the resulting posterior distribution for $\mu_i|\hat{\mu}_i$ remains in the normal family. However, this shrinkage approach does raise certain philosophical questions when applied in the context of a two-stage drop-the-loser trial. The primary issue is that we do not know *a priori* which treatment will be selected. So, assigning the $N(\mu, W_s\tau^2)$ prior to μ_s (for general values of μ and τ^2) is only possible after we have observed the first stage data. This violates the principle of 'Temporal Coherency' [24] that states that the prior must be specified in advance and constant in time. Indeed, this principle is overwhelmingly adhered to by practitioners of Bayesian inference in the interest of maintaining scientific objectivity. A consequence of this temporal violation, which becomes most apparent in Section 5, is that there is no general way to simulate data consistent with the assumptions of $\check{\mu}_s^0$. To understand this, suppose we wanted to generate trial data consistent with the shrinkage estimator $\check{\mu}_s^{\tau^2}$ instead. We simply start by simulating the μ_is from an $N(\mu, \tau^2)$ density given values for μ and τ^2, which can then be used to generate the trial data for stages one and two $(X_1, ..., X_k, Y_s)$. These data can then be used to specify the distributions $\hat{\mu}_i|\mu_i$ and $\mu_i|\hat{\mu}_i$ from equation (9). Clearly, we can not follow an equivalent data-generating procedure when the μ_is come from an $N(\mu, W_i\tau^2)$ prior density because, as previously stated, the prior can only be specified *after* seeing the data. The single exception is when $\tau^2=0$ (implying a degenerate normal prior) in which case the μ_is all take the value μ with probability 1. This is equivalent to assuming a fixed effects model with only one unknown parameter, μ.

Of course, despite these philosophical concerns, we are still free and able to evaluate $\check{\mu}_s^0$ in a simulation study without exactly mimicking the data-generating process it relies upon.

5. Simulation study

We simulate trial data under a two-stage drop-the-loser design in order to quantify the bias and MSE of four new estimators $\left(\hat{\mu}_s^{MPL}, \check{\mu}_s^{\tau^2}, \check{\mu}_s^0 \text{ and } \check{\mu}_s^0(LT)\right)$ for μ_s, alongside the existing estimators $\left(\check{\mu}_s^{CB}, \hat{\mu}_s, \tilde{\mu}_s\right)$. We use the definition of bias and MSE from equation (3), which can be simply and accurately approximated by averaging over all simulations where, in each single case, a treatment T_i out of k is ranked top at the end of stage 1 so that $\mu_s = \mu_i$. We chose four different levels of standard error

associated with the stage 1 and 2 estimates, along with four different scenarios for the six unknown means:

- Scenario *I*. True means $\sim N(0, 1)$;
- Scenario *II*: True means all 0;
- Scenario *III*: One true mean $= 1$, 5 means equal to 0;
- Scenario *IV*: One true mean $= 1.5$, 5 means equal to 0.

The underlying distribution of the treatment parameters is a key factor driving the Bayesian motivation of any shrinkage estimator. Scenario *I* is compatible with the prior assumptions of $\breve{\mu}_s^{\tau^2}$ and $\breve{\mu}_s^{CB}$. Scenario *II* is compatible with the prior assumptions of all shrinkage estimators, despite being non-stochastic, because it is equivalent to scenario *I* with $\tau^2 = 0$. Carreras and Brannath's dominance result for $\breve{\mu}_s^{CB}$ is valid for scenarios *I* and *II* . Scenarios *III* and *IV* are not compatible with any shrinkage estimator. However, all scenarios are compatible with the assumptions of the UMVCUE, in the sense that it maintains its unbiasedness for any constellation of parameter values.

We show the results for the 16 simulation scenarios in Table I. All reported figures are based on 50 000 simulations. For each simulation scenario, we show the bias and $\sqrt{MSE}$ in units of the MLE $\hat{\mu}_s$s naive standard error, $\sqrt{W_s}$, to make comparisons easier. We do not show the bias of $\tilde{\mu}$ because it is always zero, except for sampling error. All of the shrinkage estimators generally outperform the MLE in terms of bias and $\sqrt{MSE}$, the exception being simulation 15, scenario *IV*. Carreras and Brannath [10] show theoretically that the MLE is maximally biased when all treatment means are equal. The results of scenario *II* supports this. $\breve{\mu}_s^{CB}$ and $\breve{\mu}_s^{\tau^2}$s performances are fairly equal. $\breve{\mu}_s^{CB}$ tends to have a smaller bias than $\breve{\mu}_s^{\tau^2}$ but a larger $\sqrt{MSE}$. The performance of $\hat{\mu}_s^{MPL}$ varies considerably; it is the best estimator

Table I. Bias and mean squared error (MSE) of the various estimands over the 16 scenarios of a two-stage drop-the-loser trial with $k = 6$ initial treatments.

