Adaptive enrichment designs with a continuous biomarker

Nigel Stallard

Warwick Medical School, University of Warwick
Coventry, CV4 7AL, U.K.

n.stallard@warwick.ac.uk

Abstract

A popular design for clinical trials assessing targeted therapies is the two-stage adaptive enrichment design with recruitment in stage two limited to a biomarker-defined subgroup chosen based on data from stage one. The data-dependent selection leads to statistical challenges if data from both stages are used to draw inference on treatment effects in the selected subgroup.

If subgroups considered are nested, as when defined by a continuous biomarker, treatment effect estimates in different subgroups follow the same distribution as estimates in a group-sequential trial. This result is used to obtain tests controlling the familywise type I error rate (FWER) for six simple subgroup selection rules, one of which also controls the FWER for any selection rule. Two approaches are proposed; one based on multivariate normal distributions suitable if the number of possible subgroups, k, is small, and one based on Brownian motion approximations suitable for large k. The methods, applicable in the wide range of settings with asymptotically normal test statistics, are illustrated using survival data from a breast cancer trial.

Keywords

Multiple cut-points; Sequential analysis; Stratified medicine clinical trial design; Subgroup selection.

1 Introduction

An area of much interest in modern medicine, particularly in oncology, is the development of targeted therapies that benefit a specific subset of the patient population. This in turn has led to interest in the design of clinical trials to evaluate effectiveness in settings where the treatment effect may be heterogeneous (Antoniou et al. 2016, Lin et al. 2019).

One such approach is the adaptive enrichment design (Wang et al. 2009, Simon & Simon 2013, Rosenblum et al. 2020, Lin et al. 2021). This is a design for confirmatory phase III randomized clinical trials conducted in two stages. In stage 1, patients recruited from the whole population are randomized to experimental and control treatments. At the end of stage 1 the observed data are used to decide whether to continue to recruit from the entire population or to restrict recruitment on the basis of a pre-specified candidate biomarker, enriching the trial to focus on this chosen subgroup. Further patients are then recruited from the selected population or subgroup in stage 2. Ho et al. (2012) describe such a trial.

The analysis of such trials raises a number of challenges (Simon 2015). At the end of the trial it is desired to test if the experimental treatment is superior to the control in the selected subgroup. If data from stage

1 are included in this test, the use of the same data to select the subgroup and evaluate the effect of the treatment can inflate the type I error rate. Some proposed solutions to this problem in the general setting are to include only stage 2 patients in the final analysis (Renfro et al. 2014) or to include all stage 1 patients, whether or not they are in the selected subgroup, combining these with stage 2 patients in a prespecified manner (Simon & Simon 2013).

In the specific case of a single binary biomarker with prior belief that a larger treatment effect will be associated with a positive biomarker, the decision after stage 1 is to continue with the full population or enrich to include biomarker positive patients only (Wang et al. 2007). Methods for this case have been proposed by Spiessens & Debois (2010), Jenkins et al. (2011), Friede et al. (2012), Stallard et al. (2014) and Rosenblum, Luber, Thompson & Hanley (2016). Lai et al. (2019) and Chiu et al. (2018) consider a more general setting with a number of nested subgroups and selection maximizing the stage 1 likelihood ratio test statistic.

Here we consider the setting of subgroup selection based on a single pre-specified biomarker that is either continuous or has multiple levels with it assumed *a priori* that the effect of the experimental treatment increases with increasing levels of this biomarker. In this case the subgroups considered are nested as assumed by Lai et al. (2019). Selection of a subgroup corresponds to choice of a threshold value for the biomarker, with the selected subgroup comprising all patients with biomarker levels above this threshold, as explained in more detail in Section 2, where six simple methods for subgroup selection are described.

In Section 3 we show how a combination test, as proposed in a general adaptive design setting by Bauer & Köhne (1994), can be used to combine evidence from the two stages of the adaptive enrichment design. The combination testing approach requires construction of a p-value from the stage 1 data allowing for the subgroup selection. For the case of nested subgroups, analysis of data from different subgroups is analagous to that of accumulating data in a sequential trial. Together with an assumption of asymptotic normality, this result is exploited in Section 4 to obtain the distribution of test statistics for the six different selection rules introduced in Section 2, and hence expressions for the calculation of stage 1 p-values in each case. When the number of subgroups considered, k, is small or moderate, expressions based on k-dimensional multivariate normal tail areas can be used, generalizing the approach of Lai et al. (2019) to allow for different selection rules. Alternative expressions are then given for three of the selection rules based on Brownian motion approximations that are suitable for use with large k. One method controls the type I error rate for any selection rule based on the stage 1 data, though may be conservative.

The paper therefore provides methods for analysis of an adaptive enrichment design that controls the type I error allowing for the selection in stage 1, with specific methods to utilise the full type I error for a number of common simple selection rules and a conservative method for any selection rule. Asymptotic results are given that are suitable when the biomarker cutpoint can take many levels, with methods based on multivariate normal distributions available when the number of levels is small and asymptotic results may be less accurate.

Section 5 illustrates the methods through analysis of data from 686 patients from the German Breast Cancer Study (GBCS) (Schmoor et al. 1996). This trial compared survival times for women who were or were not treated with hormone therapy. Baseline characteristics include the number of progesterone receptors. We demonstrate the retrospective application of an adaptive enrichment design to investigate identification of a subgroup for whom hormone therapy is effective. Also presented are simulations based on resampled GBCS study data, demonstrating accurate type I error rate control. The paper concludes with discussion in Section 6. R code to implement the methods proposed is available from the author.

2 Subgroup selection methods

Consider a two-stage adaptive enrichment design. In stage 1 patients are recruited and randomized between an experimental treatment, T_1, and a control treatment, T_0, with patient i having biomarker value x_i. Following treatment, we observe some response Y_i. We assume higher values of x_i are believed to be associated with larger treatment effects. On the basis of stage 1 data, we will obtain a threshold, λ in some set Λ. The choice of Λ and some common methods for choosing λ are described below. In stage 2 of the trial, recruitment is restricted to patients with biomarker values exceeding λ, with patient i having biomarker value $x_i > \lambda$. At the end of stage 2 we wish to use data from both stages to assess whether T_1 is superior to T_0 for patients with biomarker values exceeding λ.

Let θ_λ denote the average effect of T_1 relative to T_0 for patients with $x_i > \lambda$. Following selection of the population of patients with $x > \lambda$, it is desired to test the null hypothesis $H_\lambda : \theta_\lambda \leq 0$. As selection of λ is data dependent, we wish to control the familywise error rate (FWER) in the strong sense for the family of hypotheses $\mathcal{H} = \{H_\lambda, \lambda \in \Lambda\}$.

At the end of stage 1, for a given choice of λ, defining the subgroup chosen, a test of H_λ may be based on comparison of responses for patients receiving T_1 with those who receive T_0 amongst patients with $x_i > \lambda$. Without loss of generality, suppose that stage 1 patients are arranged in decreasing order of biomarker value. If λ_1 and λ_2 are such that $x_i > \lambda_1 > \lambda_2 > x_{i+1}$ for some i, tests of H_{λ_1} and H_{λ_2} will give identical results. Denoting by k the size of the set Λ, and assuming $\lambda_1 > \cdots > \lambda_k$, we can therefore choose $\lambda_1, \ldots, \lambda_k$ such that for $j = 1, \ldots, k-1$, $\lambda_j > x_i > \lambda_{j+1}$ for some i, and $\lambda_k < x_i$ for all i.

Let $n_j = |\{x_i : x_i > \lambda_j\}| (j = 1, \ldots, k)$. The set of possible thresholds, Λ, may be prespecified, or chosen to be, for example, the deciles of the observed biomarker levels from stage 1, or may be chosen to come from a sufficiently finely spaced set to ensure that $n_j - n_{j-1} = 1 (j = 2, \ldots, k)$, though in order to avoid selection of a subgroup based on a very small sample, it is common to set λ_1 so as to have n_1 sufficiently large.

Denoting θ_{λ_j} by $\theta_j, (j = 1, \ldots, k)$, selection of λ based on stage 1 data is equivalent to selecting a value of j. In order to emphasize that this selection is data-dependent, we will write J for this random variable. A number of methods have been proposed for subgroup selection. Six relatively simple rules are described below and considered in detail in the remainder of this paper. More complex rules have also been suggested (see Discussion).

Selection rule 1 maximizes the test statistic in the subset. Let $\hat{\theta}_j$ be an estimate of θ_j based on data from patients $1, \ldots, n_j$, with estimated variance I_j^{-1}, and let $Z_j = \hat{\theta}_j I_j^{1/2}$ be a Wald statistic for testing H_0 based on these data. Several authors, including Lai et al. (2019), propose selecting j to correspond to the largest Z_j. This will be denoted by $J^{(1)} = \arg\max_{j=1,\ldots k}\{Z_j\}$. This is equivalent to selecting j to minimize the p-value based on Z_j.

Selection rule 2 maximizes the treatment effect estimate, taking j to be $J^{(2)} = \arg\max_{j=1,\ldots k}\{\hat{\theta}_j\}$.

Selection rule 3 maximizes the impact, the product of effect size and subgroup prevalence, as proposed by Zhao & LeBlanc (2020). Since I_j is approximately proportional to n_j, this is approximately equivalent to setting j to be $J^{(3)} = \arg\max_{j=1,\ldots k}\{S_j\}$ with $S_j = \hat{\theta}_j I_j$.

Selection rule 4 maximizes the test statistic for the interaction term giving the difference between treatment effects in the subgroup and its complement as proposed by Su et al. (2013) and Renfro et al. (2014). The standardized test statistic for this interaction when the threshold value for the biomarker is λ_j will be denoted $Z_j^{(int)}$ so that a selection rule is to choose j equal to $J^{(4)} = \arg\max_{j=1,\ldots k-1}\{Z_j^{(int)}\}$, where in this case the range of j over which the arg max is taken is $j = 1, \ldots, k-1$ since $j = k$ corresponds to the subgroup being the entire population so that its complement is empty. Let $\bar{\theta}_{\lambda_j}$ denote the treatment effect for patients with $x \leq \lambda_j$, and let $\hat{\bar{\theta}}_j$ be an estimate of $\bar{\theta}_{\lambda_j}$ based on data from patients $n_j + 1, \ldots, n$ in stage 1. The test statistic $Z_j^{(int)}$ is then approximately $Z_j^{(int)} = (\hat{\theta}_j - \hat{\bar{\theta}}_j)(\mathrm{var}(\hat{\theta}_j - \hat{\bar{\theta}}_j))^{-1/2}$.

Selection rule 5 maximizes the interaction effect estimate, the difference between the treatment effects in

the subgroup and its complement, so that j is $J^{(5)} = \arg\max_{j=1,\ldots k}\{\hat{\theta}_j - \hat{\bar{\theta}}_j\}$.

Selection rule 6 maximizes the interaction effect estimate weighted by the size of the selected subgroup, taking j to be $J^{(6)} = \arg\max_{j=1,\ldots k}\{I_j(\hat{\theta}_j - \hat{\bar{\theta}}_j)\}$.

3 Controlling the type I error when combining data from stages 1 and 2

Following selection of j, we wish to test $H_j : \theta_j \leq 0$ controlling the FWER over $\mathscr{F} = \{H_1, \ldots, H_k\}$. FWER is controlled in the strong sense if H_j is rejected when $H_{\mathscr{S}} = \bigcap_{i\in\mathscr{S}} H_i$ is rejected at nominal level for all $\mathscr{S} \subseteq \{1, \ldots, k\}$ with $\mathscr{S} \ni j$ (Markus et al. 1976). For monotonic treatment effects $\theta_i \geq \theta_j$, $i \leq j$ and $H_1 \subseteq \cdots \subseteq H_k$. Thus FWER is strongly controlled by rejecting H_j whenever H_i is rejected at nominal level for all $i \leq j$.

Let $J^{(r)}(i)$ denote the value of j from $\{i, \ldots, k\}$ selected using rule $r (r = 1, \ldots, 6)$, so that, for example, $J^{(1)}(i) = \arg\max_{j=i,\ldots k}\{Z_j\}$. For given selection rule, omitting the superscript (r) for notational convenience, as $H_i = \cap_{i'=i}^{k} H_{i'}$, a level α test of H_i can be based on $Z_{J(i)}$, with a p-value that controls the type I error at the nominal level obtained from the distribution function of $Z_{J(i)}$ under H_i, which can be calculated under the point null $H_i^* : \theta_i = \cdots = \theta_k = 0$, since this maximises the p-value over H_i. Denoting this p-value by $p_i(Z_{J(i)})$, a p-value for the test of H_J that strongly controls the FWER is $\max_{i=1,\ldots,J}\{p_i(Z_{J(i)})\}$.

Bauer & Köhne (1994) propose a method for testing a hypothesis H_i based on independent (one-sided) p-values, p_1 and p_2, from the two stages of a clinical trial with stage 2 adapted based on stage 1 data. Given pre-specified w_1, w_2 with $w_1^2 + w_2^2 = 1$, let

$$p_c = 1 - \Phi(w_1\Phi^{-1}(1 - p_1) + w_2\Phi^{-1}(1 - p_2)), \tag{1}$$

with Φ the standard normal distribution function, then $p_c \sim U[0, 1]$ under H_i, so is a p-value for H_i (Lehmacher & Wassmer 1999; Brannath et al. 2002).

Following stage 1 selection of subgroup J, let p_1 be $\max_{i=1,\ldots,J} p_i(Z_{J(i)})$, the p-value for a test of H_J based on stage 1 data allowing for subgroup selection and p_2 be a p-value from any valid test of H_J using stage 2 data, that is with $p_2 \sim U[0, 1]$ under H_J. Since p_1 and p_2 are independent the combination test will control the FWER in the strong sense.

