Statistics & Probability Letters 70 (2004) 59–70

Missing responses in adaptive allocation design

Atanu Biswas[a,*], J.N.K. Rao[b]

[a] *Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata – 700 108, India*
[b] *School of Mathematics and Statistics, Carleton University, Ottawa, Canada K1S 5B6*

Abstract

Adaptive allocation designs are used in phase III clinical trials. Sometimes, from ethical considerations, the goal may be to skew the allocation pattern in favour of the better treatment. Bandyopadhyay and Biswas (Biometrika 88 (2001) 409) studied such allocation designs for two competing treatments, when the patients heterogeneous with respect to some prognostic factors and the response from each patient was continuous. In the present paper, we extend the work to the case of missing responses. Under missing at random assumption, we impute for the missing data at every stage depending on the data available at that point in time. We obtain the conditional and unconditional allocation probabilities and the standard error of the estimated treatment difference at each stage. Through simulation, we show that imputation for missing responses under this adaptive design set-up has a clear gain over the method that uses only complete data. The gain is in the sense that the power is larger and the standard error of the estimated treatment difference is smaller.

Keywords: Efficiency; Limiting proportion of allocation; Linear regression imputation; Missing at random; Probit link

1. Introduction

Patients arrive sequentially, perhaps for a two-treatment phase III clinical trial, and are to be allocated to one of two treatments. Sometimes, from ethical considerations, the goal may be to

skew the allocation pattern in favour of the better treatment by using available responses to determine the next allocation, using an adaptive allocation design. Bandyopadhyay and Biswas (2001) studied such allocation designs for two competing treatments when the patients were heterogeneous with respect to some prognostic factors and the response from each patient was continuous. They presented an allocation design, assuming a normal theory model. They also evaluated the expected limiting proportions of allocation and studied associated inferences.

Often, in practice, not all the responses may be available. In this case, a common method is to use only complete data at each stage, assuming that the responses are missing at random. However, this method is inefficient as it fails to make use of the prognostic factors associated with the missing responses. The information in the prognostic factors may be utilized through imputation for missing responses; in particular, using regression imputation.

The main purpose of the present paper is to study adaptive allocation designs under imputation for missing responses. We assume a normal theory linear model relating a continuous response, Y, to prognostic factors, x, and treatment effect. We also assume that the probability of response depends only on x and the treatment indicator.

Section 2 provides our adaptive allocation design under regression imputation for missing responses. The associated conditional allocation probability is derived in Section 3, as well as the exact and limiting proportions of allocation. Section 4 gives the standard error of the estimated treatment difference at each stage. Finally, Section 5 reports simulation results on the power of the test of no treatment difference and the standard error of estimated treatment difference, after the completion of the sequential allocation.

2. The allocation design

Suppose patients in the clinical trial arrive sequentially and are assigned to one of two competing treatments A and B, using an adaptive design $\mathscr{D}$. Also, suppose that a predetermined number of patients, v, are to be treated by the trial. Initially, the first $2m$ patients are to be allocated at random to the two treatments, with m patients to each treatment. This ensures that at least m patients are allocated to each treatment. We choose m to ensure that the model parameters can be estimated from the initial sample of size $2m$. We also assume that none of the initial $2m$ responses are missing.

The responses of the patients are assumed to be instantaneous and follow a normal linear model. The model for the response of the ith patient is given by

$$Y_i = \xi_i \mu_A + (1 - \xi_i)\mu_B + x_i^T \beta + \varepsilon_i, \tag{2.1}$$

where ε_i's are independent and identically distributed $N(0, \sigma^2)$ random variables, μ_A and μ_B are the treatment effects and ξ_i is an indicator variable taking the value 1 or 0 according to whether the ith patient is treated by A or B. We focus on the case where some Y-values may be missing but x is completely observed for all the patients. In the presence of missing responses, the data for the ith patient may be represented as $\{Y_i, x_i, \xi_i, \delta_i\}$, where Y_i denotes the response that may or may not be available, x_i is a $p \times 1$ vector of covariates or prognostic factors, and δ_i is another indicator variable that takes the value 1 or 0 according to whether the response of the ith patient is available

or missing. Note that we assume $\delta_i = 1$ for $i = 1, \ldots, 2m$. Model (2.1) may be validated in practice, using standard linear regression diagnostics on the initial complete sample.