σ_1, σ_2 values	Bias (μ_s^*)						$\sqrt{MSE}(\mu_s^*)$						
	$\hat{\mu}_s$	$\breve{\mu}_s^{CB}$	$\breve{\mu}_s^{\tau^2}$	$\breve{\mu}_s^0$	$\breve{\mu}_s^0(LT)$	$\hat{\mu}_s^{MPL}$	$\tilde{\mu}_s$	$\hat{\mu}_s$	$\breve{\mu}_s^{CB}$	$\breve{\mu}_s^{\tau^2}$	$\breve{\mu}_s^0$	$\breve{\mu}_s^0(LT)$	$\hat{\mu}_s^{MPL}$
Scenario I: true means $\sim N(0, 1)$													
1.(1,1)	0.63	0.19	0.22	0.11	0.11	−0.17	1.21	1.12	0.97	0.96	0.95	0.94	0.97
2.(2,1)	0.51	0.18	0.29	0.11	0.11	−0.03	1.08	1.08	0.98	0.97	0.92	0.92	0.91
3.$\left(\frac{1}{2},1\right)$	0.51	0.13	0.14	0.11	0.11	−0.22	1.35	1.08	0.99	0.99	0.98	0.98	1.04
4.$\left(1,\frac{1}{2}\right)$	0.40	0.12	0.17	0.00	0.00	−0.14	1.07	1.05	0.99	0.98	0.98	0.98	1.01
Scenario II: true means all 0													
5.(1,1)	0.89	0.35	0.45	0.35	0.36	0.16	1.27	1.23	0.92	0.87	0.79	0.79	0.65
6.(2,1)	0.57	0.22	0.39	0.22	0.23	0.12	1.09	1.10	0.97	0.95	0.83	0.83	0.78
7.$\left(\frac{1}{2},1\right)$	1.14	0.45	0.48	0.45	0.47	0.17	1.64	1.35	0.86	0.84	0.81	0.82	0.58
8.$\left(1,\frac{1}{2}\right)$	0.57	0.22	0.39	0.23	0.23	0.12	1.08	1.09	0.96	0.94	0.83	0.83	0.77
Scenario III: one true mean $= 1$, $5 = 0$													
9.(1,1)	0.78	0.25	0.32	0.21	0.22	−0.03	1.24	1.19	0.94	0.93	0.88	0.88	0.84
10.(2,1)	0.55	0.20	0.36	0.19	0.19	0.08	1.08	1.09	0.97	0.95	0.85	0.85	0.81
11.$\left(\frac{1}{2},1\right)$	0.59	−0.07	−0.06	−0.10	−0.09	−0.56	1.40	1.14	1.04	1.05	1.05	1.04	1.20
12.$\left(1,\frac{1}{2}\right)$	0.50	0.16	0.28	0.11	0.11	−0.02	1.08	1.08	0.98	0.97	0.93	0.93	0.93
Scenario IV: one true mean $= 1.5$, $5 = 0$													
13.(1,1)	0.64	0.11	0.16	0.05	0.06	−0.24	1.21	1.14	0.98	0.99	0.98	0.98	1.02
14.(2,1)	0.53	0.19	0.33	0.16	0.16	0.04	1.08	1.08	0.97	0.96	0.88	0.88	0.86
15.$\left(\frac{1}{2},1\right)$	0.21	−0.40	−0.39	−0.43	−0.43	−0.94	1.17	1.04	1.16	1.17	1.19	1.18	1.49
16.$\left(1,\frac{1}{2}\right)$	0.40	0.07	0.16	−0.01	−0.01	−0.16	1.07	1.05	0.99	0.99	1.01	1.01	1.04

in terms of $\sqrt{MSE}$ in 8 out of 16 simulations but is sometimes the worst estimator by far (e.g. simulations 11 and 15). Unlike the other shrinkage estimators, it is also negatively biased in general. $\check{\mu}_s^0$ and $\check{\mu}_s^0(LT)$ perform similarly and are consistently the most reliable estimators across the 16 scenarios. It is perhaps surprising that their similarity extends to scenarios *III* and *IV*, where one would suspect the LT strategy would come into play. This implies that the difference between $\hat{\mu}_s$ and $\check{\mu}_s^0$ is still almost always less than $\sqrt{W_s}$.