4 Construction of stage 1 p-values allowing for subgroup selection

4.1 Notation and distributional assumptions

For $\mathbf{v}$ a vector of length l, let $\mathbf{v}_{i:l} = (v_i, \ldots, v_l)'$ and $\boldsymbol{\Sigma}(\mathbf{v})$, $\boldsymbol{\Sigma}_S(\mathbf{v})$ and $\boldsymbol{\Sigma}_Z(\mathbf{v})$ be $l \times l$ matrices with $\boldsymbol{\Sigma}(\mathbf{v})_{i,j} = v_{\max\{i,j\}}^{-1}$, $\boldsymbol{\Sigma}_S(\mathbf{v})_{i,j} = v_{\min\{i,j\}}$ and $\boldsymbol{\Sigma}_Z(\mathbf{v})_{i,j} = v_{\min\{i,j\}}^{1/2} v_{\max\{i,j\}}^{-1/2}$ $(i = 1, \ldots, l, j = 1, \ldots, l)$. Let $\mathbf{B}(\mathbf{v})$, $\mathbf{C}(\mathbf{v})$ and $\mathbf{D}(\mathbf{v})$ be $(l-1) \times l$ matrices with $B(\mathbf{v})_{j,j} = (v_j v_l)^{1/2}(v_l - v_j)^{-1/2}$, $B(\mathbf{v})_{j,l} = -B(\mathbf{v})_{j,j}$, $C(\mathbf{v})_{j,j} = v_l(v_l - v_j)^{-1}$, $C(\mathbf{v})_{j,l} = -C(\mathbf{v})_{j,j}$, $D(\mathbf{v})_{j,j} = v_j v_l(v_l - v_j)^{-1}$, $D(\mathbf{v})_{j,l} = -D(\mathbf{v})_{j,j}$ $(j = 1, \ldots, l-1)$ and other elements equal to zero. Let $\mathbf{A}_l^{(j)}(j = 1, \ldots, l)$ be a $l \times l$ matrix with $A_{i,i}^{(j)} = 1(i = 1, \ldots, l)$, $A_{i,j}^{(j)} = -1(i = 1, \ldots, l, i \neq j)$, and other elements equal to 0, with $\mathbf{A}_{i:l}^{(j)}$ the submatrix obtained from $\mathbf{A}^{(j)}$ by deleting columns and rows $1, \ldots, (i-1)$. Let $\mathbf{B}^{(j)}(\mathbf{v}_{i:l}) = \mathbf{A}_{i:l-1}^{(j)}\mathbf{B}(\mathbf{v}_{i:l})$ and $\mathbf{C}^{(j)}(\mathbf{v}_{i:l}) = \mathbf{A}_{i:l-1}^{(j)}\mathbf{C}(\mathbf{v}_{i:l})$.

Bonetti & Gelber (2004) give the distribution of treatment effect estimates for nested subgroups. This is identical to that of estimates in a group-sequential trial (Spiessens & Debois 2010, Rosenblum, Qian, Du, Qiu & Fisher 2016). If $\mathbf{0}, \mathbf{I}, \mathbf{n}, \hat{\theta}, \mathbf{S}$ and $\mathbf{Z}$ are vectors with j'th elements equal to $0, I_j, n_j, \hat{\theta}_j, S_j$ and

Z_j respectively, in many settings, including normal, Bernoulli and time-to-event endpoints, asymptotically under H_i^* $(i = 1, \ldots, k)$,

$$\hat{\theta}_{i:k} \sim N\left(\mathbf{0}, \mathbf{\Sigma}(\mathbf{I}_{i:k})\right), \quad \mathbf{S}_{i:k} \sim N\left(\mathbf{0}, \mathbf{\Sigma}_S(\mathbf{I}_{i:k})\right) \text{ and } \mathbf{Z}_{i:k} \sim N\left(\mathbf{0}, \mathbf{\Sigma}_Z(\mathbf{I}_{i:k})\right) \tag{2}$$

with $I_j \propto n_j$ so that $\mathbf{\Sigma}_Z(\mathbf{n})$ approximates $\mathbf{\Sigma}_Z(\mathbf{I})$ and $\mathbf{\Sigma}(\mathbf{n})$ and $\mathbf{\Sigma}_S(\mathbf{n})$ are approximately proportional to $\mathbf{\Sigma}(\mathbf{I})$ and $\mathbf{\Sigma}_S(\mathbf{I})$ respectively (Jennison & Turnbull 1997).

In this section, for given $\mathbf{n} = (n_1, \ldots, n_k)'$, we obtain approximations, $F_i^{(r)}$, for the distribution functions of $Z_{J^{(r)}(i)} (i = 1, \ldots, k, r = 1, \ldots, 6)$, that is corresponding to the selection rules introduced above, suitable for small to moderate k and further approximations, $G_i^{(r)}$, for the distribution functions of $Z_{J^{(r)}(i)} (i = 1, \ldots, k, r = 1, 2, 3)$ suitable for large k.

Given observed $Z_{J^{(r)}} = z$, a stage 1 p-value for the test of $H_i (i = 1, \ldots, k)$ is given by $1 - F_i^{(r)}(z)$ or $1 - G_i^{(r)}(z)$. A p-value for a test of $H_{J^{(r)}}$ that strongly controls the FWER is then given by $\max_{i=1,\ldots,J^{(r)}}\{1 - F_i^{(r)}(z)\}$ or $\max_{i=1,\ldots,J^{(r)}}\{1 - G_i^{(r)}(z)\}$. Since for any selection rule depending on data from patients recruited in stage 1 we must have $Z_{J(i)}$ no larger than $Z_{J^{(1)}(i)} = \max_{j \geq i}\{Z_j\}$, $1 - F_i^{(r)}(z)$ or $1 - G_i^{(r)}(z)$ also provide a conservative p-value for that test of H_i using a test statistic based on any such selection rule. In particular, for large k, $\max_{i=1,\ldots,J^{(r)}}\{1 - G_i^{(1)}(z)\}$ provides a conservative p-value for testing $H_{J^{(r)}}$ with $r = 4, 5, 6$.

4.2 Correcting for subgroup selection in stage 1 - multivariate normal approach

For small k, it is feasible to obtain the distribution of $Z_{J^{(r)}(i)} (r = 1, \ldots, 6)$ from (2).

Using selection rule 1, to test H_i we select $J^{(1)}(i) = \arg\max_{j=i,\ldots,k}\{Z_j\}$ and base a test on $Z_{J^{(1)}(i)} = \max_{j=i,\ldots,k}\{Z_j\}$, with distribution function under H_i^*

$$Pr(Z_j \leq c, j = i, \ldots, k) = \Phi^{(k-i+1)}(c\mathbf{1}, \mathbf{0}, \mathbf{\Sigma}_Z(\mathbf{I}_{i:k})) \tag{3}$$

or approximately

$$F_i^{(1)}(c) = \Phi^{(k-i+1)}(c\mathbf{1}, \mathbf{0}, \mathbf{\Sigma}_Z(\mathbf{n}_{i:k})), \tag{4}$$

where $\Phi^{(k-i+1)}(\mathbf{x}, \mu, \mathbf{\Sigma})$ is the distribution function at $\mathbf{x}$ of a $(k - i + 1)$-dimensional normal distribution with mean μ and variance $\mathbf{\Sigma}$, and $\mathbf{1}$ is a vector with all elements equal to 1.

Using selection rule 2, as $J^{(2)}(i) = \arg\max_{j=i,\ldots,k}\{\hat{\theta}_j\}$, $Z_{J^{(2)}(i)}$ has distribution function

$$\sum_{j=i}^{k} Pr(Z_j \leq c, J^{(2)}(i) = j) = \sum_{j=i}^{k} Pr\left(\hat{\theta}_j \leq cI_j^{-1/2}, J^{(2)}(i) = j\right). \tag{5}$$

Let $\hat{\theta}_l^{(j)} = \hat{\theta}_l - \hat{\theta}_j$ for $l \neq j$ and $\hat{\theta}_j^{(j)} = \hat{\theta}_j$, with $\hat{\theta}^{(j)} = (\hat{\theta}_1^{(j)}, \ldots, \hat{\theta}_k^{(j)})'$. Then $Pr(\hat{\theta}_j \leq cI_j^{-1/2}, J^{(2)}(i) = j)$ is $Pr(\hat{\theta}_j^{(j)} \leq cI_j^{-1/2}, \hat{\theta}_l^{(j)} \leq 0, l = i, \ldots, k, l \neq j)$. Since $\hat{\theta}_{i:k}^{(j)} = \mathbf{A}_{i:k}^{(j)}\hat{\theta}_{i:k}$ it follows from (2) that $\hat{\theta}_{i:k}^{(j)} \sim N(\mathbf{0}, \mathbf{A}_{i:k}^{(j)}\mathbf{\Sigma}(\mathbf{I}_{i:k})\mathbf{A}_{i:k}^{(j)'})$. Thus, since $\mathbf{A}^{(j)}\mathbf{1}$ has element j equal to 1 and all other elements equal to 0,

$$Pr(\hat{\theta}_j \leq cI_j^{-1/2}, J^{(2)}(i) = j) = \Phi^{(k-i+1)}\left(cI_j^{-1/2}\mathbf{A}_{i:k}^{(j)}\mathbf{1}, \mathbf{0}, \mathbf{A}_{i:k}^{(j)}\mathbf{\Sigma}(\mathbf{I}_{i:k})\mathbf{A}_{i:k}^{(j)^T}\right)$$

and (5) becomes $\sum_{j=i}^{k} \Phi^{(k-i+1)}\left(cI_j^{-1/2}\mathbf{A}_{i:k}^{(j)}\mathbf{1}, \mathbf{0}, \mathbf{A}_{i:k}^{(j)}\mathbf{\Sigma}(\mathbf{I}_{i:k})\mathbf{A}_{i:k}^{(j)'}\right)$. Recalling that asymptotically I_j is proportional to n_j, this can be approximated by

$$F_i^{(2)}(c) = \sum_{j=i}^{k} \Phi^{(k-i+1)}\left(cn_j^{-1/2}\mathbf{A}_{i:k}^{(j)}\mathbf{1}, \mathbf{0}, \mathbf{A}_{i:k}^{(j)}\mathbf{\Sigma}(\mathbf{n}_{i:k})\mathbf{A}_{i:k}^{(j)'}\right). \tag{6}$$

Using selection rule 3, as $J^{(3)}(i) = \arg\max_{j=i,\ldots,k}\{S_j\}$, $Z_{J^{(3)}(i)}$ has distribution function

$$\sum_{j=i}^{k} Pr(Z_j \leq c, J^{(3)}(i) = j) = \sum_{j=1}^{k} Pr\left(S_j \leq cI_j^{1/2}, J^{(3)}(i) = j\right). \tag{7}$$

If $\mathbf{S}^{(j)} = \mathbf{A}^{(j)}\mathbf{S}$, $Pr(Z_j \leq c, J^{(3)}(i) = j)$ is $Pr(S_j^{(j)} \leq cI_j^{1/2}, S_l^{(j)} \leq 0, l = i, \ldots, k, l \neq j) =$ $\Phi^{(k-i+1)}\left(cI_j^{1/2}\mathbf{A}_{i:k}^{(j)}\mathbf{1}, \mathbf{0}, \mathbf{A}_{i:k}^{(j)}\mathbf{\Sigma}_S(\mathbf{I}_{i:k})\mathbf{A}_{i:k}^{(j)'}\right)$ and $Z_{J^{(3)}(i)}$ has approximate distribution function

$$F_i^{(3)}(c) = \sum_{j=i}^{k} \Phi^{(k-i+1)}\left(cI_j^{1/2}\mathbf{A}_{i:k}^{(j)}\mathbf{1}, \mathbf{0}, \mathbf{A}_{i:k}^{(j)}\mathbf{\Sigma}_S(\mathbf{I}_{i:k})\mathbf{A}_{i:k}^{(j)'}\right). \tag{8}$$

Using selection rule 4, we select $J^{(4)}(i) = \arg\max_{j=i,\ldots,k-1} Z_j^{(int)}$. Since I_j is approximately proportional to n_j, we have $(\text{var}(\hat{\hat{\theta}}_j))^{-1}$ approximately $\bar{I}_j = I_k - I_j$, and Z_j^{int} approximately

$$(\hat{\theta}_j - \hat{\hat{\theta}}_j)(I_j^{-1} + \bar{I}_j^{-1})^{-1/2} = \left(\hat{\theta}_j - \hat{\hat{\theta}}_j\right)(I_j\bar{I}_j/I_k)^{1/2}. \tag{9}$$

As the treatment effect estimate from the whole sample is approximately the weighted average of the estimates from a subgroup and its complement, $\hat{\theta}_k \approx (I_j\hat{\theta}_j + \bar{I}_j\hat{\hat{\theta}}_j)/I_k$ and

$$\hat{\hat{\theta}}_j \approx (I_k\hat{\theta}_k - I_j\hat{\theta}_j)/\bar{I}_j. \tag{10}$$

We will thus take $J^{(4)}(i) = \arg\max_{j=i,\ldots,k-1}\{\tilde{Z}_j\}$ with

$$\tilde{Z}_j = \left(\hat{\theta}_j - (I_k\hat{\theta}_k - I_j\hat{\theta}_j)/\bar{I}_j\right)\left(I_j\bar{I}_j/I_k\right)^{-1/2} = (\hat{\theta}_j - \hat{\theta}_k)\left(I_kI_j/\bar{I}_j\right)^{-1/2}. \tag{11}$$

The distribution function of $Pr(Z_{J^{(4)}(i)} \leq c)$ can be written as $\sum_{j=i}^{k-1}\int Pr(Z_j \leq c, J^{(4)}(i) = j \mid \hat{\theta}_k = q)dF_{\hat{\theta}_k}(q)$, where $F_{\hat{\theta}_k}(q) = \Phi(qI_k^{1/2})$ is the distribution function of $\hat{\theta}_k$.