Adaptive allocation is carried out from the $(2m+1)$st patient onwards. The allocation of $(n+1)$st patient, $2m \leqslant n \leqslant v-1$, depends on all the previously observed responses, all the previous allocation indicators $\{\xi_1, \ldots, \xi_n\}$, all the previous covariate history $\{x_1, \ldots, x_n\}$, and all the previous response indicators $\{\delta_1, \ldots, \delta_n\}$. Denote the estimator of treatment difference, $\mu_A - \mu_B$, at the $(n+1)$st stage, after imputation for all the previous missing responses and eliminating the effects of prognostic factors x, as $\hat{\mu}^*_{An} - \hat{\mu}^*_{Bn}$. We use a suitable cumulative distribution function $G(\cdot)$, which is symmetric about 0, to implement the allocation of $(n+1)$st patient. Allocate to treatment A with probability $G(\hat{\mu}^*_{An} - \hat{\mu}^*_{Bn})$ and to treatment B with probability $1 - G(\hat{\mu}^*_{An} - \hat{\mu}^*_{Bn}) = G(\hat{\mu}^*_{Bn} - \hat{\mu}^*_{An})$. A natural choice for $G(\cdot)$ is the probit link $G(a) = \Phi(a/T)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function and T is a tuning constant.

We assume that Y is missing at random, that is $P(\delta = 1 | Y, \xi, x) = P(\delta = 1 | \xi, x)$ or δ and Y are conditionally independent given ξ and x. This assumption is reasonable in many practical situations; see Little and Rubin (1987, Chapter 1) for some discussion.

3. Allocation probabilities

3.1. Conditional allocation probability

Let $N_{An} = \sum_{i=1}^n \xi_i$ and $Q_{An} = \sum_{i=1}^n \xi_i \delta_i$ denote the number of allocations and the number of available responses to treatment A, respectively, based on the first n patients. Similarly, let $N_{Bn} = \sum_{i=1}^n (1-\xi_i)$ and $Q_{Bn} = \sum_{i=1}^n (1-\xi_i)\delta_i$ for treatment B. Further let

$$\bar{Y}^C_{An} = \frac{\sum_{i=1}^n \xi_i \delta_i Y_i}{Q_{An}}, \qquad \bar{Y}^C_{Bn} = \frac{\sum_{i=1}^n (1-\xi_i)\delta_i Y_i}{Q_{Bn}},$$

$$\bar{x}^C_{An} = \frac{\sum_{i=1}^n \xi_i \delta_i x_i}{Q_{An}}, \qquad \bar{x}^C_{Bn} = \frac{\sum_{i=1}^n (1-\xi_i)\delta_i x_i}{Q_{Bn}},$$

$$\bar{x}^M_{An} = \frac{\sum_{i=1}^n \xi_i (1-\delta_i) x_i}{N_{An} - Q_{An}}, \qquad \bar{x}^M_{Bn} = \frac{\sum_{i=1}^n (1-\xi_i)(1-\delta_i) x_i}{N_{Bn} - Q_{Bn}}.$$

Here $\bar{Y}^C_{An}$ is the sample mean of the available responses to A, $\bar{x}^C_{An}$ and $\bar{x}^M_{An}$ are the sample mean vectors of the covariates corresponding to the A-treated patients whose responses are available and whose responses are missing, respectively. Similar interpretations are for B-treated patients.

Now, based on the available responses and associated covariates from the first n patients, define

$$S^{(n)}_{xx} = \sum_{i=1}^n \delta_i x_i x_i^T - Q_{An} \bar{x}^C_{An} \bar{x}^{C\,T}_{An} - Q_{Bn} \bar{x}^C_{Bn} \bar{x}^{C\,T}_{Bn},$$

$$S^{(n)}_{xy} = \sum_{i=1}^n \delta_i Y_i x_i - Q_{An} \bar{Y}^C_{An} \bar{x}^C_{An} - Q_{Bn} \bar{Y}^C_{Bn} \bar{x}^C_{Bn}.$$

Estimates of β, μ_A and μ_B are then given by

$$\hat{\beta}_n = S_{xx}^{(n)-1} S_{xy}^{(n)},$$

$$\hat{\mu}_{An} = \bar{Y}_{An}^{C} - (\bar{x}_{An}^{C})^T \hat{\beta}_n,$$

$$\hat{\mu}_{Bn} = \bar{Y}_{Bn}^{C} - (\bar{x}_{Bn}^{C})^T \hat{\beta}_n.$$

For any missing Y_i, we impute it by $\hat{\mu}_{An} + x_i^T \hat{\beta}_n$ if the ith patient is treated by A, and by $\hat{\mu}_{Bn} + x_i^T \hat{\beta}_n$ if the ith patient is treated by B. Thus, after imputation at the nth stage, the imputed ith observation is