Figures 1–3 show the results of three further simulation studies. In each case, $\sigma_1 = \sigma_2 = 1$. Figure 1 shows the scaled bias and $\sqrt{MSE}$ of the estimators for a trial with $k = 6$ treatments, 5 of which have true mean 0 and one of which has true mean δ, as δ is varied between 0 and 5. In order to highlight the strength of the selection effect as a function of δ, we also plot the average value of μ_s (labelled as '$E[\mu_s]$'). One can see that the bias of the shrinkage estimators changes sign as δ increases whereas the bias of the MLE decreases from 0.8 to 0 as δ increases. Of the shrinkage estimators, $\check{\mu}_s^0$ and $\check{\mu}_s^0(LT)$ have the smallest bias and MSE (and are indistinguishable) for δ up to 1.8. For $\delta \leqslant 2.2$, the shrinkage estimators dominate the MLE in terms of $\sqrt{MSE}$. $\hat{\mu}_s^{MPL}$ has the smallest bias of all for $\delta \leqslant 1.5$ but, by far, has the largest (negative) bias as δ increases. $\hat{\mu}_s^{MPL}$ also has the smallest $\sqrt{MSE}$ of all estimators for small δ, but as δ increases, its MSE increases dramatically.
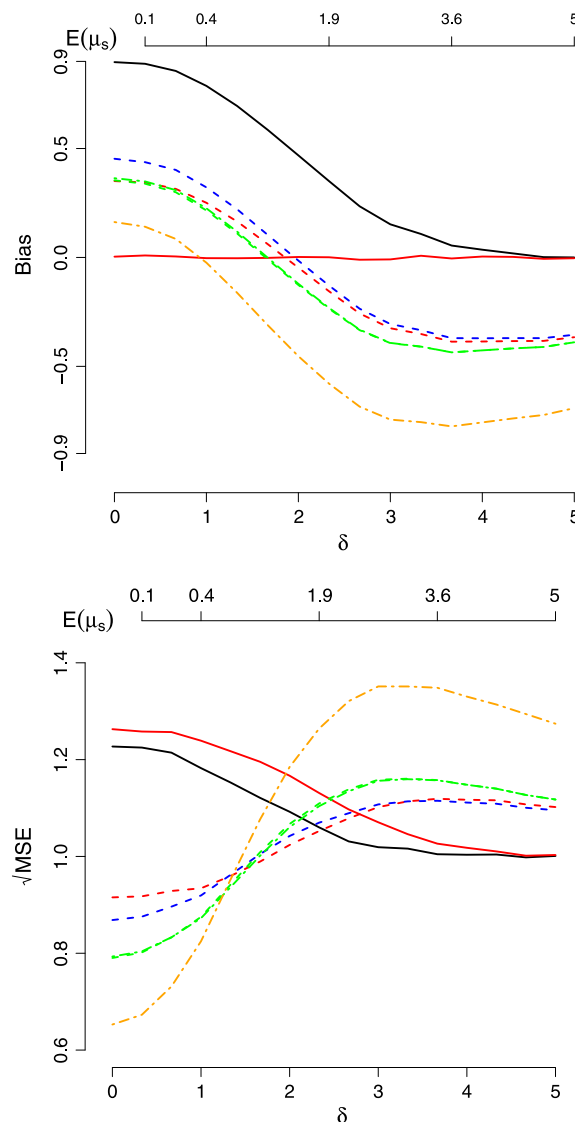
Figure 1. Bias and mean squared error (MSE) of the estimators as a function of δ. Key: maximum likelihood estimate (black), UMVCUE (red), $\check{\mu}_s^{CB}$ (red-dashed), $\check{\mu}_s^{\tau^2}$ (blue-dashed), $\check{\mu}_s^0$ (green dashed), $\check{\mu}_s^0(LT)$ (green dot-dashed) and $\hat{\mu}_s^{MPL}$ (orange dot-dashed).