From (11), and recalling that $Z_j = \hat{\theta}_jI_j^{1/2}$, we have $Z_j = \tilde{Z}_j\bar{I}_j^{1/2}I_k^{-1/2} + \hat{\theta}_kI_j^{1/2}$ so that $Z_j \leq c$ if and only if $\tilde{Z}_j \leq c(I_k/\bar{I}_j)^{1/2} - \hat{\theta}_k(I_jI_k/\bar{I}_j)^{1/2}$. Thus, setting $\tilde{Z}_i^{(j)} = \tilde{Z}_i - \tilde{Z}_j, i \neq j$ and $\tilde{Z}_j^{(j)} = \tilde{Z}_j$, $Z_{J^{(4)}(i)}$ has distribution function

$$\sum_{j=i}^{k-1}\int Pr\left(\tilde{Z}_j^{(j)} \leq c(I_k/\bar{I}_j)^{1/2} - q(I_jI_k/\bar{I}_j)^{1/2}, \tilde{Z}_l^{(j)} \leq 0, l = i, \ldots, k, l \neq j\right) d\Phi(qI_k^{1/2}).$$

Setting $\tilde{\mathbf{Z}} = (\tilde{Z}_1, \ldots, \tilde{Z}_{k-1})'$ and $\tilde{\mathbf{Z}}^{(j)} = (\tilde{Z}_1^{(j)}, \ldots, \tilde{Z}_{k-1}^{(j)})'$, we have $\tilde{\mathbf{Z}}^{(j)} = \mathbf{A}_{1:k-1}^{(j)}\tilde{\mathbf{Z}}$ and $\tilde{\mathbf{Z}} = \mathbf{B}(\mathbf{I})\hat{\theta}$. The distribution function of $Z_{J^{(4)}(i)}$ is thus

$$\sum_{j=i}^{k-1}\int \Phi^{(k-i)}\left(\left(c(I_k/\bar{I}_j)^{1/2} - q(I_jI_k/\bar{I}_j)^{1/2}\right)\mathbf{A}_{i:k-1}^{(j)}\mathbf{1}, \mathbf{0}, \mathbf{B}^{(j)}(\mathbf{I}_{i:k})\mathbf{\Sigma}(\mathbf{I}_{i:k})\mathbf{B}^{(j)}(\mathbf{I}_{i:k})'\right) d\Phi(qI_k^{1/2}).$$

Lemma 1 in Appendix 1 in the Supplementary material shows that this is

$$\sum_{j=i}^{k-1}\Phi^{(k-i)}\left(c\left(\frac{I_k}{\bar{I}_j}\right)^{1/2}\mathbf{A}_{i:k-1}^{(j)}\mathbf{1}, \mathbf{0}, \mathbf{B}^{(j)}(\mathbf{I}_{i:k})\mathbf{\Sigma}(\mathbf{I}_{i:k})\mathbf{B}^{(j)}(\mathbf{I}_{i:k})' + \left(\frac{I_j}{\bar{I}_j}\right)\mathbf{A}_{i:k-1}^{(j)}\mathbf{1}\mathbf{1}'\mathbf{A}_{i:k-1}^{(j)'}\right),$$

noting that $\mathbf{A}^{(j)}_{i:k-1}\mathbf{1}\mathbf{1}'\mathbf{A}^{(j)}_{i:k-1}{}'$ is a $(k-i)\times(k-i)$ matrix with element $(j+1-i, j+1-i)$ equal to 1 and all other elements equal to zero. Given that I_j is approximately proportional to n_j, with $\bar{n}_j = n_k - n_j$, this is approximately

$$F_i^{(4)}(c) = \sum_{j=i}^{k-1} \Phi^{(k-i)}\Big(c\,(n_k/\bar{n}_j)^{1/2}\,\mathbf{A}^{(j)}_{i:k-1}\mathbf{1}, \mathbf{0}, \mathbf{B}^{(j)}(\mathbf{n}_{i:k})\boldsymbol{\Sigma}(\mathbf{n}_{i:k})\mathbf{B}^{(j)}(\mathbf{n}_{i:k})' + n_j\bar{n}_j^{-1}\mathbf{A}^{(j)}_{i:k-1}\mathbf{1}\mathbf{1}'\mathbf{A}^{(j)}_{i:k-1}{}'\Big). \tag{12}$$

Using selection rule 5, we select $J^{(5)}(i) = \arg\max_{j=i,\ldots,k-1}\{\hat{\theta}_j - \hat{\bar{\theta}}_j\}$. Using (10), this may be approximated by the rule that selects j corresponding to the largest value of $\hat{\delta}_j = \left(\hat{\theta}_j - \hat{\theta}_k\right) I_k/\bar{I}_j$. Since $\hat{\theta}_j = \hat{\delta}_j\bar{I}_j/I_k + \hat{\theta}_k$ and $Z_j = \hat{\theta}_j I_j^{1/2}$, $Z_j \le c$ if and only if $\hat{\delta}_j \le (c - \hat{\theta}_k I_j^{1/2})I_k/(\bar{I}_j I_j^{1/2})$ and the distribution function of $Z_{J^{(5)}(i)}$ is

$$\sum_{j=i}^{k-1}\int Pr\left(\hat{\delta}_j^{(j)} \le \frac{(c - qI_j^{1/2})I_k}{\bar{I}_j I_j^{1/2}}, \hat{\delta}_l^{(j)} \le 0, l = i,\ldots,k, l\ne j\right) d\Phi(qI_k^{1/2})$$

where $\hat{\delta}^{(j)}$ is the vector with $\hat{\delta}_i^{(j)} = \hat{\delta}_i - \hat{\delta}_j, i \ne j$ and $\hat{\delta}_j^{(j)} = \hat{\delta}_j$ so that $\hat{\delta}^{(j)} = \mathbf{C}^{(j)}(\mathbf{I})\hat{\theta}$. The distribution function of $Z_{J^{(5)}(i)}$ is thus

$$\sum_{j=i}^{k-1}\int \Phi^{(k-i)}\left(\left(\frac{(c - qI_j^{1/2})I_k}{\bar{I}_j I_j^{1/2}}\right)\mathbf{A}^{(j)}_{i:k-1}\mathbf{1}, \mathbf{0}, \mathbf{C}^{(j)}(\mathbf{I}_{i:k})\boldsymbol{\Sigma}(\mathbf{I}_{i:k})\mathbf{C}^{(j)}(\mathbf{I}_{i:k})'\right) d\Phi(qI_k^{1/2})$$

which, using the result in Lemma 1 and replacing I_j by n_j, is approximately

$$F_i^{(5)}(c) = \sum_{j=i}^{k-1} \Phi^{(k-i)}\Big(cn_k\bar{n}_j^{-1}n_j^{-1/2}\mathbf{A}^{(j)}_{i:k-1}\mathbf{1}, \mathbf{0}, \mathbf{C}^{(j)}(\mathbf{n}_{i:k})\boldsymbol{\Sigma}(\mathbf{n}_{i:k})\mathbf{C}^{(j)}(\mathbf{n}_{i:k})' + n_k\bar{n}_j^{-2}\mathbf{A}^{(j)}_{i:k-1}\mathbf{1}\mathbf{1}'\mathbf{A}^{(j)}_{i:k-1}{}'\Big). \tag{13}$$

Using selection rule 6, we select $J^{(6)}(i) = \arg\max_{j=i,\ldots,k-1}\{I_j(\hat{\theta}_j - \hat{\bar{\theta}}_j)\}$. From (10), this is approximated by selecting j corresponding to the largest value of $\hat{\epsilon}_j = \left(\hat{\theta}_j - \hat{\theta}_k\right) I_j I_k/\bar{I}_j$. As $Z_j \le c$ whenever $\hat{\epsilon}_j \le (c - \hat{\theta}_k I_j^{1/2})I_k I_j^{1/2}/(\bar{I}_j)$, $Z_{J^{(6)}(i)}$ has distribution function

$$\sum_{j=i}^{k-1}\int Pr\left(\hat{\epsilon}_j^{(j)} \le \frac{(c - qI_j^{1/2})I_j^{1/2}I_k}{\bar{I}_j}, \hat{\epsilon}_l^{(j)} \le 0, l = i,\ldots,k, l\ne j\right) d\Phi(qI_k^{1/2})$$

where $\hat{\epsilon}^{(j)}$ is the vector with $\hat{\epsilon}_i^{(j)} = \hat{\epsilon}_i - \hat{\epsilon}_j, i \ne j$ and $\hat{\epsilon}_j^{(j)} = \hat{\epsilon}_j$ so that $\hat{\epsilon}^{(j)} = \mathbf{D}^{(j)}(\mathbf{I})\hat{\theta}$.

The distribution function of $Z_{J^{(6)}(i)}$ is thus

$$\sum_{j=i}^{k-1}\int \Phi^{(k-i)}\left(\left(\frac{(c - qI_j^{1/2})I_j^{1/2}I_k}{\bar{I}_j}\right)\mathbf{A}^{(j)}_{i:k-1}\mathbf{1}, \mathbf{0}, \mathbf{D}^{(j)}(\mathbf{I}_{i:k})\boldsymbol{\Sigma}(\mathbf{I}_{i:k})\mathbf{D}^{(j)}(\mathbf{I}_{i:k})'\right) d\Phi(qI_k^{1/2})$$

which, again using the result in Lemma 1 and replacing I_j by n_j, is approximately

$$F_i^{(6)}(c) = \sum_{j=i}^{k-1} \Phi^{(k-i)} \Big(c n_k n_j^{1/2} \bar{n}_j^{-1} \mathbf{A}_{i:k-1}^{(j)} \mathbf{1}, \mathbf{0}, \\ \mathbf{D}^{(j)}(\mathbf{n}_{i:k}) \mathbf{\Sigma}(\mathbf{n}_{i:k}) \mathbf{D}^{(j)}(\mathbf{n}_{i:k})' + n_j^2 n_k \bar{n}_j^{-2} \mathbf{A}_{i:k-1}^{(j)} \mathbf{1}\mathbf{1}' \mathbf{A}_{i:k-1}^{(j)\prime} \Big). \tag{14}$$

4.3 Correcting for subgroup selection in stage 1 - Brownian motion approach

The expressions for the distribution functions, $F_i^{(r)}$ given by (4), (6), (8), (12) and (13) can be used for small to moderate values of k using standard software such as `mvtnorm` (Genz et al. 2020) in R. For larger k, however, this can be computationally infeasible, particularly for $r > 1$ that require up to k or $k-2$ evaluations of multivariate normal tail areas. We now obtain approximations to the distribution functions $F_i^{(r)}(r = 1, 2, 3)$ for k large.

Assume that the subgroups of the stage 1 data considered increase in size in equal increments of some size g, with some minimum size, a multiple of g; that is that $n_j = g(j_0 + j)(j = 1, \ldots, k)$. This arises with $g > 1$ if $\lambda_1, \ldots, \lambda_k$ are chosen to divide the stage 1 patients into equal groups such as deciles, or with $g = 1$ where any subgroup of patients with biomarker value exceeding some threshold can be considered, again subject to some minimum size.

With $n_j = g(j_0 + j)$, the off-diagonal terms in $\mathbf{\Sigma}_Z(\mathbf{n})$ are of the form $(j_0 + j_1)^{1/2}(j_0 + j_2)^{-1/2}$ so that $F_i^{(1)}$ does not depend on g. Since $F_i^{(2)}$ and $F_i^{(3)}$ depend only on the distribution of $\mathbf{Z}$ and $J^{(2)}$ and $J^{(3)}$ and the latter are unaffected by scaling of θ and $\mathbf{S}$, these also do not depend on g. We thus set $g = 1$ and take $n_j = j_0 + j(j = 1, \ldots, k)$. From (2), $S_i, \ldots, S_k$ have the same distribution as values of a zero-drift Brownian motion at times $I_i, \ldots, I_k$ (Siegmund 1985). Using the argument above, this is the same distribution as values at times $j_0 + i, \ldots, j_0 + k$. Approximations to (4), (6) and (8) can thus be based on this Brownian motion.

For selection rule 1, let B_t denote the value of a zero-drift Brownian motion at time t and let $t_j = j_0 + j$, then (3) is equal to $Pr(B_{t_j} t_j^{-1/2} \leq c, j = i, \ldots, k)$, that is $1 - Pr(B_{t_j} \geq c t_j^{1/2}$, some $j = i \ldots, k)$, the probability that a Brownian motion observed at unit time increments crosses a boundary of the form $ct^{1/2}$ in $[t_i, t_k]$. Siegmund (1985) gives an approximation to this probability, leading to an approximation to $F_i^{(1)}(c)$ of

$$G_i^{(1)}(c) = \Phi(c) - c\phi(c) \int_{c(j_0+k)^{-1/2}}^{c(j_0+i)^{-1/2}} x^{-1} e^{-\rho x} dx \tag{15}$$

with $\rho = 0.583$ and the integral evaluated numerically.

For selection rule 2 we wish to approximate (5). As $\hat{\theta}_j = S_j / I_j$, these have the same distribution as the slopes of chords to points on B_t at times $I_1, \ldots, I_k$. The distribution of $(\hat{\theta}_{J^{(2)}(i)}, J^{(2)}(i))$ is thus approximately the joint distribution of the maximum slope of a chord to B_t with $t \in (t_i, t_k)$ and the time of this maximum, denoted by $\tilde{\Theta}_{t_i,t_k}$ and R_{t_i,t_k} respectively, with density, $f_{\tilde{\Theta}_{t_i,t_k}, R_{t_i,t_k}}(\tilde{\theta}, r)$, given in Theorem 3 in the Supplementary material. Since $Z_j \geq c$ when $\tilde{\theta}_j \geq cj^{-1/2}$, $F_i^{(2)}(c)$ can be approximated by

$$1 - \int_{t_i}^{t_k} \int_{cr^{-1/2}}^{\infty} f_{\tilde{\Theta}_{t_i,t_k}, R_{t_i,t_k}}(\tilde{\theta}, r) d\tilde{\theta} dr.$$

Integrating with respect to $\tilde{\theta}$ (see Supplementary material Theorem 3), and approximating the integral over

$r \in (t_i, t_k)$ by a sum over $j = j_0 + i + 1, \ldots, j_0 + k - 1$, $F_i^{(2)}(c)$ is approximately

$$G_i^{(2)}(c) = 1 - \sum_{j=i+1}^{k-1} \frac{1}{j + j_0} \left[\sqrt{\frac{2(j_0 + i)}{\pi(j - i)}} \phi(c) \Phi\left(c \sqrt{\frac{k - j}{j + j_0}} \right) + \sqrt{\frac{(j_0 + i)(j_0 + k)}{\pi^2 (k - j)(j - i)}} \left\{ 1 - \Phi\left(c \sqrt{\frac{j_0 + k}{j + j_0}} \right) \right\} \right]. \tag{16}$$

For selection rule 3, we wish to approximate (7). This can be approximated via the joint distribution of the maximum of B_t in (t_i, t_k) and the time of this maximum, denoted by $\tilde{B}_{t_i,t_k}$ and T_{t_i,t_k} respectively. The density, $f_{\tilde{B}_{t_i,t_k}, T_{t_i,t_k}}(b, t)$, of this joint distribution is given in Theorem 2 in the Supplementary material.

Since $Z_j \geq c$ when $S_j \geq cj^{1/2}$, analagous to (8), $F_i^{(3)}(c)$ can be approximated by

$$1 - \int_{t_i}^{t_k} \int_{ct^{1/2}}^{\infty} f_{\tilde{B}_{t_i,t_k}, T_{t_i,t_k}}(b, t) db dt.$$

Integrating with respect to b (see Theorem 2 in Supplementary material) and approximating the integral over $t \in (t_i, t_1)$ by a sum over $j = j_0 + i + 1, \ldots, j_0 + k - 1$, $F_i^{(3)}(c)$ is approximately

$$G_i^{(3)}(c) = 1 - \sum_{j=i+1}^{k-1} \left\{ \frac{1 - \Phi(c\sqrt{(j + j_0)/(i + j_0)})}{\pi \sqrt{(j - i)(k - j)}} + \sqrt{\frac{2}{\pi j (k - j)}} \phi(c) \Phi\left(c \sqrt{\frac{j - i}{j_0 - i}} \right) \right\}. \tag{17}$$

Note that (15), (16) and (17) are such that $G_i^{(r)}$ $(r = 1, 2, 3)$ are increasing in i. Thus $\max_{i=1,\ldots,J^{(r)}}\{1 - G_i^{(r)}(z)\} = 1 - G_1^{(r)}(z)$ so that this provides a p-value for testing $H_{J^{(r)}}$.

5 Example and simulation study

This section applies the approaches described above to real and resampled data from the German Breast Cancer Study (GBCS) comparing survival times with or without hormone therapy treatment. Data are available at `ftp://ftp.wiley.com/public/sci_tech_med/survival/` (see Schmoor et al. 1996; Hosmer et al. 2008). We investigate if the baseline number of progesterone receptors can be used to identify a subgroup for whom hormone therapy is effective. We first consider construction of corrected p-values for a subgroup selected on the basis of the number of progesterone receptors from all 686 women in the GBCS dataset, that is treating the full dataset as stage 1 of an adaptive enrichment design.