$$Z_{in}^{(1)} = \delta_i Y_i + (1 - \delta_i)\{\bar{Y}_{An} + (x_i - \bar{x}_{An}^{C})^T \hat{\beta}_n\},$$

if the ith patient is treated by A, and it is

$$Z_{in}^{(2)} = \delta_i Y_i + (1 - \delta_i)\{\bar{Y}_{Bn} + (x_i - \bar{x}_{Bn}^{C})^T \hat{\beta}_n\},$$

if treated by B. We write the imputed estimator of $\mu_A - \mu_B$, after eliminating the effects of the covariates, as

$$\hat{\mu}_{An}^{*} - \hat{\mu}_{Bn}^{*} = \frac{1}{N_{An}} \sum_{i=1}^{n} \xi_i Z_{in}^{(1)} - \frac{1}{N_{Bn}} \sum_{i=1}^{n} (1 - \xi_i) Z_{in}^{(2)} - (\bar{x}_{An}^{*} - \bar{x}_{Bn}^{*})^T \hat{\beta}_n^{*},$$

where

$$\bar{x}_{An}^{*} = \frac{Q_{An} \bar{x}_{An}^{C} + (N_{An} - Q_{An}) \bar{x}_{An}^{M}}{N_{An}}$$

is the mean of all the x_i's corresponding to the A-treated patients. Similarly, $\bar{x}_{Bn}^{*}$ is defined. The estimator of β based on the imputed data is given by

$$\hat{\beta}_n^{*} = S_{xx}^{(n)*-1} S_{xy}^{(n)*},$$

with

$$S_{xx}^{(n)*} = \sum_{i=1}^{n} \xi_i x_i x_i^T - N_{An} \bar{x}_{An}^{*} \bar{x}_{An}^{*\,T} - N_{Bn} \bar{x}_{Bn}^{*} \bar{x}_{Bn}^{*\,T}, \tag{3.1}$$

and

$$S_{xy}^{(n)*} = \sum_{i=1}^{n} \xi_i x_i (Y_i - \bar{Y}_{An}^{*}) + \sum_{i=1}^{n} (1 - \xi_i) x_i (Y_i - \bar{Y}_{Bn}^{*}), \tag{3.2}$$

where

$$\bar{Y}^*_{An} = \frac{Q_{An}\bar{Y}_{An} + (N_{An} - Q_{An})\bar{Y}_{An} + \left\{\sum_{i:\delta_i=0} \xi_i(x_i - \bar{x}^C_{An})^T\right\}\hat{\beta}_n}{N_{An}}$$

$$= \bar{Y}_{An} + \left(\frac{N_{An} - Q_{An}}{N_{An}}\right)(\bar{x}^M_{An} - \bar{x}^C_{An})^T\hat{\beta}_n.$$

We have a similar expression for $\bar{Y}^*_{Bn}$.

It is easy to show that

$$\hat{\beta}^*_n = \hat{\beta}_n. \tag{3.3}$$

Result (3.3) implies that the covariates associated with the missing responses provide no additional information in estimating β. However, those covariates contribute to the imputed estimators $\hat{\mu}^*_{An}$ and $\hat{\mu}^*_{Bn}$ of μ_A and μ_B, respectively.

We have

$$\hat{\mu}^*_{An} - \hat{\mu}^*_{Bn} = \frac{1}{N_{An}}\sum_{i=1}^{n} \xi_i Z^{(1)}_{in} - \frac{1}{N_{Bn}}\sum_{i=1}^{n}(1 - \xi_i)Z^{(2)}_{in} - (\bar{x}^*_{An} - \bar{x}^*_{Bn})^T\hat{\beta}_n. \tag{3.4}$$

Now, conditionally given $\{\xi_1, \ldots, \xi_n; \delta_1, \ldots, \delta_n; x_1, \ldots, x_n\}$, we write

$$Z^{(1)}_{in} \sim N\left(\mu_A + x_i^T\beta, \sigma^2\left[\delta_i + (1 - \delta_i)\left\{\frac{1}{Q_{An}} + (x_i - \bar{x}^C_{An})^T S^{(n)-1}_{xx}(x_i - \bar{x}^C_{An})\right\}\right]\right),$$

under missing at random assumption. Further, conditionally

$$cov(Z^{(1)}_{in}, Z^{(1)}_{jn}) = (1 - \delta_i\delta_j)\sigma^2\left[\frac{1}{Q_{An}} + (x_i - \bar{x}^C_{An})^T S^{(n)-1}_{xx}(x_j - \bar{x}^C_{An})\right],$$

noting that

$$cov(\bar{Y}^C_{An}, \hat{\beta}_n) = 0,$$

and

$$cov(Y_i, \bar{Y}^C_{An} + (x_j - \bar{x}^C_{An})^T\hat{\beta}_n) = \sigma^2\left[\frac{1}{Q_{An}} + (x_i - \bar{x}^C_{An})^T S^{(n)-1}_{xx}(x_j - \bar{x}^C_{An})\right]$$