Figure 2 shows the results of a simulation for $k = 6$ treatments with mean parameters drawn from a $N(0, \tau^2)$ distribution as τ^2 is varied between 0 and 4 (the choice of $\mu = 0$ is clearly unimportant). The bias of all shrinkage estimators decreases towards 0 as τ^2 increases, although this happens most rapidly for $\check{\mu}_s^0$ and $\check{\mu}_s^0(LT)$ so that they are the least biased. For values of τ^2 greater than 1, $\hat{\mu}_s^{MPL}$ is consistently negatively biased. The $\sqrt{MSE}$ of the shrinkage estimators appears to asymptote upwards towards $\sqrt{W_s}$ (towards 1 after scaling) as τ^2 increases, but remain below that of the MLE in this range.

In an effort to separate $\check{\mu}_s^0$ and $\check{\mu}_s^0(LT)$, Figure 3 shows the results of a simulation assuming that the treatment mean parameters are drawn from a $N(0, 1)$ distribution, but the number of treatments, k, is varied between 5 and 20. As k increases, the positive bias exhibited by $\check{\mu}_s^0$ decreases, quickly becoming large and negative. $\check{\mu}_s^{CB}$ and $\check{\mu}_s^{\tau^2}$ do not to suffer in the same way; their biases asymptote towards 0 as k increases. $\check{\mu}_s^0(LT)$ appears to protect $\check{\mu}_s^0$ well from its tendency for negative bias beyond $k = 10$. From the right-hand panel, one can see that the price $\check{\mu}_s^0(LT)$ pays for this bias protection is an increase in $\sqrt{MSE}$. Interestingly, as k increases beyond 15, even the UMVCUE has a smaller $\sqrt{MSE}$ than the MLE.

5.1. Summary of findings

Across all simulations, the performance of $\check{\mu}_s^{\tau^2}$ is most similar to Carreras and Brannath's original estimator $\check{\mu}_s^{CB}$, but the two estimators that performed the best were $\check{\mu}_s^0$ and $\hat{\mu}_s^{MPL}$. However, the reasons for the latter's apparent success are now qualified. Estimation of the heterogeneity parameter τ^2 is
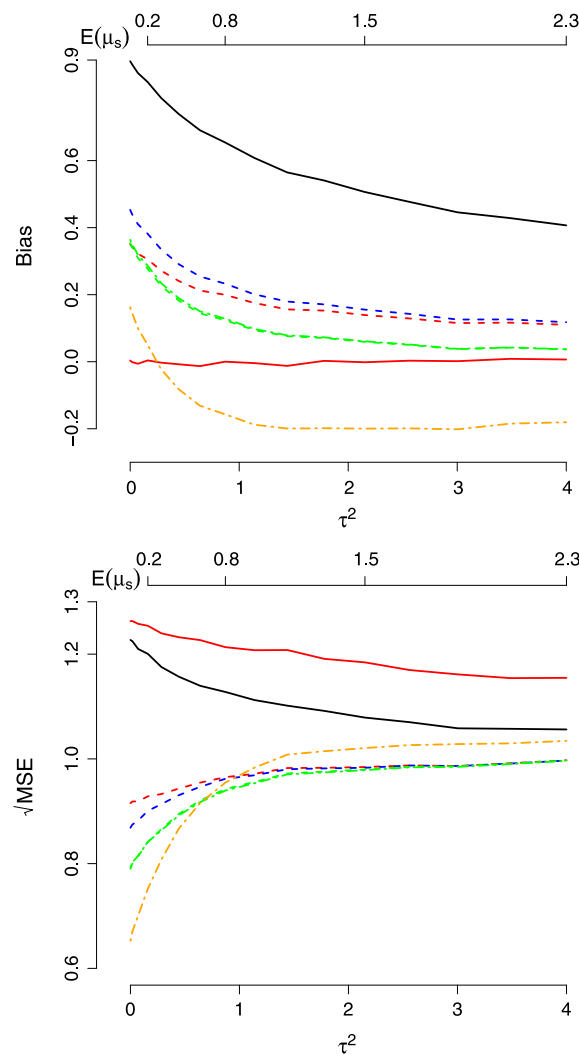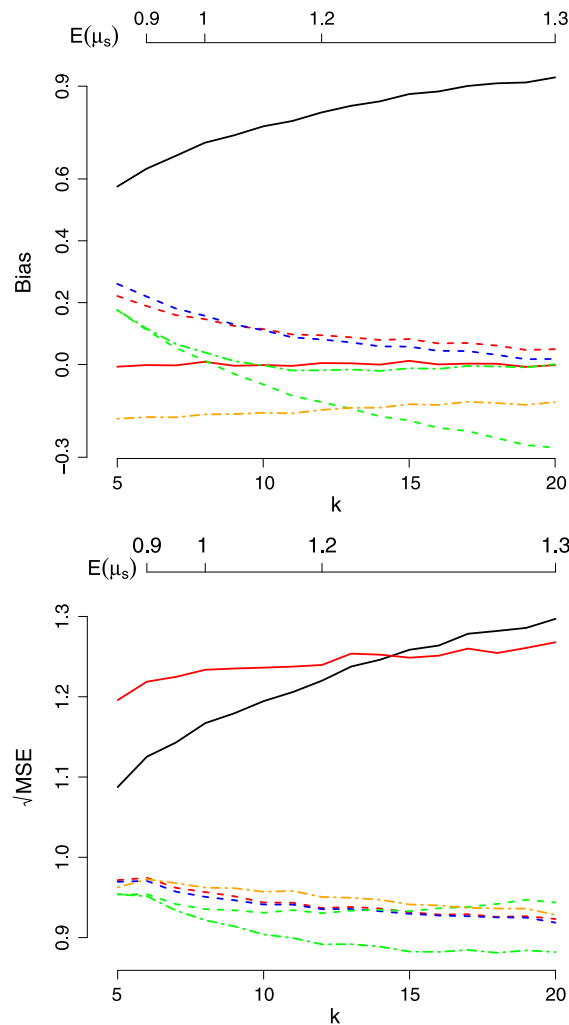