We first consider a case with k small. Approximate deciles of the observed number of progesterone receptors, X, are 0, 5, 10, 20, 30, 60, 100, 160 and 300. To ensure that subgroups are of a reasonable size, we will consider $k = 9$ subgroups with $X > \lambda_i, i = 1, \ldots, 9$ for $(\lambda_1, \ldots, \lambda_9) = (160, 100, 60, 30, 20, 10, 5, 0, -1)$, the last value to include all patients.

Table 1 gives results from a Cox model comparing survival times for women in these subgroups who did and did not receive hormone therapy. For subgroup $j, j = 1, \ldots, 9$, the table gives the number of observations, n_j, estimated log-hazard ratio, $\hat{\theta}_j$ (positive values corresponding to improved survival in the hormone therapy group), corresponding Wald statistic, Z_j, for a one-sided test of the null hypothesis $H_j : \theta_j \leq 0$, $n_j\hat{\theta}_j$, a measure of impact from selecting this subgroup, interaction test statistic for comparing treatment effects in this group and its complement, $Z_j^{(int)}$ and unweighted and weighted estimated difference in treatment effects, $(\hat{\theta}_j - \hat{\bar{\theta}}_j)$ and $n_j(\hat{\theta}_j - \hat{\bar{\theta}}_j)$. Maxima of Z_j, $\hat{\theta}_j$, $n_j\hat{\theta}_j$, $Z_j^{(int)}$, $(\hat{\theta}_j - \hat{\bar{\theta}}_j)$ and $n_j(\hat{\theta}_j - \hat{\bar{\theta}}_j)$ are given in

Table 1: Results of analysis of the GBCS data for $k = 9$ subgroups

j	1	2	3	4	5	6	7	8	9
n_j	144	208	277	352	409	475	531	598	686
Z_j	2.83	3.36	3.41	3.10	**3.41**	3.22	3.35	3.28	2.91
$\hat{\theta}_j$	**1.08**	1.06	0.85	0.63	0.64	0.53	0.51	0.46	0.36
$n_j\hat{\theta}_j$	155.5	219.9	236.7	223.2	262.2	250.7	269.8	**272.9**	249.7
$Z^{(int)}$	2.01	**2.53**	2.27	1.68	2.07	1.83	1.85	2.23	-
$\hat{\theta}_j - \hat{\bar{\theta}}_j$	0.80	**0.87**	0.66	0.43	0.52	0.46	0.50	0.71	-
$n_j(\hat{\theta}_j - \hat{\bar{\theta}}_j)$	115.6	180.2	182.0	152.8	213.7	220.8	263.9	**423.9**	-

Table 2: Results of analysis of the GBCS data with $k = 9$ and $k = 637$

		$k = 9$				$k = 637$	
Selection rule (r)	Statistic maximized	$J^{(r)}$	$Z_{J^{(r)}}$	p-value using $F^{(r)}$	p-value using $G^{(r)}$ $^{(a)}$	$Z_{J^{(r)}}$	p-value using $G^{(r)}$ $^{(a)}$
1	Z_j	5	3.41	0.0016	0.0016	3.86	0.0010
2	$\hat{\theta}_j$	1	2.83	0.0065	0.0071	2.85	0.0133
3	$n_j\hat{\theta}_j$	8	3.28	0.0016	0.0024	3.37	0.0027
4	$Z^{(int)}$	2	3.36	0.0017	0.0019	3.86	0.0010
5	$\hat{\theta}_j - \hat{\bar{\theta}}_j$	2	3.36	0.0015	0.0019	3.08	0.0130
6	$n_j(\hat{\theta}_j - \hat{\bar{\theta}}_j)$	8	3.28	0.0012	0.0025	3.08	0.0130

$^{(a)}$ using $G^{(1)}$ for $r \geq 4$

bold. The resulting values of $J^{(r)}$ and $Z_{J^{(r)}}(r = 1, \ldots, 6)$ are given in the left hand part of Table 2 together with p-values given by $\max_{i=1,\ldots,J^{(r)}}\{1 - F_i^{(r)}(Z_{J^{(r)}(i)})\}(r = 1, \ldots, 6)$ and $1 - G_1^{(r)}(Z_{J^{(r)}})(r = 1, 2, 3)$. For selection rules 4, 5 and 6 values of $1 - G_1^{((1))}(Z_{J^{(r)}})$ are given, providing a conservative test. The values using $F_i^{(r)}$ were calculated using the observed values of $n_1, \ldots, n_k$, whilst those using $G_1^{(r)}$ were calculated with $j_0 = 1$ and $k = 9$. The p-values obtained using the two approximations are very similar for selection rules 1 or 2, but are slightly more different for selection rule 3, with $G^{(3)}$ leading to a more conservative test than $F^{(3)}$. The conservatism from using $G^{(1)}$ for selection rules 4 or 5 and particularly rule 6 is illustrated by the larger p-values.

To illustrate an analysis with large k, the same data were analysed considering subgroups with a minimum size, n_1, of 50 and $n_j - n_{j-1} = 1(j = 2, \ldots, k)$ with $k = 637$ so that $n_k = 686$ and the largest subgroup is the whole sample. In this case calculation of the multivariate normal probabilities in $F_i^{(r)}$ would be infeasible. Figure 1 shows values of Z_j, $\hat{\theta}_j$, $j\hat{\theta}_j$, $Z_j^{(int)}$ and $(\hat{\theta}_j - \hat{\bar{\theta}}_j)$ $(j = 1, \ldots, k)$. Values of $J^{(r)}(r = 1, \ldots, 6)$, corresponding to the largest values of Z_j, $\hat{\theta}_j$, $j\hat{\theta}_j$, $Z_j^{(int)}$ and $(\hat{\theta}_j - \hat{\bar{\theta}}_j)$, are indicated by dashed vertical lines. The right hand part of Table 2 gives values of $Z_{J^{(r)}}(r = 1, \ldots, 6)$ and the corresponding p-values given by $1 - G_1^{(r)}(Z_{J^{(r)}})(j = 1, 2, 3)$ and $1 - G_1^{(1)}(Z_{J^{(r)}})(r = 4, 5, 6)$.

To estimate type I error rates for the approaches proposed an adaptive enrichment design was simulated with 400 stage 1 patients with event/censored times and censoring indicator drawn with replacement from

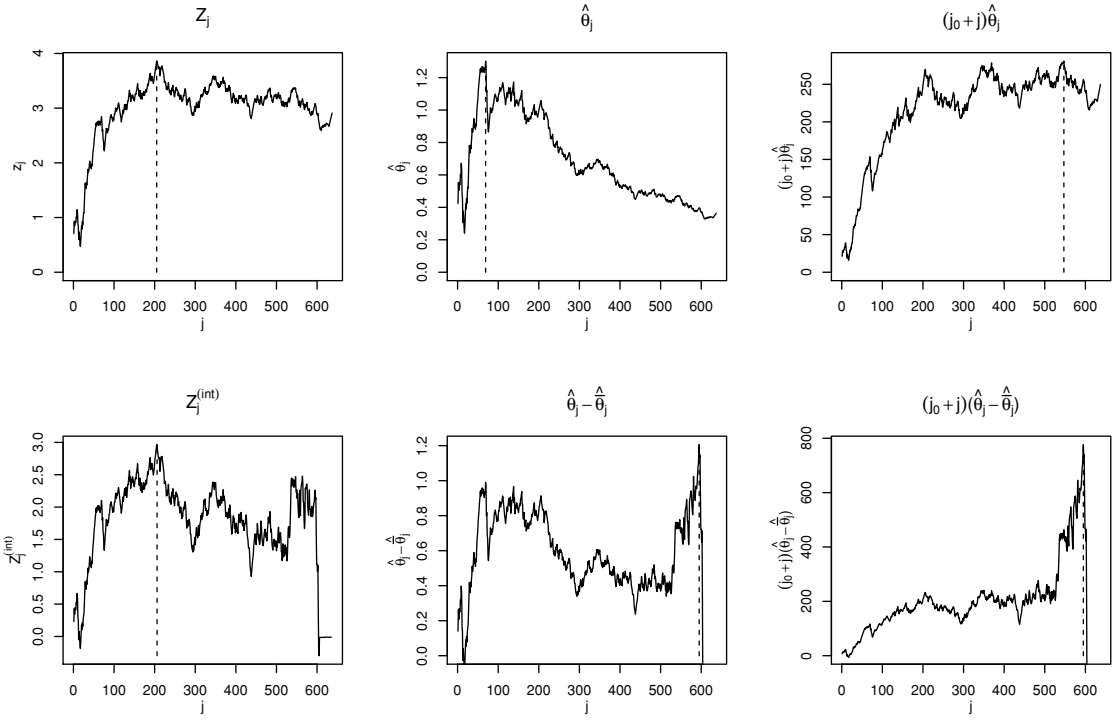

Figure 1: Results of analysis of GBCS data for $k = 637$

the non-hormone replacement therapy group in the GBCS dataset, with these equally assigned at random to treated and control groups. Based on these data a subgroup of patients with the number of progesterone receptors exceeding some threshold was selected using the rules above, and p-values calculated as proposed. A further 400 stage 2 patients were resampled from the non-hormone replacement therapy GBCS patients with the number of progesterone receptors exceeding the selected threshold, again equally assigned to treated and control groups. The p-values from stage 2 and the selected subgroup in stage 1 were then combined using (1) with $w_1^2 = w_2^2 = 1/2$.

Columns three and four of Table 3 give estimated type I error rates for one-sided level 0.025 tests based on F_r $(r = 1, \ldots, 6)$, $G^{(r)}$ $(r = 1, 2, 3)$, and $G^{(1)}$ $(r = 4, 5, 6)$ for $k = 9$, using the values $\lambda_1, \ldots, \lambda_9$ given above. Using p-values based on $F^{(r)}$ accurately controls type I error rates, possibly with the exception of $r = 3$ where there may be a very slight error rate inflation. The approximation $G^{(r)}$ leads to error rate inflation for $r = 1, 2$ suggesting that $F^{(r)}$ should be used when computationally feasible.

Columns five and six of Table 3 give estimated type I error rates for tests based on $G^{(r)}$ using selection rules $r = 1, \ldots, 3$ and $G^{(1)}$ for selection rules 4, 5 and 6 with $(j_0, k) = (9, 41)$ and $(j_0, k) = (39, 361)$ respectively with 400 patients per stage in each case. Type I errors are reasonably controlled in all cases. The test using $G^{(1)}$ is conservative for selection rules 5 and 6, but less so for rule 4. Empirical distribution functions for simulated p-values (see Appendix 3 in Supplementary material) indicate that while the approximation $G^{(r)}$ controls the type I error rate reasonably for $\alpha = 0.025$, larger p-values based on $G^{(1)}$ for selection rules 1, 4, 5 or 6 may be inaccurate even for large k, as considered in the Discussion.

Table 3: Simulated type I error and power values

		Type I error rates[a]				Power[b]	
Selection rule (r)	Statistic maximized	$k=9$ using $F^{(r)}$	$k=9$ using $G^{(r)}$	$k=41$ using $G^{(r)}$	$k=361$ using $G^{(r)}$	$k=9$ using $F^{(r)}$	$k=9$ Simon & Simon
1	Z_j	0.0250	0.0286	0.0256	0.0266	0.842	0.764
2	$\hat{\theta}_j$	0.0242	0.0307	0.0222	0.0214	0.926	0.864
3	$n_j\hat{\theta}_j$	0.0269	0.0254	0.02425	0.0262	0.732	0.649
4	$Z^{(int)}$	0.0245	0.0251[c]	0.0223[c]	0.0219[c]	0.798	0.716
5	$\hat{\theta}_j - \hat{\bar{\theta}}_j$	0.0244	0.0240[c]	0.0193[c]	0.0149[c]	0.816	0.738
6	$n_j(\hat{\theta}_j - \hat{\bar{\theta}}_j)$	0.0256	0.0194[c]	0.0157[c]	0.0104[c]	0.957	0.924

[a]from 100,000 simulations, [b]from 1,000 simulations, [c]using $G^{(1)}$

To obtain an indication of the power of the proposed method, simulations were conducted with a treatment effect imposed. In a resampling approach as described above, all event and censoring times for patients assigned to the treated group with the number of progesterone receptors more than 100 were multiplied by exp(0.5). The penultimate columns of Table 3 gives an estimate of the power to reject the global null hypothesis H_1 using $F^{(r)}(r = 1, \ldots, 6)$ for $k = 9$. For comparison, estimated power for the combination test using all stage 1 patients as proposed by Simon & Simon (2013) is also given, showing the gain in power.

6 Discussion

This paper provides a p-value from an adaptive enrichment design with selection of a subgroup with values of a continuous biomarker above some threshold. The methods could also be used in single-stage trials with subgroup selection (see Mandrekar & Sargent 2009; Freidlin et al. 2010). FWER is specifically controlled for the selection rules described. More complex rules have been proposed (see Ballarini et al. 2020; Ohwada & Morita 2016; Ondra et al. 2016, 2017; Zhang et al. 2017; Thall 2020; Antoniou et al. 2016; Ondra, Dmitrienko, Friede, Graf, Miller, Stallard & Posch 2017). As noted, FWER is controlled for any selection rule using $F^{(1)}$ or $G^{(1)}$ to give a conservative test. Constructing less conservative tests may be challenging, though approaches like cross-validation (see Zhang et al. 2017) could be used.

Methods proposed are based on normality of test statistics comparing treatments in subgroups of the stage 1 data. Sample sizes will be sufficiently large for asymptotic approximations to hold in most confirmatory adaptive enrichment trials unless selection of very small subgroups is considered. This is unlikely to arise in practice, both because of the unreliable performance of selection methods with small subgroups and because there is unlikely to be a desire to develop treatments for very small patient groups (Lin et al. 2019). Asymptotic normality also holds when adjusting for covariates, including prognostic biomarkers. The use of the combination test requires the test statistics from stages 1 and 2 to satisfy the p-clud condition described by Brannath et al. (2002). This is satisfied, at least asymptotically, if patients recruited in stage 1 analysis but either excluded from the stage 1 analysis because they did not have data available at that time, or, with a time to event endpoint, censored in the stage 1 anaylsis, are included in the stage 2 analysis (Wassmer 2006).