Consequently, the conditional variance of $\sum_{i=1}^n \xi_i Z^{(1)}_{in}$ is given by

$$Var\left(\sum_{i=1}^{n} \xi_i Z^{(1)}_{in}\right) = \sum_{i:\xi_i=1} \sigma^2\left[\delta_i + (1 - \delta_i)^2\left\{\frac{1}{Q_{An}} + (x_i - \bar{x}^C_{An})^T S^{(n)-1}_{xx}(x_i - \bar{x}^C_{An})\right\}\right]$$

$$+ \sum_{i,j:i\neq j,\xi_i=1,\xi_j=1} \sigma^2(1 - \delta_i\delta_j)\left\{\frac{1}{Q_{An}} + (x_i - \bar{x}^C_{An})^T S^{(n)-1}_{xx}(x_j - \bar{x}^C_{An})\right\}$$

$$= Q_{An}\sigma^2 + \sigma^2 \sum_{i,j:\xi_i=1,\xi_j=1} (1 - \delta_i\delta_j)\left\{\frac{1}{Q_{An}} + (x_i - \bar{x}^C_{An})^T S^{(n)-1}_{xx}(x_j - \bar{x}^C_{An})\right\}.$$

A similar expression for the conditional variance of $\sum_{i=1}^{n}(1-\xi_i)Z_{in}^{(2)}$ is obtained and the conditional variance of $\widehat{\beta}_n$ is $S_{xx}^{(n)}\sigma^2$. Now noting that the three components of $\widehat{\mu}_{An}^{*}-\widehat{\mu}_{Bn}^{*}$ in (3.4) are conditionally independent, we obtain the conditional variance of $\widehat{\mu}_{An}^{*}-\widehat{\mu}_{Bn}^{*}$ as

$$
\begin{aligned}
V_n &= Var(\hat{\mu}_{An}^{*}-\hat{\mu}_{Bn}^{*}) \\
&= \frac{1}{N_{An}^2}\left\{Q_{An}\sigma^2+\sigma^2\sum_{i,j:\xi_i=1,\xi_j=1}(1-\delta_i\delta_j)\left[\frac{1}{Q_{An}}+(x_i-\bar{x}_{An}^{C})^{T}S_{xx}^{(n)-1}(x_j-\bar{x}_{An}^{C})\right]\right\} \\
&\quad+\frac{1}{N_{Bn}^2}\left\{Q_{Bn}\sigma^2+\sigma^2\sum_{i,j:\xi_i=0,\xi_j=0}(1-\delta_i\delta_j)\left[\frac{1}{Q_{Bn}}+(x_i-\bar{x}_{Bn}^{C})^{T}S_{xx}^{(n)-1}(x_j-\bar{x}_{Bn}^{C})\right]\right\} \\
&\quad+\sigma^2(\bar{x}_{An}^{*}-\bar{x}_{Bn}^{*})^{T}S_{xx}^{(n)-1}(\bar{x}_{An}^{*}-\bar{x}_{Bn}^{*}) \\
&= \sigma^2\left\{\frac{1}{Q_{An}}+\frac{1}{Q_{Bn}}+(\bar{x}_{An}^{*}-\bar{x}_{Bn}^{*})^{T}S_{xx}^{(n)-1}(\bar{x}_{An}^{*}-\bar{x}_{Bn}^{*})\right. \\
&\quad+\left(1-\frac{Q_{An}}{N_{An}}\right)^2(\bar{x}_{An}^{M}-\bar{x}_{An}^{C})^{T}S_{xx}^{(n)-1}(\bar{x}_{An}^{M}-\bar{x}_{An}^{C}) \\
&\quad\left.+\left(1-\frac{Q_{Bn}}{N_{Bn}}\right)^2(\bar{x}_{Bn}^{M}-\bar{x}_{Bn}^{C})^{T}S_{xx}^{(n)-1}(\bar{x}_{Bn}^{M}-\bar{x}_{Bn}^{C})\right\}, \\
&=: \sigma_n^{*2}, \qquad (3.5)
\end{aligned}
$$

noting that

$$\sum_{i,j:\xi_i=1,\xi_j=1}\delta_i\delta_j(x_i-\bar{x}_{An}^{C})^{T}S_{xx}^{(n)-1}(x_j-\bar{x}_{An}^{C})=0$$

and

$$\sum_{i,j:\xi_i=0,\xi_j=0}\delta_i\delta_j(x_i-\bar{x}_{Bn}^{C})^{T}S_{xx}^{(n)-1}(x_j-\bar{x}_{Bn}^{C})=0.$$