Figure 2. Bias and mean squared error (MSE) of the estimators as a function of τ^2. Key: Same as Figure 1.

Figure 3. Bias and mean squared error (MSE) of the estimators as a function of k. Key: Same as Figure 1.

challenging when its true value is small or the number of treatments is small. When k is small or there is little apparent variation between treatment effect estimates $(\hat{\mu}_1, ..., \hat{\mu}_k)$, the estimate for τ^2 is often likely to be at or close to zero, even when its true value is larger. This fact is well documented in the meta-analysis literature, where quantification of the (between trial) heterogeneity parameter is often not recommended unless the number of studies is sufficiently large (at least 10). Although poor estimation of τ^2 impacts $\check{\mu}_s^{\tau^2}$ and $\hat{\mu}_s^{MPL}$, the latter is most strongly affected because from equation (10) when $\tau^2 = 0$, it reduces via equation (11) to the fixed effects estimate for μ, $\hat{\mu}(0)$. However, when the Paule–Mandel estimate for τ^2 is zero, $\check{\mu}_s^{\tau^2}$ does not shrink to exactly $\hat{\mu}(0)$ because $\hat{C}_s$ in equation (14) is not equal to 0. This explains why $\hat{\mu}_s^{MPL}$ performs so well in Scenario *II*, Table I, and for small values of δ, τ^2, k in Figures 1–3 because in these situations, the true value of μ_s is always close (or equal) to μ. Conversely, it also explains why it performs badly when μ_s is truly very different from μ (e.g. Scenario *IV*, Table I, and large values of δ, τ^2, k in Figures 1–3).

The LT version of $\check{\mu}_s^0$ only helped to improve its performance in simulations when the number of treatments rose above 10, which is unrealistically large for most clinical trials settings. It therefore appears to be an unnecessary extension in this context. However, it could potentially be implemented in a more sophisticated manner than we have here. For example, the width of the protection region around the MLE can be tuned to control its bias and MSE, rather than being fixed at a specific value as we did. Efron and Morris [22] provide the theoretical framework for doing this in a general shrinkage estimation context, but their method would need to be altered before application to drop-the-loser designs, in order to account for the selection mechanism. This is a topic for further research. One could also argue that by shrinking the prior variance for μ_s after selection, $\check{\mu}_s^0$ already contains and in-built form of LT.

6. Discussion

In this paper, we have reviewed the shrinkage estimation strategy of Carreras and Brannath [10] for two-stage drop-the-loser designs. As an extension, we propose that rather than using only the first stage data, the stage two data should also be used to furnish a single shrinkage equation. Although this strategy replaces two formulae with one, evaluation of this single shrinkage formula is much harder because the variance of its k estimated components are no longer equal. By incorporating the methods of Carter and Rolph [16] and Efron and Morris [22], we identified several alternative procedures. The alternative approaches necessitate either explicit estimation of an additional between treatment arm heterogeneity parameter τ^2 (for $\hat{\mu}_s^{MPL}$ and $\check{\mu}_s^{\tau^2}$) or a different (and quite unnatural) prior distribution for the mean treatment effects (for $\check{\mu}_s^0$). The new methods tend to outperform Carreras and Brannath's original estimator, but unfortunately, no equivalent dominance results could be shown.