Expressions for $G^{(1)}, G^{(2)}$ and $G^{(3)}$ are based on Brownian motion approximations. For the analysis of the data in Section 5.1 with $k = 9$, p-values are similar whether based on the multivariate normal or Brownian

motion approximations. The expression for $G^{(1)}$ is based on an assumption that c is large (see Siegmund 1985), so that the approximation might be expected to be less accurate for large p-values. Simulation results in the Supplementary material show that this is the case, though this may not be a problem in practice, as interest generally focusses on small p-values. Brownian motion approximations to $F^{(r)}, r \geq 4$, have not been obtained. Although $G^{(1)}$ may be used, this can be conservative for small p-values, particularly for selection rule 5. The interaction test statistic can be related to a Brownian bridge. James et al. (1987), give the distribution of the maximum, but not the joint distribution with the time at which it occurs as needed for the distributions of $Z_{J^{(r)}}, r \geq 4$.

This paper has focussed on FWER. Practical questions, including choice of k, the challenge of interim analysis with long-term follow-up and the properties of selection rules, particularly with stage 1 small, warrant further research. Brownian motion approximations might also prove fruitful for unbiased estimation (Kunzmann et al. 2017; Kimani et al. 2015, 2018).

Acknowledgements

The author is grateful to Peter Kimani and anonymous reviewers for their helpful comments.

References

Antoniou, M., Jorgensen, A. L. & Kolamunnage-Dona, R. (2016), 'Biomarker-guided adaptive trial designs in phase II and phase III: A methodological review', *PLoS ONE* **11**, e0149803.

Ballarini, N. M., Burnett, T., Jaki, T., Jennison, C., König, F. & Posch, M. (2020), 'Evaluation of experiments with adaptive interim analyses', *Statistics in Medicine* **40**, 2939–2956.

Bauer, P. & Köhne, K. (1994), 'Evaluation of experiments with adaptive interim analyses', *Biometrics* **50**, 1029–1041.

Bonetti, M. & Gelber, R. D. (2004), 'Patterns of treatment effects in subsets of patients in clinical trials', *Biostatistics* **5**, 465–481.

Brannath, W., Posch, M. & Bauer, P. (2002), 'Recursive combination tests', *Journal of the American Statistical Association* **97**, 236–244.

Chiu, Y.-D., Koenig, F., Posch, M. & Jaki, T. (2018), 'Design and estimation in clinical trials with subpopulation selection', *Statistics in Medicine* **37**, 4335–4352.

Freidlin, B., McShane, L. M. & Korn, E. L. (2010), 'Randomized clinical trials with biomarkers: Design issues', *Journal of the National Cancer Institute* **102**, 152–160.

Friede, T., Parsons, N. & Stallard, N. (2012), 'A conditional error function approach for subgroup selection in adaptive clinical trials', *Statistics in Medicine* **31**, 4309–4320.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Bornkamp, B., Mæchler, M. & Hothorn, T. (2020), 'Package 'mvtnorm''.
URL: *http://CRAN.R-project.org/web/packages/mvtnorm/mvtnorm.pdf*

Ho, T. W., Pearlman, E., Lewis, D., Hämäläinen, M., Connor, K., Michelson, D., Zhang, Y., Assaid, C., Mozley, L. H., Strickler, N., Bachman, R., Mahoney, E., Lines, C. & Hewitt, D. J. (2012), 'Efficacy and tolerability of rizatriptan in pediatric migraineurs: Results from a randomized, double-blind, placebo-controlled trial using a novel adaptive enrichment design', *Cephalagia* **32**, 750–765.

Hosmer, D. W., Lemeshow, S. & May, S. (2008), *Applied Survival Analysis*, Wiley.

James, B., James, K. L. & Siegmund, D. (1987), 'Tests for a change-point', *Biometrika* **74**, 71–83.

Jenkins, M., Stone, A. & Jennison, C. (2011), 'An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints', *Pharmaceutical Statistics* **10**, 347–356.

Jennison, C. & Turnbull, B. (1997), 'Group sequential analysis incorporating covariate information', *Journal of the American Statistical Association* **92**, 1330–1341.

Kimani, P., Todd, S., Renfro, L. & Stallard, N. (2018), 'Point estimation following two-stage adaptive threshold enrichment clinical trials', *Statistics in Medicine* **37**, 3179–3196.

Kimani, P., Todd, S. & Stallard, N. (2015), 'Estimation after subpopulation selection in adaptive seamless trials', *Statistics in Medicine* **34**, 2581–2601.

Kunzmann, K., Benner, L. & Kieser, M. (2017), 'Point estimation in adaptive enrichment designs', *Statistics in Medicine* **36**, 3935–3947.

Lai, T. L., P.W.Lavori & Tsang, K. W. (2019), 'Adaptive enrichment designs for confirmatory trials', *Statistics in Medicine* **38**, 613–624.

Lehmacher, W. & Wassmer, G. (1999), 'Adaptive sample size calculations in group sequential trials', *Biometrics* **55**, 1286–1290.

Lin, J., Bunn, V. & Liu, R. (2019), 'Practical considerations for subgroups quantification, selection and adaptive enrichment in confirmatory trials', *Statistics in Biopharmaceutical Research* **11**, 407–418.

Lin, Z., Flournoy, N. & Rosenberger, W. F. (2021), 'Inference for a two-stage enrichment design', *Annals of Statistics* . (To appear).

Mandrekar, S. M. & Sargent, D. J. (2009), 'Clinical trial designs for predictive biomarker validation: one size does not fit all', *Journal of Biopharmaceutical Statistics* **19**, 530–542.

Markus, R., Pertiz, E. & Gabriel, K. R. (1976), 'On closed testing procedures with special reference to ordered analysis of variance', *Biometrika* **63**, 655–660.

Ohwada, S. & Morita, S. (2016), 'Bayesian adaptive patient enrollment restriction to identify a sensitive subpopulation using a continuous biomarker in a randomized phase 2 trial', *Pharmaceutical Statistics* **15**, 420–429.

Ondra, T., Dmitrienko, A., Friede, T., Graf, A., Miller, F., Stallard, N. & Posch, M. (2017), 'Methods for indentification and confirmation of targeted subgroups in clinical trials: a systematic review', *Journal of Biopharmaceutical Statistics* **26**, 99–119.

Ondra, T., Jobjörnsson, S., Beckman, R. A., Burman, C.-F., König, F., Stallard, N. & Posch, M. (2016), 'Optimizing trial designs for targeted therapies', *PLoS ONE* **11**, e0163726.

Ondra, T., Jobjörnsson, S., Beckman, R. A., Burman, C.-F., König, F., Stallard, N. & Posch, M. (2017), 'Optimized adaptive enrichment designs', *Statistical Methods in Medical Research* **28**, 2096–2111.

Renfro, L. A., Coughlin, C. M., Grothey, A. M. & Sargent, D. J. (2014), 'Adaptive randomized phase II design for biomarker threshold selection and independent evaluation', *Chinese Clinical Oncology* **3**, 3.

Rosenblum, M., Fang, E. X. & Liu, H. (2020), 'Optimal, two-stage, adaptive enrichment designs for randomized trials, using sparse linear programming', *Journal of the Royal Statistical Society, Series B* **82**, 749 772.

Rosenblum, M., Luber, B., Thompson, R. E. & Hanley, D. (2016), 'Group sequential designs with prospectively planned rules for subpopulation enrichment', *Statistics in Medicine* **21**, 3776–3791.

Rosenblum, M., Qian, T., Du, Y., Qiu, H. & Fisher, A. (2016), 'Multiple testing procedures for adaptive enrichment designs: combining group sequential and reallocation approaches', *Biostatistics* **17**, 650–662.

Schmoor, C., Olschweski, M. & Schumacher, M. (1996), 'Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies', *Statistics in Medicine* **15**, 263–271.

Siegmund, D. (1985), *Sequential Analysis*, Springer Verlag.

Simon, N. (2015), 'Adaptive enrichment designs: applications and challenges', *Clinical Investigations* **5**, 383–391.

Simon, N. & Simon, R. (2013), 'Adaptive enrichment designs for clinical trials', *Biostatistics* **14**, 613–625.

Spiessens, B. & Debois, M. (2010), 'Adjusted significance levels for subgroup analyses in clinical trials', *Contemporary Clinical Trials* **31**, 647–656.

Stallard, N., Hamborg, T., Parsons, N. & Friede, T. (2014), 'Adaptive designs for confirmatory clinical trials with subgroup selection', *Journal of Biopharmaceutical Statistics* **24**, 168–187.

Su, M., Fang, L. & Su, Z. (2013), 'A likelihood and resampling based approach to dichotomizing a continuous biomarker in medical research', *Journal of Biopharmaceutical Statistics* **23**, 637–647.

Thall, P. F. (2020), 'Bayesian cancer clinical trial designs with subgroup-specific decisions', *Contemporary Clinical Trials* **90**, 105860.

Wang, S.-J., Hung, H. M. J. & O'Neill, R. T. (2009), 'Adaptive patient enrichment designs in therapeutic trials.', *Biometrical Journal* **51**, 358–74.

Wang, S.-J., O'Neill, R. T. & Hung, H. M. J. (2007), 'Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset.', *Pharmaceutical Statistics* **6**, 227–44.

Wassmer, G. (2006), 'Planning and analyzing adaptive group sequential survival trials.', *Biometrical Journal* **48**, 714–729.

Zhang, Z., Li, M., Lin, M., Soon, G., Greene, T. & Shen, C. (2017), 'Subgroup selection in adaptive signature designs of confirmatory clinical trials', *Journal of the Royal Statistical Society, Series C* **66**, 345–361.

Zhao, Y.-Q. & LeBlanc, M. (2020), 'Designing precision medicine trials to yield a greater population impact', *Biometrics* **76**, 643–653.

Adaptive enrichment designs with a continuous biomarker

Nigel Stallard

Supplementary material

Appendix 1 Derivation of result used in Section 4.2

Let $\phi^{(k)}(\mathbf{x}, \mu, \mathbf{\Sigma})$ and $\Phi^{(k)}(\mathbf{x}, \mu, \mathbf{\Sigma})$ denote respectively the density and distribution functions of a k-dimensional multivariate normal distribution with mean μ and variance-covariance matrix $\mathbf{\Sigma}$ evaluated at $\mathbf{x}$.

Lemma 1. *Let $\mathbf{1}^{(1)}$ and $\mathbf{1}^{(1,1)}$ denote respectively the k-vector and $k \times k$ matrix given by $\mathbf{1}^{(1)} = (1, 0, \ldots, 0)'$ and*

$$\mathbf{1}^{(1,1)} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

then for scalar A and B ($A \neq 0$),

$$\int_{-\infty}^{\infty} \Phi^{(k)}\left((Aq + B)\mathbf{1}^{(1)}, \mu, \mathbf{\Sigma}\right) \phi^{(1)}(q, \mu_q, \sigma_q^2) dq = \Phi^{(k)}\left((A\mu_q + B)\mathbf{1}^{(1)}, \mu, \mathbf{\Sigma} + A^2\sigma_q^2\mathbf{1}^{(1,1)}\right).$$

Proof.

$$\int_{-\infty}^{\infty} \Phi^{(k)}\left((Aq + B)\mathbf{1}^{(1)}, \mu, \mathbf{\Sigma}\right) \phi^{(1)}(q, \mu_q, \sigma_q^2) dq =$$

$$\int_{-\infty}^{\infty} \Phi^{(k)}\left(\mathbf{0}, \mu - (Aq + B)\mathbf{1}^{(1)}, \mathbf{\Sigma}\right) \phi^{(1)}(q, \mu_q, \sigma_q^2) dq =$$

$$\int_{\mathbf{x} \in (-\infty, 0]^k} \int_{-\infty}^{\infty} \phi^{(k)}\left(\mathbf{x}, \mu - (Aq + B)\mathbf{1}^{(1)}, \mathbf{\Sigma}\right) \phi^{(1)}(q, \mu_q, \sigma_q^2) dq d\mathbf{x} =$$

$$\int_{\mathbf{x} \in (-\infty, 0]^k} \int_{-\infty}^{\infty} \phi^{(k)}\left((Aq + B)\mathbf{1}^{(1)}, \mu - \mathbf{x}, \mathbf{\Sigma}\right) \phi^{(1)}(q, \mu_q, \sigma_q^2) dq d\mathbf{x} =$$

$$\int_{\mathbf{x} \in (-\infty, 0]^k} \int_{-\infty}^{\infty} A^{-k} \phi^{(k)}\left(q\mathbf{1}^{(1)}, (\mu - \mathbf{x} - B\mathbf{1}^{(1)})/A, \mathbf{\Sigma}/A^2\right) \phi^{(1)}(q, \mu_q, \sigma_q^2) dq d\mathbf{x}.$$

The inner integral is a standard convolution so that the expression is equal to

$$\int_{\mathbf{x} \in (-\infty, 0]^k} A^{-k} \phi^{(k)}\left(\mathbf{0}, (\mu - \mathbf{x} - B\mathbf{1}^{(1)})/A - \mu_q\mathbf{1}^{(1)}, \mathbf{\Sigma}/A^2 + \sigma_q^2\mathbf{1}^{(1,1)}\right) d\mathbf{x} =$$

$$\int_{\mathbf{x} \in (-\infty, 0]^k} \phi^{(k)}\left(\mathbf{0}, \mu - \mathbf{x} - (A\mu_q + B)\mathbf{1}^{(1)}, \mathbf{\Sigma} + A^2\sigma_q^2\mathbf{1}^{(1,1)}\right) d\mathbf{x} =$$

$$\int_{\mathbf{x}\in(-\infty,0]^k} \phi^{(k)}\left(\mathbf{x}, \mu - (A\mu_q + B)\mathbf{1}^{(1)}, \mathbf{\Sigma} + A^2\sigma_q^2\mathbf{1}^{(1,1)}\right) d\mathbf{x} =$$

$$\Phi^{(k)}\left(\mathbf{0}, \mu - (A\mu_q + B)\mathbf{1}^{(1)}, \mathbf{\Sigma} + A^2\sigma_q^2\mathbf{1}^{(1,1)}\right) =$$

$$\Phi^{(k)}\left((A\mu_q + B)\mathbf{1}^{(1)}, \mu, \mathbf{\Sigma} + A^2\sigma_q^2\mathbf{1}^{(1,1)}\right).$$

□

Appendix 2 Detailed derivation of the expressions given in Section 4.3

A2.1 Preliminary results

Let ϕ and Φ denote respectively the density and distribution function for a standard normal distribution. We have the following results which will be used in the proofs of Theorems in sections A1.2 to A1.4.