Thus, conditionally we have

$$\widehat{\mu}_{An}^{*}-\widehat{\mu}_{Bn}^{*}\sim N(\mu_A-\mu_B,\sigma_n^{*2}). \qquad (3.6)$$

If the responses are not missing, then $\delta_i=1$ for $i=1,\ldots,n$, and (3.5) reduces to expression (4) of Bandyopadhyay and Biswas (2001).

At this stage, we assume that x_i's are independent and identically distributed as $N_p(\mu_x, \Sigma)$. Then, conditionally, σ_n^{*2} has the same distribution as

$$\sigma^2\left[\left(\frac{1}{Q_{An}}+\frac{1}{Q_{Bn}}\right)+\left(\frac{1}{N_{An}}+\frac{1}{N_{Bn}}\right)\frac{p}{Q_n-p-1}W_1\right.$$
$$+\left(1-\frac{Q_{An}}{N_{An}}\right)^2\left(\frac{1}{Q_{An}}+\frac{1}{N_{An}-Q_{An}}\right)\frac{p}{Q_n-p-1}W_2$$
$$\left.+\left(1-\frac{Q_{Bn}}{N_{Bn}}\right)^2\left(\frac{1}{Q_{Bn}}+\frac{1}{N_{Bn}-Q_{Bn}}\right)\frac{p}{Q_n-p-1}W_3\right],$$

where $Q_n = Q_{An} + Q_{Bn}$ and W_1, W_2 and W_3 have an F-distribution with degrees of freedom p and $Q_n - p - 1$. Note that $(Q_n - p - 1)^{-1} W_i \to 0$ in probability as $n \to \infty$, $i = 1, 2, 3$.

3.2. *Unconditional allocation probability*

Let $\psi(n) = P(\xi_{n+1} = 1)$ be the unconditional probability that the $(n+1)$st patient will be allocated to treatment A. Clearly,

$$\begin{aligned}\psi(n) &= E\{P(\xi_{n+1}=1|\text{ data based on the first } n \text{ patients})\}\\ &= E\left\{\Phi\left(\frac{\hat{\mu}_{An}^* - \hat{\mu}_{Bn}^*}{T}\right)\right\}\\ &= E\left[E\left\{\Phi\left(\frac{\hat{\mu}_{An}^* - \hat{\mu}_{Bn}^*}{T}\right)\middle|\delta_1,\ldots,\delta_n;\xi_1,\ldots,\xi_n,W_1,W_2,W_3\right\}\right]\\ &= E\left[P\left(U<\frac{\hat{\mu}_{An}^* - \hat{\mu}_{Bn}^*}{T}\right)\right],\end{aligned}$$

where U follows a standard normal distribution. Thus, using (3.6), the unconditional probability reduces to

$$\psi(n) = E\left\{\Phi\left(\frac{\mu_A - \mu_B}{\sqrt{T^2+\sigma_n^{*2}}}\right)\right\}. \tag{3.7}$$

It follows from (3.7) that $\psi(n) = \frac{1}{2}$ when $\mu_A = \mu_B$. If treatment A is better than B, that is $\mu_A > \mu_B$, then the unconditional probability is skewed in favour of the better treatment, A.

Now, following the proof of Theorem 1 of Bandyopadhyay and Biswas (2001), we get

$$\psi(n) \leqslant \Phi\left(\frac{\mu_A - \mu_B}{T}\right),$$

and it is enough to study the limiting behaviour of $\{\psi(n),\ n \geqslant 2m+1\}$ only at $W_1 = W_2 = W_3 = 0$, i.e. to study the limiting behaviour of

$$\psi^*(n) = E\left\{\Phi\left(\frac{\mu_A - \mu_B}{\sqrt{T^2 + \sigma^2\left(\frac{1}{Q_{An}} + \frac{1}{Q_{Bn}}\right)}}\right)\right\}$$

The sequence $\{\psi^*(n)\}$ is non-decreasing and bounded, as in Bandyopadhyay and Biswas (2001), and hence the sequence converges to a number π^* in the interval (0, 1). Similar to Bandyopadhyay and Biswas (2001), we can show that

$$\pi^* = \Phi\left(\frac{\mu_A - \mu_B}{T}\right).$$

Further,

$$\frac{N_{An}}{n} \to \pi^*$$

in probability, as $n \to \infty$.