From looking at the simulations in totality, the estimator that consistently performs well when μ_s is close to and far from the overall mean of all treatments is $\check{\mu}_s^0$. Some may object philosophically to its use in this context because of concerns over Temporal Coherency. However, as Cox [25] states: There may be certain situations where it is perfectly right to modify one's prior beliefs as more data become available. Furthermore, when one's prior uncertainty about a parameter *is* allowed to change over time in a manner proportional to the (increasing) size of the data sample, then the resulting Bayesian inference starts to approximate a classical significance test [26]. This is exactly what occurs in the drop-the-loser context when, at the point of selection, the variance of the prior for μ_s shrinks from $\sigma_1^2 \tau^2$ to $W_s \tau^2$ – for example, by a factor of 2 when $\sigma_1^2 = \sigma_2^2$. It is therefore pertinent to note that the formula for $\check{\mu}_s^0$ can also be arrived at by applying Lindley's original equal variance shrinkage formula (5) to the standardised MLEs, $\hat{\mu}_i / \sqrt{W_i}$, [16] because they are sufficient test statistics for the null hypothesis $\mu = 0$.

We chose to illustrate the different estimation approaches for trials involving five or more treatments. Apart from $\hat{\mu}_s^{MPL}$, all of the shrinkage estimators discussed in this paper are only defined for $k \geqslant 4$, as indicated by the factors of $(k-3)$ they contain. However, we can crudely apply them for the $k = 3$ case by simply replacing these terms with $(k-2)$ instead – as performed by Carreras and Brannath [10]. We repeated the simulations shown in Figures (1) and (2) for $k = 3$ using this crude fix to see how it affected the performance of the various estimators. The results (not shown) were qualitatively very similar.

Throughout this paper, we have attempted to stress the link between shrinkage estimation in the adaptive trial context with that of meta-analysis. We have shown that $\check{\mu}_s^0$, which incorporates the fixed effects estimate $\hat{\mu}(0)$, works well as an estimator for μ_s under drop-the-losers selection. It is therefore interesting to note the following: In meta-analyses that exhibit substantial amounts of between study heterogeneity, the random effects estimate for μ is known to be unreliable when the heterogeneity is thought to be driven by dissemination bias (i.e. selective reporting and publication of extreme findings) [18, 27]. In order to address this, it has been advocated that the fixed effects estimate $\hat{\mu}(0)$ be used as the preferred measure of overall effect instead [28, 29]. Thus, despite the fact that in the adaptive trial setting, the parameter of interest is $\hat{\mu}_s$ and in meta-analysis, it is the overall grand mean μ, when the data are affected by some form of selection, Carter and Rolph's proportional prior approach appears to be an effective solution to both problems.

Bowden and Glimm [30] have extended the idea of a two-stage drop-loser-trial to allow the best performing treatment to be identified over multiple stages. The motivation for adding further stages of selection is that one can markedly increase the probability of selecting the truly best treatment (and subsequently declaring it effective in a confirmatory analysis), whilst keeping trial costs to a minimum. Many other multi-arm multi-stage (MAMS) designs incorporating treatment selection rules have also been proposed with a similar motivation in mind, see for example [31, 32]. Because they ignore the selection process altogether and use all of the data to define a target posterior distribution or shrinkage equation, $\hat{\mu}_s^{MPL}$, $\check{\mu}_s^{\tau^2}$ and $\check{\mu}_s^0$ should be simple to apply in any of these contexts. It is not so obvious to see how Carreras and Brannath's original estimator, $\check{\mu}_s^{CB}$, would generalise to the multi-stage context or if the dominance results that make it attractive in the two-stage case would remain in intact.

A simple, straightforward translation of the shrinkage estimators proposed here to other multi-arm trial designs is only immediate if the k treatment effect estimates are independent before selection. If, as in the MAMS design of Royston *et al.* [33], the treatment effect summarised time-to-event data in the form of a log-hazard ratio, then the effect estimates would be intrinsically correlated across treatment arms because of their shared control group data. Extending shrinkage estimation to account for inter-dependence of this sort is another topic for further research.