Lemma 2.

$$\frac{1}{\sigma_1}\phi\left(\frac{x-b_1}{\sigma_1}\right)\frac{1}{\sigma_2}\phi\left(\frac{x-b_2}{\sigma_2}\right) =$$
$$\frac{1}{\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}}\phi\left(\frac{x-\frac{b_1\sigma_2^2+b_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}}{\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}}\right)\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\phi\left(\frac{b_1-b_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\right).$$

Proof. The left hand side of the expression in the statement of Lemma 2 is equal to

$$\frac{1}{2\pi\sigma_1^2\sigma_2^2}\exp\left\{-\frac{1}{2\sigma_1^2\sigma_2^2}\left[\sigma_2^2(b_1^2-2b_1x+x^2)+\sigma_1^2(b_2^2-2b_2x+x^2)\right]\right\}. \tag{A.1}$$

The term $[\sigma_2^2(b_1^2-2b_1x+x^2)+\sigma_1^2(b_2^2-2b_2x+x^2)]$ is equal to

$$(\sigma_1^2+\sigma_2^2)\left[\left(x-\frac{b_1\sigma_2^2+b_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2+\frac{\sigma_1^2\sigma_2^2(b_1-b_2)^2}{(\sigma_1^2+\sigma_2^2)^2}\right],$$

and substituting into (A.1) and rearranging gives the required result. □

Lemma 3.

$$\int_{b=u}^{\infty} b\frac{1}{\sigma_1}\phi\left(\frac{b}{\sigma_1}\right)\Phi\left(\frac{b}{\sigma_2}\right)db =$$
$$\sigma_1\phi\left(\frac{u}{\sigma_1}\right)\Phi\left(\frac{u}{\sigma_2}\right)+\frac{\sigma_1^2}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\left(1-\Phi\left(\frac{u\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}\right)\right).$$

Proof. The left hand side of the expression in the statement in Lemma 3 can be rewritten as

$$\int_{b=u}^{\infty}\int_{x=-\infty}^{b} bf_{B,X}(b,x)dxdb$$

where $f_{B,X}(b,x)$ is the bivariate normal density of $(B,X)'$ with

$$\begin{pmatrix} B \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right).$$

We can rewrite this using transformation of variables in terms of B and $Y = B - X$ with

$$\begin{pmatrix} B \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2+\sigma_2^2 \end{pmatrix}\right)$$

to get

$$\int_{b=u}^{\infty}\int_{x=-\infty}^{b} bf_{B,X}(b,x)dxdb = \int_{b=u}^{\infty}\int_{y=0}^{\infty} bf_{B,Y}(b,y)dydb.$$

This is an expectation of a truncated bivariate normal, and is given by Kan and Robotti (2017) to be $\Sigma_{11}(c_1 + c_2)$ where Σ is the variance matrix of $(B, Y)'$, so that $\Sigma_{11} = \sigma_1^2$, and

$$c = \begin{pmatrix} \frac{1}{\sigma_1}\phi\left(\frac{u}{\sigma_1}\right)\Phi\left(\frac{u}{\sigma_2}\right) \\ \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\left(1-\Phi\left(\frac{u\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}\right)\right) \end{pmatrix},$$

giving the stated result. □

A2.2 Derivation of joint distribution of value and time of the maximum of a Brownian motion with drift in an interval

Theorem 1. *Consider a Brownian motion with drift μ, B_t^μ. Let $\tilde{B}_{t_0,t_1}^\mu = \max_{t_0<t<t_1} B_t^\mu$ and $S_{t_0,t_1}^\mu = \arg\max_{t_0<t<t_1} B_t^\mu$.*

The joint density of $(\tilde{B}_{t_0,t_1}^\mu, S_{t_0,t_1}^\mu)$ given by

$$f_{\tilde{B}_{t_0,t_1}^\mu, S_{t_0,t_1}^\mu}(b,s) = \frac{2}{\sqrt{s}}\phi\left(\frac{b-\mu s}{\sqrt{s}}\right)\left[\frac{b}{s}\Phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)+\sqrt{\frac{t_0}{s(s-t_0)}}\phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)\right]$$
$$\left[\frac{1}{\sqrt{t_1-s}}\phi\left(\mu\sqrt{t_1-s}\right)-\mu\left(1-\Phi\left(\mu\sqrt{t_1-s}\right)\right)\right].$$

where $s \in (t_0, t_1)$.

Proof. Buffet (2003) considers this problem for $t_0 = 0$, and gives the joint density function for $(B_{t_1}^\mu, \tilde{B}_{0,t_1}^\mu, S_{0,t_1}^\mu)$ at (a_1, b, s) to be

$$f_{B_{t_1}^\mu, \tilde{B}_{0,t_1}^\mu, S_{0,t_1}^\mu}(a_1,b,s) =$$
$$e^{-\mu^2 t_1/2+\mu a_1}\frac{2b(b-a_1)}{s(t_1-s)}\frac{1}{\sqrt{t_1-s}}\phi\left(\frac{b-a_1}{\sqrt{t_1-s}}\right)\frac{1}{\sqrt{s}}\phi\left(\frac{b}{\sqrt{s}}\right).$$

To get the distribution of the maximum in (t_0, t_1), we can consider a Brownian motion starting at time t_0 with a value $B_{t_0}^\mu = a_0$ with $(B_{t_1}^\mu - B_{t_0}^\mu, \tilde{B}_{t_0,t_1}^\mu - B_{t_0}^\mu, S_{t_0,t_1}^\mu - t_0)$ having the same distribution as $(B_{t_1-t_0}^\mu, \tilde{B}_{0,t_1-t_0}^\mu, S_{0,t_1-t_0}^\mu)$.

The joint distribution of $(B_{t_0}^\mu, B_{t_1}^\mu, \tilde{B}_{t_1,t_0}^\mu, S_{t_0,t_1}^\mu)$ at (a_0, a_1, b, s) is thus given by

$$f_{B_{t_0}^\mu, B_{t_1}^\mu, \tilde{B}_{t_0,t_1}^\mu, S_{t_0,t_1}^\mu}(a_0,a_1,b,s) =$$
$$e^{-\mu^2 t_1/2+\mu a_1}\frac{2(b-a_0)(b-a_1)}{(s-t_0)(t_1-s)}\frac{1}{\sqrt{s-t_0}}\phi\left(\frac{b-a_0}{\sqrt{s-t_0}}\right)$$
$$\frac{1}{\sqrt{t_1-s}}\phi\left(\frac{b-a_1}{\sqrt{t_1-s}}\right)\frac{1}{\sqrt{t_0}}\phi\left(\frac{a_0-\mu t_0}{\sqrt{t_0}}\right)$$

with $a_0 \in (-\infty, \infty)$, $a_1 \in (-\infty, \infty)$, $b \geq \max\{a_0, a_1\}$ and $s \in (t_0, t_1)$.

The joint distribution $f_{\tilde{B}_{t_0,t_1}^\mu, S_{t_0,t_1}^\mu}(b,s)$ is then given by integrating this density over $a_0 \leq b$ and $a_1 \leq b$, giving

$$\int_{-\infty}^{b} e^{-\mu^2 t_1/2+\mu a_1}\frac{(b-a_1)}{(t_1-s)^{3/2}}\phi\left(\frac{b-a_1}{\sqrt{t_1-s}}\right)da_1$$
$$\int_{-\infty}^{b}\frac{2(b-a_0)}{s-t_0}\frac{1}{\sqrt{t_0}}\phi\left(\frac{a_0-\mu t_0}{\sqrt{t_0}}\right)\frac{1}{\sqrt{s-t_0}}\phi\left(\frac{b-a_0}{\sqrt{s-t_0}}\right)da_0. \tag{A.2}$$

Consider first the first integral, that is

$$\int_{-\infty}^{b} e^{-\mu^2 t_1/2+\mu a_1} \frac{(b-a_1)}{(t_1-s)^{3/2}} \phi\left(\frac{b-a_1}{\sqrt{t_1-s}}\right) da_1 =$$

$$\exp\left(\frac{2\mu b-\mu^2 s}{2}\right) \int_{-\infty}^{b} \frac{(b-a_1)}{(t_1-s)^{3/2}\sqrt{2\pi}} \exp\left(\frac{-(b+\mu(t_1-s)-a_1)^2}{2(t_1-s)}\right) da_1.$$

Setting $z = b - a_1 + \mu(t_1 - s)$, this is equal to

$$\exp\left(\frac{2\mu b-\mu^2 s}{2}\right) \int_{z=\mu(t_1-s)}^{\infty} \frac{z}{(t_1-s)^{3/2}} \phi\left(\frac{z}{\sqrt{t_1-s}}\right) - \frac{\mu}{\sqrt{t_1-s}} \phi\left(\frac{z}{\sqrt{t_1-s}}\right) dz.$$

Evaluating the integral of the first term using the expression for the expected value of a truncated normal given by Kan and Robotti (2017), this is equal to

$$\exp\left(\frac{2\mu b-\mu^2 s}{2}\right) \left[\frac{1}{\sqrt{t_1-s}} \phi\left(\frac{\mu(t_1-s)}{\sqrt{t_1-s}}\right) - \mu\left(1-\Phi\left(\frac{\mu(t_1-s)}{\sqrt{t_1-s}}\right)\right)\right] =$$

$$\phi\left(\frac{b-\mu s}{\sqrt{s}}\right) \left(\phi\left(\frac{b}{\sqrt{s}}\right)\right)^{-1} \left[\frac{1}{\sqrt{t_1-s}} \phi\left(\mu\sqrt{t_1-s}\right) - \mu\left(1-\Phi\left(\mu\sqrt{t_1-s}\right)\right)\right].$$

Consider now the second integral in (A.2), that is

$$\int_{-\infty}^{b} \frac{2(b-a_0)}{s-t_0} \frac{1}{\sqrt{t_0}} \phi\left(\frac{a_0}{\sqrt{t_0}}\right) \frac{1}{\sqrt{s-t_0}} \phi\left(\frac{b-a_0}{\sqrt{s-t_0}}\right) da_0.$$

Applying Lemma 2 and writing $b^* = bt_0/s$ and $t_0^* = \sqrt{t_0(s-t_0)/s}$, this is equal to

$$\frac{2}{s-t_0} \frac{1}{\sqrt{s}} \phi\left(\frac{b}{\sqrt{s}}\right) \int_{-\infty}^{b} \frac{b-a_0}{t_0^*} \phi\left(\frac{a_0-b^*}{t_0^*}\right) da_0$$

which, using the result from Kan and Robotti (2017), is equal to

$$\frac{2}{s-t_0} \frac{1}{\sqrt{s}} \phi\left(\frac{b}{\sqrt{s}}\right) \left[(b-b^*)\Phi\left(\frac{b-b^*}{t_0^*}\right) + t_0^* \phi\left(\frac{b-b^*}{t_0^*}\right)\right].$$

Since $b - b^* = b(s-t_0)/s$ and $(b-b^*)/t_0^* = b\sqrt{(s-t_0)/(st_0)}$, this is equal to

$$\frac{2}{\sqrt{s}} \phi\left(\frac{b}{\sqrt{s}}\right) \left[\frac{b}{s}\Phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right) + \sqrt{\frac{t_0}{s(s-t_0)}} \phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)\right].$$

Hence

$$f_{\tilde{B}^{\mu}_{t_0,t_1}, S^{\mu}_{t_0,t_1}}(b,s) = \phi\left(\frac{b-\mu s}{\sqrt{s}}\right) \left(\phi\left(\frac{b}{\sqrt{s}}\right)\right)^{-1} \frac{2}{\sqrt{s}} \phi\left(\frac{b}{\sqrt{s}}\right)$$
$$\left[\frac{b}{s}\Phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right) + \sqrt{\frac{t_0}{s(s-t_0)}} \phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)\right]$$
$$\left[\frac{1}{\sqrt{t_1-s}} \phi\left(\mu\sqrt{t_1-s}\right) - \mu\left(1-\Phi\left(\mu\sqrt{t_1-s}\right)\right)\right]$$

which can be rearranged to give the required expression. □

A2.3 Results for the joint distribution of value and time of the maximum of a zero-drift Brownian motion in an interval

Theorem 2. *Consider a zero-drift Brownian motion, B_t with $\tilde{B}_{t_0,t_1} = \max_{t_0<t<t_1} B_t$ and $S_{t_0,t_1} = \arg\max_{t_0<t<t_1} B_t$. Then*

(i) $(\tilde{B}_{t_0,t_1}, S_{t_0,t_1})'$ has joint density

$$f_{\tilde{B}_{t_0,t_1},S_{t_0,t_1}}(b,s) =$$

$$\frac{\sqrt{2}}{\sqrt{\pi s(t_1-s)}}\phi\left(\frac{b}{\sqrt{s}}\right)\left[\frac{b}{s}\Phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)+\sqrt{\frac{t_0}{s(s-t_0)}}\phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)\right]$$

(ii)

$$\int_u^\infty f_{\tilde{B}_{t_0,t_1},\theta_{t_0,t_1}}(b,s)db =$$

$$\frac{1}{\pi\sqrt{(s-t_0)(t_1-s)}}\left(1-\Phi\left(\frac{u}{\sqrt{t_0}}\right)\right)+\sqrt{\frac{2}{\pi s(t_1-s)}}\phi\left(\frac{u}{\sqrt{s}}\right)\Phi\left(\frac{u}{\sqrt{t_0 s/(s-t_0)}}\right).$$

Proof. (i) This follows directly from Theorem 1 with $\mu = 0$.