4. Gain in precision

If we ignore the missing observations completely and proceed with the $Q_n = Q_{An} + Q_{Bn}$ complete observations, we get the conditional variance of the estimator of $\mu_A - \mu_B$, the treatment difference, as

$$Var(\hat{\mu}_{An} - \hat{\mu}_{Bn}) = \sigma^2\left\{\frac{1}{Q_{An}} + \frac{1}{Q_{Bn}} + (\bar{x}^C_{An} - \bar{x}^C_{Bn})^T S^{(n)-1}_{xx}(\bar{x}^C_{An} - \bar{x}^C_{Bn})\right\}. \tag{4.1}$$

Note that

$$\bar{x}^C_{An} - \bar{x}^C_{Bn} = (\bar{x}^*_{An} - \bar{x}^*_{Bn}) + \left(1 - \frac{Q_{An}}{N_{An}}\right)(\bar{x}^C_{An} - \bar{x}^M_{An}) + \left(1 - \frac{Q_{Bn}}{N_{Bn}}\right)(\bar{x}^M_{Bn} - \bar{x}^C_{Bn}).$$

Hence,

$$\begin{aligned}
&(\bar{x}^C_{An} - \bar{x}^C_{Bn})^T S^{(n)-1}_{xx}(\bar{x}^C_{An} - \bar{x}^C_{Bn}) \\
&\quad = (\bar{x}^*_{An} - \bar{x}^*_{Bn})^T S^{(n)-1}_{xx}(\bar{x}^*_{An} - \bar{x}^*_{Bn}) + \left(1 - \frac{Q_{An}}{N_{An}}\right)^2 (\bar{x}^M_{An} - \bar{x}^C_{An})^T S^{(n)-1}_{xx}(\bar{x}^M_{An} - \bar{x}^C_{An}) \\
&\qquad + \left(1 - \frac{Q_{Bn}}{N_{Bn}}\right)^2 (\bar{x}^M_{Bn} - \bar{x}^C_{Bn})^T S^{(n)-1}_{xx}(\bar{x}^M_{Bn} - \bar{x}^C_{Bn}) + R_n,
\end{aligned}$$

Table 1
Exact expectations and standard deviations, SD, of the proportion of allocation to treatment A, and standard errors of the estimated treatment difference based on the imputed data, SE $(\widehat{\Delta}_v^*)$, and only on the available responses, SE $(\widehat{\Delta}_v)$: $T = 2$

Δ	$E(N_{Av}/v)$	SD (N_{Av}/v)	SE $(\widehat{\Delta}_v^*) \times 100$	SE $(\widehat{\Delta}_v) \times 100$	$\frac{\text{SE}(\widehat{\Delta}_v)}{\text{SE}(\widehat{\Delta}_v^*)}$
Missing type A					
0	0.5000	0.1405	0.825	0.900	1.09
1	0.6527	0.1409	1.023	1.124	1.10
2	0.7674	0.1484	1.745	1.862	1.07
3	0.8294	0.1482	2.711	2.838	1.05
Missing type B					
0	0.5000	0.1402	0.872	1.016	1.16
1	0.6552	0.1406	1.143	1.294	1.13
2	0.7648	0.1412	1.872	2.082	1.08
3	0.8258	0.1403	2.902	3.065	1.06
Missing type C					
0	0.5000	0.1412	0.910	1.106	1.21
1	0.6472	0.1411	1.211	1.409	1.16
2	0.7529	0.1430	1.924	2.125	1.10
3	0.8282	0.1450	2.957	3.177	1.07

where

$$R_n = 2\left\{\left(1-\frac{Q_{An}}{N_{An}}\right)(\bar{x}^*_{An}-\bar{x}^*_{An})^{\mathrm{T}} S_{xx}^{(n)-1}(\bar{x}^{\mathrm{C}}_{An}-\bar{x}^{\mathrm{M}}_{An})\right.$$
$$\left(1-\frac{Q_{Bn}}{N_{Bn}}\right)(\bar{x}^*_{Bn}-\bar{x}^*_{Bn})^{\mathrm{T}} S_{xx}^{(n)-1}(\bar{x}^{\mathrm{M}}_{Bn}-\bar{x}^{\mathrm{C}}_{Bn})$$
$$\left.\left(1-\frac{Q_{An}}{N_{An}}\right)\left(1-\frac{Q_{Bn}}{N_{Bn}}\right)(\bar{x}^{\mathrm{C}}_{An}-\bar{x}^{\mathrm{M}}_{An})^{\mathrm{T}} S_{xx}^{(n)-1}(\bar{x}^{\mathrm{M}}_{An}-\bar{x}^{\mathrm{C}}_{An})\right\}. \tag{4.2}$$