Appendix A: Details for the calculation of $\hat{\mu}_s^{MPL}$

Assume without loss of generality that $s = 1$, so that μ_1 is the mean of the best performing treatment. Let $\mathbf{Z} = (X_1, Y_1, X_2, ..., X_k,)$ equal the complete vector of data from the two-stage trial. Under the framework of equation (9), the model induces the unconditional distribution $\mathbf{Z} \sim \mathrm{N}(\mu \mathbf{1}_{k+1}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_2 & \mathbf{0} \\ \mathbf{0} & (\sigma_1^2 + \tau^2)\,\mathbf{I}_{k-1} \end{pmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} \sigma_1^2 + \tau^2 & \tau^2 \\ \tau^2 & \sigma_2^2 + \tau^2 \end{pmatrix}$$

Assuming that μ and τ^2 are known, then the best linear prediction of $\mu_1 | \mu$ is given in equation (10), which is equivalent to

$$\frac{\sigma_1^2 \sigma_1^2 \mu + \sigma_2^2 \tau^2 X_1 + \sigma_1^2 \tau^2 Y_1}{\sigma_1^2 \sigma_2^2 + \sigma_1^2 \tau^2 + \sigma_2^2 \tau^2}$$

The empirical question is how to estimate μ and τ^2 as accurately as possible. If we were to assume that τ^2 is known, then $\boldsymbol{\Sigma}$ becomes a completely known matrix. Then, it is easily seen from differentiating $\boldsymbol{a}' \boldsymbol{\Sigma} \boldsymbol{a} + \lambda \boldsymbol{a}' \mathbf{1}_{k+1}$ with respect to $\mathbf{a}$ and λ that the linear combination $\boldsymbol{a}' \mathbf{Z}$, which minimises $\mathrm{Var}(\boldsymbol{a}' \mathbf{Z}) = \boldsymbol{a}' \boldsymbol{\Sigma} \boldsymbol{a}$, is given by

$$\boldsymbol{a} = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}_{k+1}}{\mathbf{1}_{k+1}' \boldsymbol{\Sigma}^{-1} \mathbf{1}_{k+1}}$$

As this also maximises the multivariate normal likelihood, $\boldsymbol{a}' \mathbf{Z}$ is also equivalent to the MLE for μ. Because Σ is diagonal, its inverse is trivially

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_2^{-1} & \mathbf{0} \\ \mathbf{0} & (\sigma_1^2 + \tau^2)^{-1}\,\mathbf{I}_{k-1} \end{pmatrix}$$

with

$$\boldsymbol{\Sigma}_2^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 + \sigma_1^2 \tau^2 + \sigma_2^2 \tau^2} \begin{pmatrix} \sigma_2^2 + \tau^2 & -\tau^2 \\ -\tau^2 & \sigma_1^2 + \tau^2 \end{pmatrix}$$

Hence, we obtain the estimate for $\hat{\mu}_{\tau^2}$ given in equation (11). We then plug $\hat{\mu}_{\tau^2}$ into the log-likelihood for τ^2 given μ:

$$\begin{aligned} ll(\tau^2) &= -\frac{1}{2}\left(\log Det(\boldsymbol{\Sigma}) + (\mathbf{Z} - \hat{\mu}_{\tau^2})' \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \hat{\mu}_{\tau^2}) + (\mathbf{Z} - \hat{\mu}_{\tau^2})\right) \\ &= \frac{1}{2}\left((k-1)\log\left(\sigma_1^2 + \tau^2\right) + \log\left(\sigma_1^2 \sigma_2^2 + \sigma_1^2 \tau^2 + \sigma_2^2 \tau^2\right)\right. \\ &\quad \left. + (\mathbf{Z} - \hat{\mu}_{\tau^2})' \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \hat{\mu}_{\tau^2}) + (\mathbf{Z} - \hat{\mu}_{\tau^2})\right) \end{aligned}$$

and maximise to obtain an estimate for τ^2

Acknowledgements

References

1. Sampson A, Sill M. Drop-the-losers design: normal case. *Biometrical Journal* 2005; **47**:257–268.
2. Wu S, Wang W, Yang M. Interval estimation for drop-the-losers designs. *Biometrika* 2010; **97**:406–418.
3. US Food and Drug Administration (FDA) 2010. Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics.
4. Cohen A, Sackrowitz H. Two stage conditionally unbiased estimators of the selected mean. *Statistics and Probability Letters* 1989; **8**:273–278.