(ii) Integrating the density with respect to b, we get

$$\int_u^\infty \frac{\sqrt{2}}{\sqrt{\pi s(t_1-s)}}\phi\left(\frac{b}{\sqrt{s}}\right)\sqrt{\frac{t_0}{s(s-t_0)}}\phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)db \tag{A.3}$$

$$+\int_u^\infty \frac{\sqrt{2}}{\sqrt{\pi s(t_1-s)}}\phi\left(\frac{b}{\sqrt{s}}\right)\frac{b}{s}\Phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)db.$$

The first integral in (A.3) is

$$\int_u^\infty \sqrt{\frac{2t_0^2}{\pi(t_1-s)(s-t_0)^2}}\frac{1}{\sqrt{s}}\phi\left(\frac{b}{\sqrt{s}}\right)\sqrt{\frac{s-t_0}{st_0}}\phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)db$$

which, from Lemma 2, with $\sigma_1 = \sqrt{s}$ and $\sigma_2 = \sqrt{t_0 s/(s-t_0)}$, so that $\sqrt{\sigma_1^2+\sigma_2^2} = s/\sqrt{s-t_0}$ and $\sigma_1\sigma_2/\sqrt{\sigma_1^2+\sigma_2^2} = \sqrt{t_0}$, is equal to

$$\int_u^\infty \sqrt{\frac{2t_0^2}{\pi(t_1-s)(s-t_0)^2}}\frac{1}{\sqrt{t_0}}\phi\left(\frac{b}{\sqrt{t_0}}\right)\frac{\sqrt{s-t_0}}{s}\phi(0)db =$$

$$\frac{t_0}{\pi s\sqrt{(s-t_0)(t_1-s)}}\left(1-\Phi\left(\frac{u}{\sqrt{t_0}}\right)\right).$$

The second integral in (A.3) is

$$\frac{\sqrt{2}}{\sqrt{\pi s^2(t_1-s)}}\int_u^\infty \frac{b}{\sqrt{s}}\phi\left(\frac{b}{\sqrt{s}}\right)\Phi\left(b\sqrt{\frac{s-t_0}{st_0}}\right)db$$

which, from Lemma 3, again with $\sigma_1 = \sqrt{s}$ and $\sigma_2 = \sqrt{t_0 s/(s-t_0)}$, so that $\sigma_1^2/\sqrt{\sigma_1^2+\sigma_2^2} = \sqrt{s-t_0}$, is equal to

$$\sqrt{\left(\frac{2}{\pi s^2(t_1-s)}\right)}\left[\sqrt{s}\phi\left(\frac{u}{\sqrt{s}}\right)\Phi\left(\frac{u}{\sqrt{t_0 s/(s-t_0)}}\right)+\sqrt{\frac{s-t_0}{2\pi}}\left(1-\Phi\left(\frac{u}{\sqrt{t_0}}\right)\right)\right].$$

Summing these expressions and collecting together the $(1-\Phi(u/\sqrt{t_0}))$ terms gives the required result. □

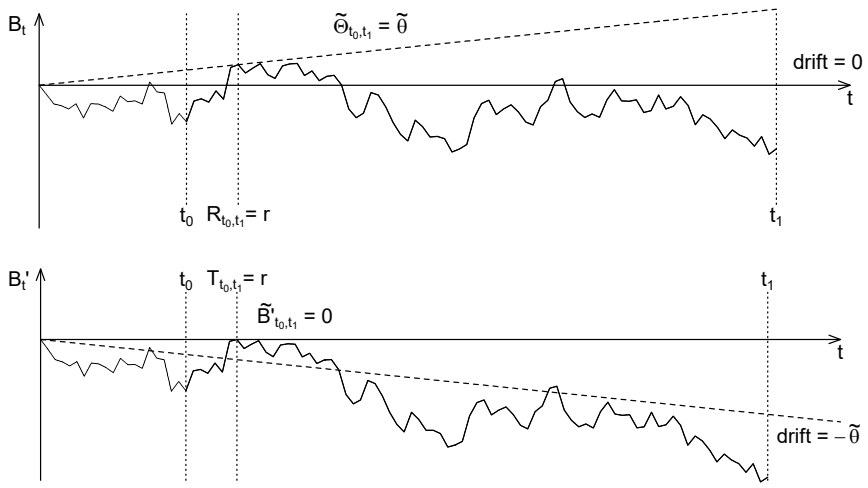

Figure A.1: An example Brownian motion path B_t with steepest chord in (t_0, t_1) with slope $\tilde{\theta}$ at some time r, and $B'_t = B_t - \tilde{\theta}t$ with maximum in (t_0, t_1) of 0 also at time r

A2.4 Results for the joint distribution of value and time of the slope of the steepest chord to a zero-drift Brownian motion in an interval

Theorem 3. *Consider a zero-drift Brownian motion, B_t. Let $\Theta_t = B_t/t$ be the slope of a chord to B_t, and let $\tilde{\Theta}_{t_0,t_1} = \max_{t_0<t<t_1} \Theta_t$ and $R_{t_0,t_1} = \arg\max_{t_0<t<t_1} \Theta_t$. Then*

(i) The joint density of $(\tilde{\Theta}_{t_0,t_1}, R_{t_0,t_1})'$, $f_{\tilde{\Theta}_{t_0,t_1},R_{t_0,t_1}}$, is equal to

$$\sqrt{\frac{2t_0}{\pi(r-t_0)}}\left[\frac{\phi(\tilde{\theta}\sqrt{t_1})}{\sqrt{2\pi(t_1-r)}} + \tilde{\theta}\phi(\tilde{\theta}\sqrt{r})\Phi\left(\tilde{\theta}\sqrt{t_1-r}\right)\right]$$

(ii)

$$\int_u^\infty f_{\tilde{\Theta}_{t_0,t_1},R_{t_0,t_1}}(\tilde{\theta}, r)d\tilde{\theta} =$$

$$\sqrt{\frac{2t_0}{\pi(r-t_0)r^2}}\phi(u\sqrt{r})\Phi(u\sqrt{t_1-r}) + \sqrt{\frac{t_0t_1}{\pi^2r^2(t_1-r)(r-t_0)}}(1-\Phi(u\sqrt{t_1})).$$

Proof. (i) For given $(\tilde{\theta}, r)$, to have $(\tilde{\Theta}_{t_0,t_1}, R_{t_0,t_1}) = (\tilde{\theta}, r)$, we must have $B_r/r = \tilde{\theta}$ and $B_t/t \leq \tilde{\theta}_t$ for all $t \in (t_0, t_1)$. Equivalently, writing $B'_t = B_t - \tilde{\theta}t$, we must have $B'_t \leq 0$ for all $t \in (t_0, t_1)$ and $B'_r = 0$. The event $(\tilde{\Theta}_{t_0,t_1}, R_{t_0,t_1}) = (\tilde{\theta}, r)$ can thus be written as $(\tilde{B}'_{t_0,t_1}, T'_{t_0,t_1}) = (0, r)$ where $\tilde{B}'_{t_0,t_1}$ and T'_{t_0,t_1} are the value and time of the maximum in (t_0, t_1) of B', a Brownian motion with drift $-\tilde{\theta}$. Figure A.1 shows an example path B_t and B'_t with $\tilde{\Theta}_{t_0,t_1} = \tilde{\theta}$ and hence $\tilde{B}'_{t_0,t_1} = 0$.

The event $(\tilde{\Theta}_{t_0,t_1}, R_{t_0,t_1}) = (\tilde{\theta}, r)$ thus corresponds to that of a Brownian motion with drift $-\tilde{\theta}$ having maximum value 0 occuring at time r, so that the density of $(\tilde{\Theta}_{t_0,t_1}, R_{t_0,t_1})'$ can be obtained from that of the value and time of the maximum for a Brownian motion with non-zero drift given by Theorem 1.

Including the Jacobian for a change of variables, we thus have

$$f_{\tilde{\Theta}_{t_0,t_1},R_{t_0,t_1}}(\tilde{\theta},r) = r f_{\tilde{B}^{-\tilde{\theta}}_{t_0,t_1},S^{-\tilde{\theta}}_{t_0,t_1}}(0,r)$$

which, from Theorem 1 is

$$\frac{2r}{\sqrt{r}}\phi\left(\frac{\tilde{\theta}r}{\sqrt{r}}\right)\sqrt{\frac{t_0}{r(r-t_0)}}\phi(0)\left[\frac{1}{\sqrt{t_1-r}}\phi\left(\tilde{\theta}\sqrt{t_1-r}\right)+\tilde{\theta}\Phi(\left(\tilde{\theta}\sqrt{t_1-r}\right)\right] =$$

$$\sqrt{\frac{2t_0}{\pi(r-t_0)}}\phi(\tilde{\theta}\sqrt{r})\left[\frac{1}{\sqrt{t_1-r}}\phi\left(\tilde{\theta}\sqrt{t_1-r}\right)+\tilde{\theta}\Phi\left(\tilde{\theta}\sqrt{t_1-r}\right)\right].$$

From Lemma 2 we have $\phi(\tilde{\theta}\sqrt{r})\phi\left(\tilde{\theta}\sqrt{t_1-r}\right) = \phi(\tilde{\theta}\sqrt{t_1})/\sqrt{2\pi}$ which gives the required result.

(ii) The integral is equal to

$$\sqrt{\frac{t_0}{\pi^2(t_1-r)(r-t_0)t_1}}\int_u^\infty \sqrt{t_1}\phi(c\sqrt{t_1})dc + \sqrt{\frac{2t_0}{\pi(r-t_0)}}\int_u^\infty \tilde{\theta}\phi(\tilde{\theta}\sqrt{r})\Phi\left(\tilde{\theta}\sqrt{t_1-r}\right). \tag{A.4}$$

The integral in the first term is equal to $1-\Phi(u\sqrt{t_1})$. By Lemma 3, taking $\sigma_1 = 1/\sqrt{r}$ and $\sigma_2 = 1/\sqrt{t_1-r}$ so that $\sqrt{\sigma_1^2+\sigma_2^2} = \sqrt{t_1/(r(t_1-r))}$ and $\sigma_1\sigma_2/\sqrt{\sigma_1^2+\sigma_2^2} = 1/\sqrt{t_1}$, the second integral is equal to

$$\frac{1}{r}\phi(u\sqrt{r})\Phi(u\sqrt{t_1-r}) + \sqrt{\frac{t_1-r}{2\pi r^2 t_1}}(1-\Phi(u\sqrt{t_1})).$$

Substituting into (A.4) and collecting the $1-\Phi(u\sqrt{t_1})$ terms gives the required result. □

References

Buffet, E. (2003). On the time of the maximum of Brownian motion with drift. *Journal of Applied Mathematics and Stochastic Analysis 16*, 236 - 207.

Kan, R. and C. Robotti (2017) On moments of folded and truncated multivariate normal distributions. *Journal of Computational and Graphical Statistics 26*, 930 - 934.

Appendix 3 Additional results from simulations in Section 5

Figures A.2 and A.3 shows empirical distribution functions for p-values calculated using $F^{(1)}, \ldots, F^{(6)}$ (heavy black lines) and $G^{(1)}, G^{(2)}, G^{(3)}, G^{(1)}, G^{(1)}$ and $G^{(1)}$ (heavy grey lines) for selection rules 1 to 6 from 100,000 trials resampled with no difference between the treatment groups as described in Section 5 with $k = 9$. As small p-values are generally considered to be of more interest than larger ones, the Figure A.3 is focussed on the lower tail of the distributions. Also shown on the plots are a reference line with zero intercept and unit slope (light line).

Figures A.4 and A.5 shows empirical distribution functions (heavy lines) for p-values calculated using $G^{(1)}, G^{(2)}, G^{(3)}, G^{(1)}, G^{(1)}$ and $G^{(1)}$ from 100,000 trials resampled with no difference between the treatment groups with $(j_0, k) = (39, 361)$, $n_1 = 40$ and $n_i - n_{i-1} = 1, i = 2, \ldots, k$ using different selection rules as described in Section 5, together with a reference line with zero intercept and unit slope (light line). Figure A.5 is focussed on the lower tail of the distributions.

Figures A.6 and A.7 shows empirical distribution functions (heavy lines) for p-values calculated using $G^{(1)}, G^{(2)}, G^{(3)}, G^{(1)}, G^{(1)}$ and $G^{(1)}$ from 100,000 trials resampled with no difference between the treatment groups with $k = 41$ using different selection rules as described in Section 5, together with a reference line with zero intercept and unit slope (light line). Figure A.7 is focussed on the lower tail of the distributions.

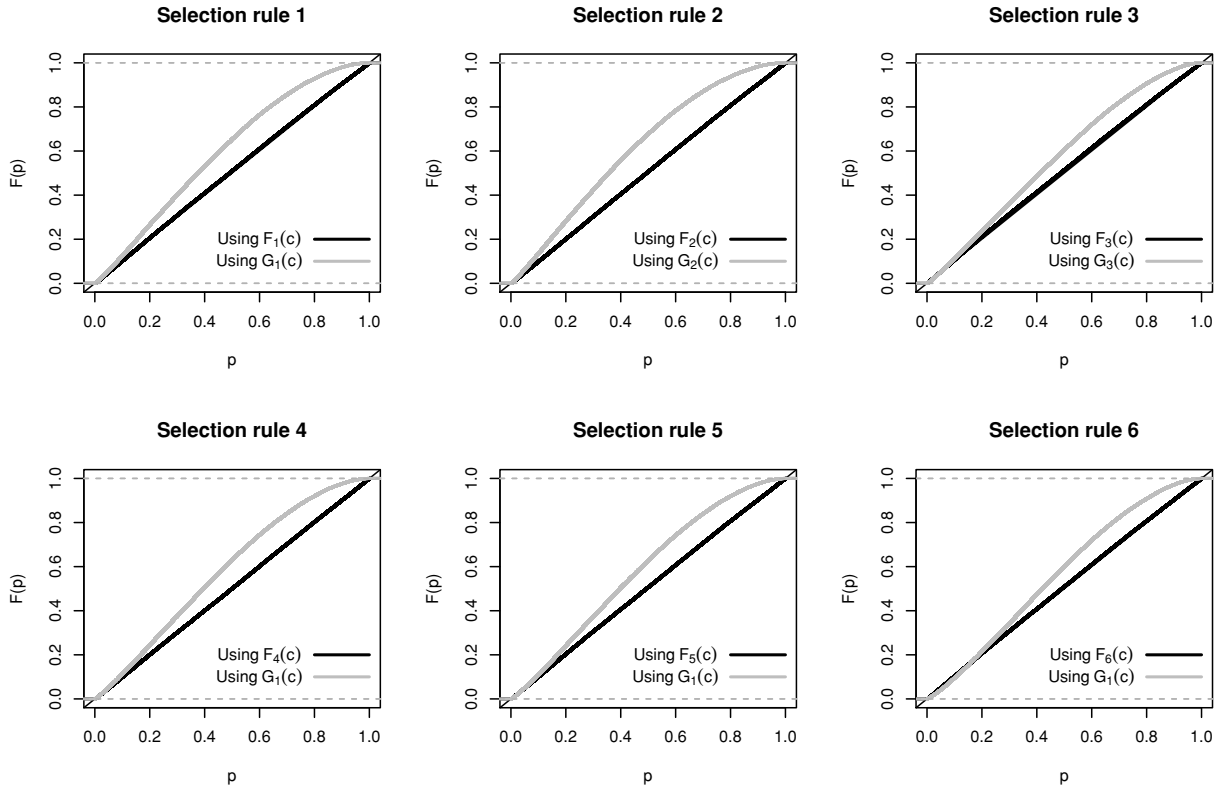

Figure A.2: Emprirical distribution functions for 100,000 simulated p-values in example with $k = 9$