It follows from (3.5), (4.1) and (4.2) that we get benefit from the imputation in terms of efficiency if $R_n \geqslant 0$. But it is difficult to determine R_n for each n. We, therefore, conducted a simulation study in Section 5 to evaluate the gain in efficiency due to imputation for missing responses; see Table 1.

5. Simulation study

We conducted a small simulation study to evaluate the expected proportion of allocation to treatment A and the power of a two-sided test of the null hypothesis $\Delta = \mu_A - \mu_B = 0$ after completion of the experiment with v patients. The two-sided test is based on the imputed estimator

$\hat{\Delta}_v^* = \hat{\mu}_{Av}^* - \hat{\mu}_{Bv}^*$, and the conditional distribution of $\hat{\Delta}_v^*$ is given by (3.6) with $n = v$. We also evaluated the standard error of $\hat{\Delta}_v^*$.

We studied three missing types, denoted by A, B and C, under the missing at random assumption, using a single covariate generated from N(2, 4). We followed Wang and Rao (2002) to specify the missing types A, B and C.

Type A: $P_1(x) = P(\delta = 1|x) = 0.8 + 0.2|x - 1|$ if $|x - 1| \leqslant 1$, and $P_1(x) = \max\{1 - 0.05|x - 1|, 0\}$ otherwise.

Type B: $P_1(x) = P(\delta = 1|x) = 0.9 - 0.2|x - 1|$ if $|x - 1| \leqslant 4.5$, and $P_1(x) = 0$ otherwise.

Type C: $P_1(x) = P(\delta = 1|x) = 0.6$ for all x.

Note that, $EP_1(x) \simeq 0.9$, $EP_2(x) \simeq 0.74$ and $EP_3(x) \simeq 0.6$. Thus, missing type A has the lowest missing rate and missing type C has the highest missing rate.

For each missing type, we generated $R = 10,000$ simulated samples, each of size $v = 40$ from model (2.1) with $\mu_B = 0$, $\beta = 2$, $\sigma = 1$ and specified μ_A, using the single covariate $x \sim N(2, 4)$. We used the probit link $G(a) = \Phi(a/T)$ with $T = 2$ and initial complete sample of size $2m = 8$. The simulated samples were used to evaluate (i) the expected proportion of allocation to treatment A, $E(N_{Av}/v) = \pi_A$ and the standard deviation of N_{Av}/v; (ii) the standard error of the imputed estimator $\hat{\Delta}_v^*$; (iii) the power of the two-sided test: $\Delta = 0$ versus $\Delta \neq 0$ based on $\hat{\Delta}_v^*$. We also calculated the standard error of the estimated treatment difference and the power of the two-sided test based only on the observed responses. Tables 1 and 2 report the above values for $\mu_A = 0, 1, 2$ and 3.

Table 2
Exact and approximate powers of the two-sided test based on the imputed data and only on the available responses: $T = 2$

Δ	Imputed data		Available responses only		Ratio of powers (1)/(3)
	Exact power (1)	Approximate power (2)	Exact power (3)	Approximate power (4)	
Missing type A					
0	0.050	0.061	0.050	0.062	1.00
1	0.282	0.302	0.247	0.264	1.14
2	0.796	0.825	0.759	0.783	1.05
3	0.900	0.930	0.830	0.857	1.08
Missing type B					
0	0.050	0.060	0.050	0.062	1.00
1	0.221	0.245	0.179	0.190	1.23
2	0.683	0.701	0.662	0.641	1.10
3	0.784	0.806	0.720	0.737	1.09
Missing type C					
0	0.050	0.063	0.050	0.061	1.00
1	0.193	0.212	0.130	0.157	1.48
2	0.616	0.633	0.551	0.568	1.18
3	0.726	0.742	0.672	0.689	1.08

The limiting proportion $\pi^* = \Phi\{(\mu_A - \mu_B)/T\}$ for $T = 2$ is given by $\pi^* = 0.50$ for $\Delta = 0$, $\pi^* = 0.69$ for $\Delta = 1$, $\pi^* = 0.84$ for $\Delta = 2$, and $\pi^* = 0.93$ for $\Delta = 3$. Comparing the π^*-values with the corresponding exact (simulated) values $E(N_{Av}/v)$ in Table 1, we see that the exact values are always smaller than the corresponding limiting values. We also note that as Δ increases, the expected proportion $E(N_{Av}/v)$, and the limiting proportion π^*, of allocation to the better treatment increases from $\frac{1}{2}$. Table 1 shows that the standard deviation of N_{Av}/v is not dependent on either Δ or the missing type, while the expected proportion, $E(N_{Av}/v)$, increases with Δ but is not dependent on the missing type.