5. Sill M, Sampson A. Extension of a two-stage conditionally unbiased estimator of the selected population to the bivariate normal case. *Communications in Statistics-Theory and Methods* 2007; **36**:801–813.

6. Bowden J, Glimm E. Unbiased estimation of selected treatment means in two-stage trials. *Biometrical Journal* 2008; **50**:515–527.

7. Kimani P, Stallard N, Todd S. Conditionally unbiased estimation in phase II/III clinical trials with early stopping for futility. *Statistics in Medicine* 2013. Published online DOI: doi/10.1002/sim.5757/pdf.

8. Stallard N, Todd S. Point estimators and confidence regions for sequential trials involving selection. *Journal of Statistical Planning and Inference* 2005; **135**:402–419.

9. Bebu I, Luta G, Dragalin V. Likelihood inference for a two-stage design with treatment selection. *Biometrical Journal* 2010; **52**:811–822.

10. Carreras M, Brannath W. Shrinkage estimation in two-stage adaptive designs with midtrial treatment selection. *Statistics in Medicine* 2013; **32**:1677–90. Published online DOI: doi/10.1002/sim.5463.

11. James W, Stein C. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Univ. California Press, Berkeley, 1961; 361–379.

12. Posch M, Koenig F, Branson M, Brannath W, Dunger-Baldauf C, Bauer P. Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine* 2005; **24**:3697–3714.

13. Lindley D. Discussion of professor Stein's paper "confidence sets for the mean of a multivariate normal distribution". *Journal of the Royal Statistical Society, series B* 1962; **24**:265–296.

14. Efron B. Biased versus unbiased estimation. *Advances in Mathematics* 1975; **16**:259–277.

15. Hwang J. Empirical Bayes estimation for the means of the selected populations. *The Indian Journal of Statistics* 1993; **55**:285–311.

16. Carter G, Rolph J. Emprical Bayes methods applied to estimating fire alarm probabilities. *Journal of the American Statistical Association* 1974; **69**:880–885.

17. Viechtbauer W. Confidence intervals for the amount of heterogeneity in a meta-analysis. *Statistics in Medicine* 2007; **26**:37–52.

18. Bowden J, Tierney J, Copas A, Burdett S. Quantifying, displaying and accounting for heterogeneity in the meta-analysis of rcts using standard and generalised Q statistics. *Medical Research Methodology* 2011; **11**:41.

19. Paule R, Mandel J. Concensus values and weighting factors. *Journal of Research for the National Bureau of Standards* 1982; **87**:377–85.

20. DerSimonian R, Kacker R. Random effect models for meta-analysis of clinical trials: an update. *Contemporary Clinical Trials* 2007; **28**:105–114.

21. Higgins J, Thompson S, Deeks J, Altman D. Measuring inconsistency in meta-analyses. *British Medical Journal* 2003; **327**:557–560.

22. Efron B, Morris C. Limting the risk of Bayes and empirical Bayes estimators-part II: the empirical Bayes case. *Journal of the American Statistical Association* 1972; **67**:130–139.

23. Johnson D. An empirical Bayes approach to analyzing recurring animal surveys. *Ecology* 1989; **70**:945–952.

24. Hacking I. Slightly more realistic personal probability. *Philosophy of Science* 1967; **34**:311–325.

25. Cox D. Foundations of statistical inference: the case for eclecticism. *Australian Journal of Statistics* 1978; **20**:43–59.

26. Cox D, Hinkley D. *Theoretical Statistics*. Chapman and Hall: London, 1974.

27. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology* 1999; **150**:469–475.

28. Henmi M, Copas J. Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in Medicine* 2010; **29**:2969–2983.

29. Bowater R, Escarela G. Heterogeneity and study size in random-effects meta-analysis. *Journal of Applied Statistics* 2013; **40**:2–16.

30. Bowden J, Glimm E. Conditionally unbiased and near unbiased estimation of the selected treatment mean for multi-stage drop-the-losers trials. *Biometrical Journal* 2013. In press.

31. Stallard N, Friede T. A group-sequential design for clinical trials with treatment selection. *Statistics in Medicine* 2008; **27**:6209–6227.

32. Magirr D, Jaki T, Whitehead J. A generalized Dunnett test for multi-arm multi-stage clinical studies with treatment selection. *Biometrika* 2012; **99**:494–501.

33. Royston P, Parmar M, Qian W. Novel designs for multi-arm clinical trials with survival outcomes, with an application in ovarian cancer. *Statistics in Medicine* 2003; **22**:2239–2256.