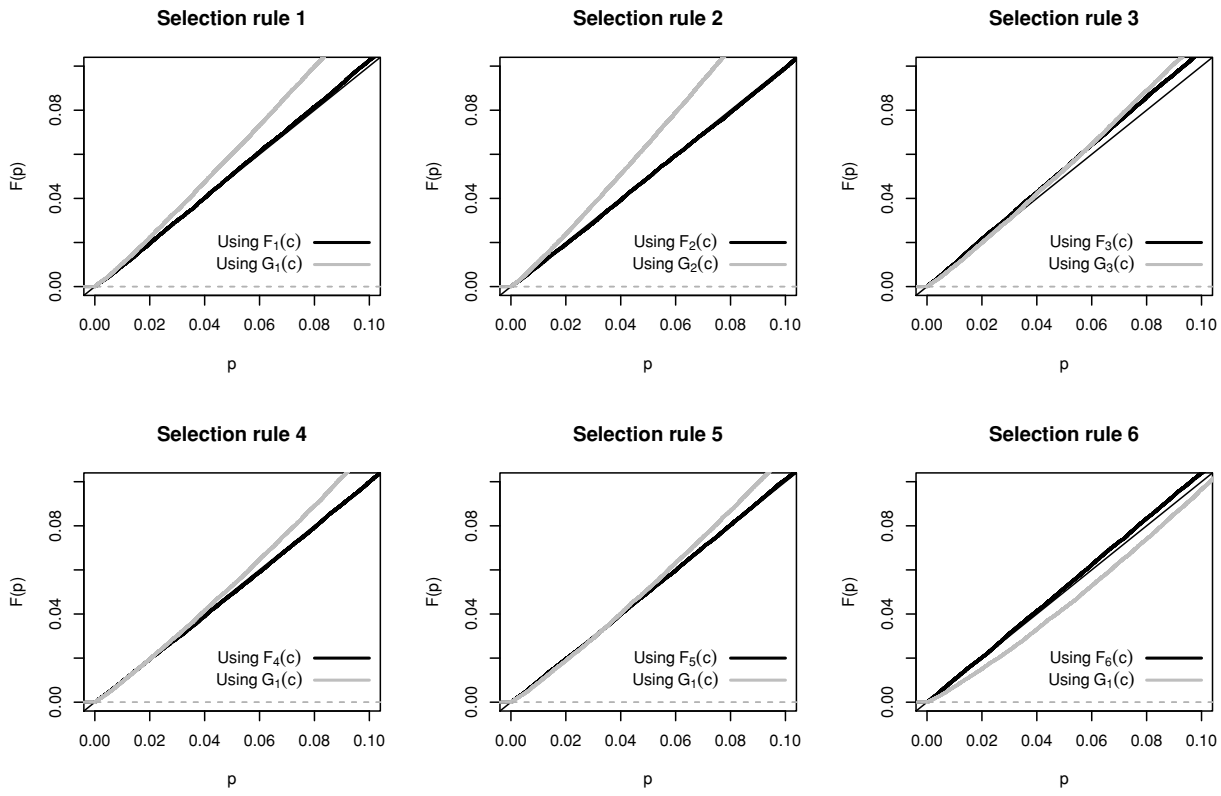

Figure A.3: Emprirical distribution functions for 100,000 simulated p-values in example with $k = 9$ (lower tail only)

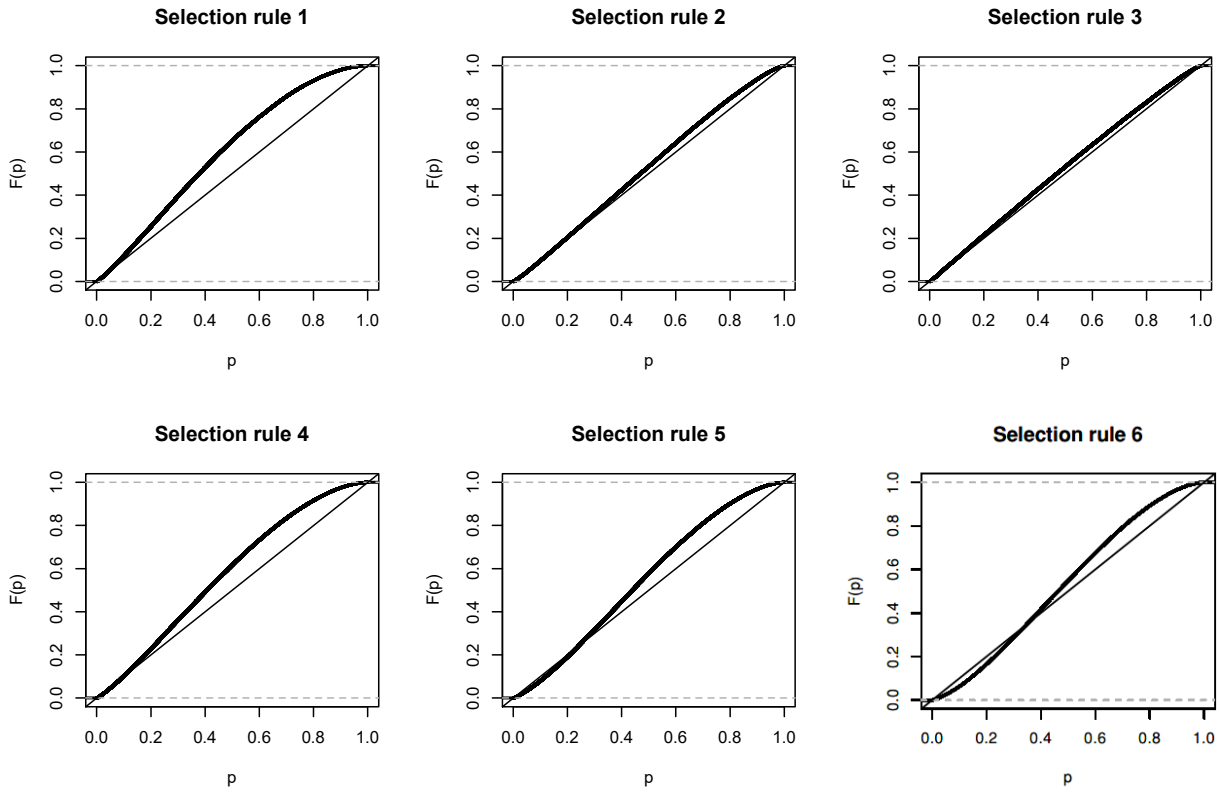

Figure A.4: Emprirical distribution functions for 100,000 simulated p-values in example with $k = 361$

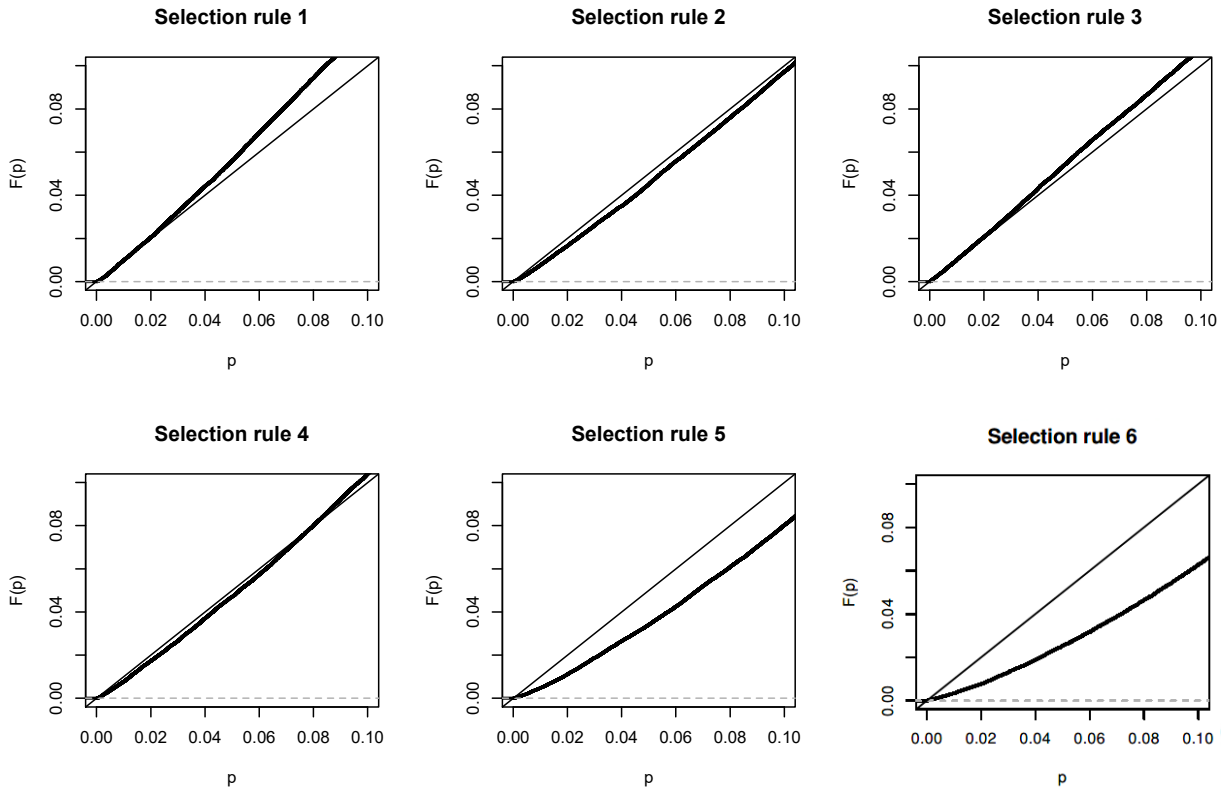

Figure A.5: Emprirical distribution functions for 100,000 simulated p-values in example with $k = 361$ (lower tail only)

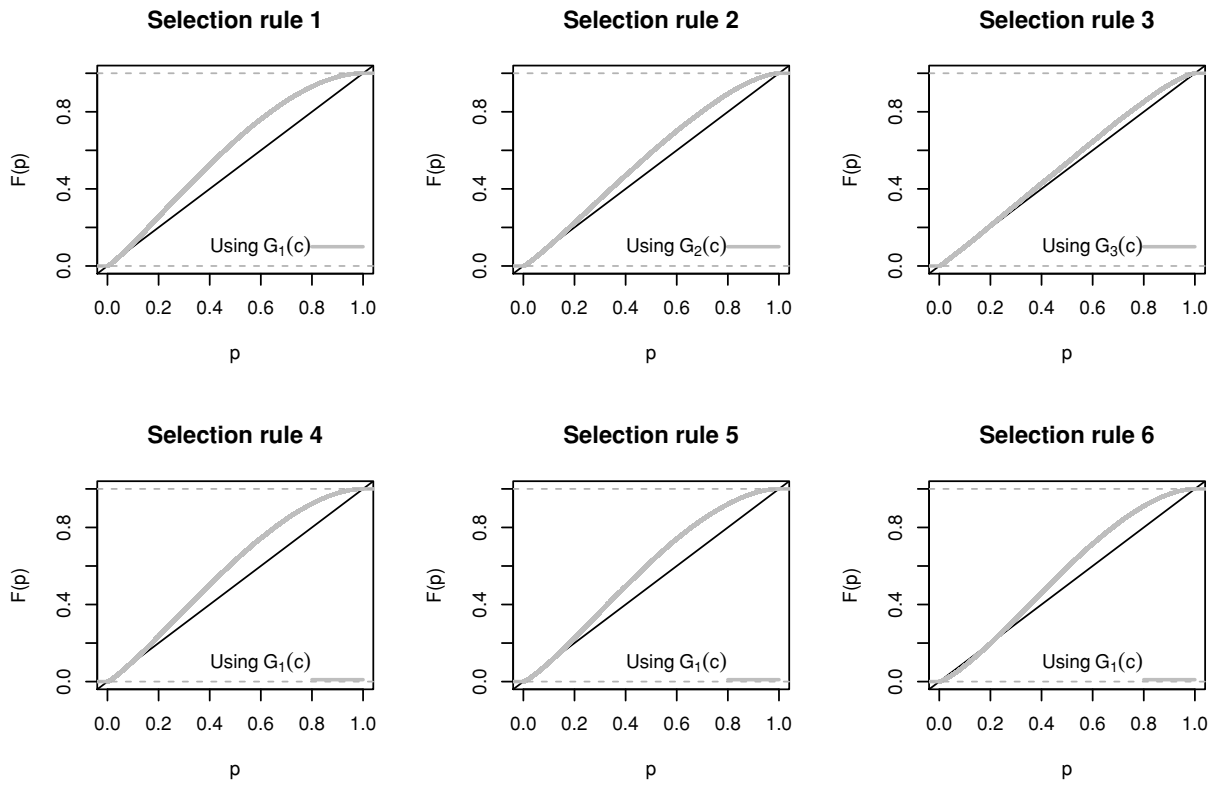

Figure A.6: Emprirical distribution functions for 100,000 simulated p-values in example with $k = 41$

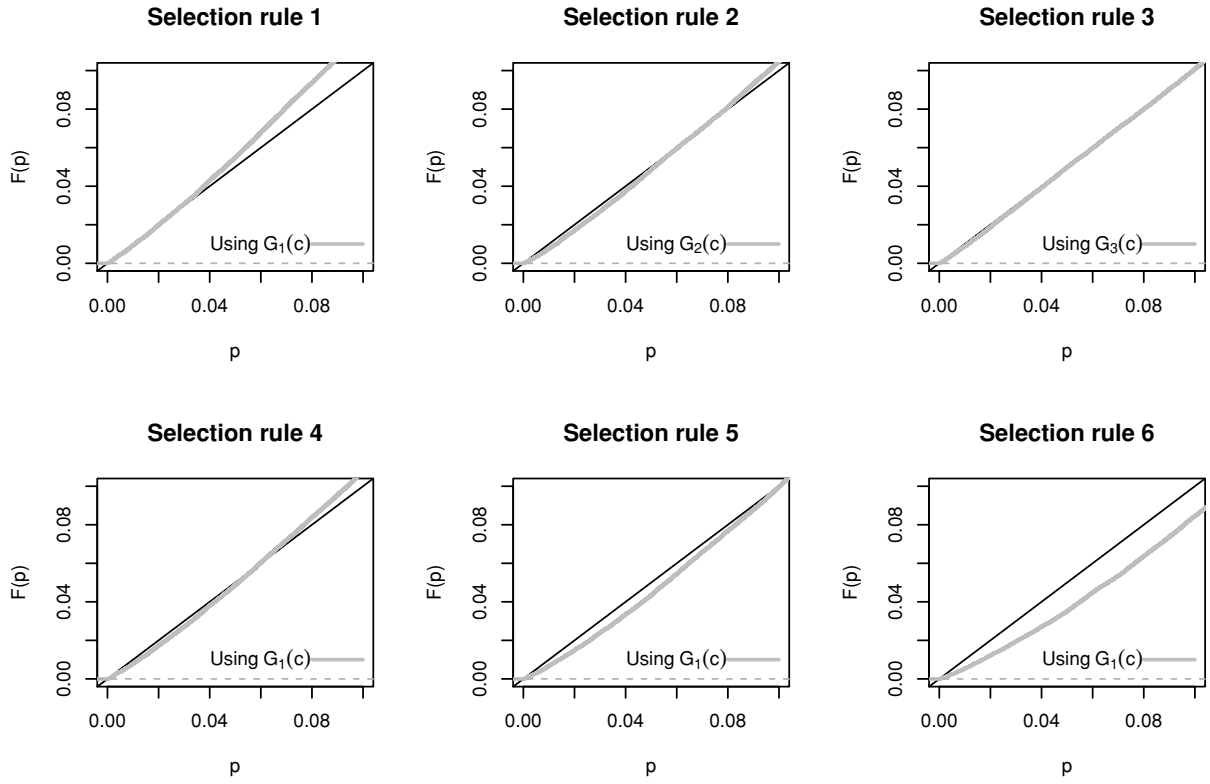

Figure A.7: Emprirical distribution functions for 100,000 simulated p-values in example with $k = 41$ (lower tail only)