Turning to the standard error of $\widehat{\Delta}_v^*$, $\mathrm{SE}(\widehat{\Delta}_v^*)$, we see from Table 1 that $\mathrm{SE}(\widehat{\Delta}_v^*)$ increases with Δ and the missing rate. Further, comparing $\mathrm{SE}(\widehat{\Delta}_v^*)$ to $\mathrm{SE}(\widehat{\Delta}_v)$, the standard error of the estimator, $\widehat{\Delta}_v$, based only on the available responses, we see that the imputed estimator $\widehat{\Delta}_v^*$ leads to significant reduction in standard error, especially for higher missing rate and smaller Δ. For example, $\mathrm{SE}(\widehat{\Delta}_v)/\mathrm{SE}(\widehat{\Delta}_v^*) = 1.16$ for $\Delta = 1$ and missing type C.

To calculate the exact power based on the imputed estimator $\widehat{\Delta}_v^*$, we first obtained the 5% level cut-off point $c_{0.05}^*$ from the null distribution of $|\widehat{\Delta}_v^*|$ such that the proportion of $|\widehat{\Delta}_v^*|$-values, obtained from the 10,000 simulation runs, exceeding $c_{0.05}^*$ equals 0.05. We then generated $|\widehat{\Delta}_v^*|$ from the alternative distributions specified by Δ, and computed the exact power as the proportion of $|\widehat{\Delta}_v^*|$-values exceeding $c_{0.05}^*$. We also calculated the approximate power using the critical value $d_{0.05}^*$ based on the normal approximation, $\widehat{\Delta}_v^* \simeq \mathrm{N}(0, \sigma_n^{*2})$, under the null hypothesis $\Delta = 0$, using $\sigma = 1$ specified in the simulation model. In practice, σ^2 is estimated by the residual mean sum of squares based on the Q_v observations with no missing responses. The normal approximation will be valid if Q_v is sufficiently large. Otherwise a t-distribution with $(v - p - 2)$ degrees of freedom will work. For comparison, we also calculated the corresponding exact and approximate power values using only the available responses. Table 2 reports the power values.

The use of $d_{0.05}^*$ as the critical value leads to a slightly inflated size of the test. As a result, the approximate power based on $d_{0.05}^*$ is also larger than the corresponding exact power. For larger sample sizes, v, the use of normal approximation critical value $d_{0.05}^*$ should be satisfactory. Note that in practice $c_{0.05}^*$ is not available. It is clear from Table 2 that the exact power increases with Δ and it decreases as the missing rate increases. We compared the power based on the imputed data with the corresponding power based only on the available data. The ratio of powers, given in the last column of Table 2, indicates that the test based on the imputed estimator $\widehat{\Delta}_v^*$ leads to significant increase in power for smaller Δ and larger missing rates. For example, the ratio is 1.23 for $\Delta = 1$ and missing type B and 1.48 for $\Delta = 1$ and missing type C.

All in all, our simulation study indicates that the use of imputed data can lead to significant reduction in standard error of the estimated treatment difference and significant increase in power.

6. Concluding remarks

The present paper has demonstrated that imputation for missing responses at each stage in adaptive allocation designs can lead to significant gain in terms of standard error of estimated treatment difference and power of two-sided test, relative to the method that uses only available responses. We used a normal linear model in this paper. We propose to study linear models

without the normality assumption by using empirical likelihood methods. We are also extending the results to the case where auxiliary information on X in the form $Eg(X) = 0$ is available.

Acknowledgements

The research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors wish to thank a referee for constructive suggestions.

References

Bandyopadhyay, U., Biswas, A., 2001. Adaptive designs for normal responses with prognostic factors. Biometrika 88 409–419.

Little R.J.A., Rubin, D.B., 1987. Statistical Analysis with Missing Data. Wiley, New York.

Wang, Q., Rao, J.N.K., 2002. Empirical likelihood-based inference in linear models with missing data. Scand. J Statist. 29.