Sample size re-estimation without breaking the blind in clinical trial

Chien-Hua Wu
Department of Applied Mathematics
Chung-Yuan Christian University

Dec. 16, 2011

Table of contents

How many participants we will need to reliably answer the clinical question?

1. Too many participants are a needless waste of resources.
2. Too few participants will not produce a precise, reliable, and definitive answer.
3. Thus, to provide an accurate and reliable sample size, an appropriate method of sample size determination is desirable.

Sample Size Calculation: References

1. A few general references are (1) Chow, Shao, and Wang, (2) Cohen, (3) Desu and Raghavarao, and (4) Lachin.
2. Given the general difficulty of working out an exact answer or a good approximation algebraically, it would be useful, particularly when faced with a new or nonstandard situation, to have an alternative approach that is simple to use yet versatile enough to give an exact solution for a broad range of problems.
3. With the current availability of inexpensive high-speed computing, the proposed procedure (sample size algorithm) is suggested as another computer approach.

Sample Size Algorithms in Clinical Trials: references

1. Chien-Hua Wu*, Shu-Mei Wan, Yu-Chun Yang and Chiung-Yu Huang (2008), Sample size algorithms in clinical trials, Drug Infromation Journal, Vol.42 , No.0 p.429-439.
2. Shu-Mei Wan, Chien-Hua Wu*, Ya-Min Tseng and Ming-Jie Wang (2009), An Improved Algorithm for Sample Size Determination of Ordinal Response by Two Groups, Communications in Statistics - Simulation and Computation , Vol.38 , No.10 p.2235-2242.

Sample Size Algorithms in Clinical Trials

Step 1: Provide an initial sample size.

Step 2: Obtain the critical value under H_0.

Step 3: Calculate the power based on the critical value given by Step 2.

Step 4: If this power is smaller than the desired power, increase the sample size by one unit. Otherwise, reduce the sample size by one unit.

Step 5: Do the iterations from Step 2 to Step 4 until the sample size converges.

Why consider sample size re-estimation?

1. Minimize number of patients exposed to inferior or highly toxic treatment
2. Stop the trial for futility if insufficient benefit
3. Right-size the trial to demonstrate efficacy
4. Reduce or increase sample size
5. Incorporate new internal or external information into a trial design during the course of the trial

Unblind vs. Blind

1. Most work in the literatures is based on unblinded re-estimation of the variance: e.g. Schwartz TA, Denne JS (2003), Friede T, Kieser M. (2006), Proschan MA. (2009)
2. there are also procedures available that do not require unblinding: e.g. Zucker DM, Wittes JT,(1999), Kieser M, Friede T. (2003), Waksman JA. (2007)
3. The European Union of a Note for Guidance (1995) from the Commission for Proprietary Medical Products (CPMP) highlights the importance of blind.

EM algorithm

1. The randomized response model preserved blinding procedure to re-estimate the sample size for binary outcome is proposed by Shih and Zhao (1997).
2. Gould and Shih (1992) uses EM algorithm to estimate individual group means and estimate variances for continuous case
3. Some controversy over appropriateness of EM by Friede and Kieser (2002, 2005) and Gould and Shih (2005).

Randomized response model

1. Given the controversy over appropriateness of EM algorithm, it would be useful to have an alternative approach that is simple to use yet versatile enough for a broad range of models.
2. The proposed procedure is such tool to produce the sample sizes in the middle of the trial without breaking the blind.
3. The proposed model is equivalent to the randomized response model introduced by Shih and Zhao (1997) if we have dichotomized outcomes of a two-arms study.
4. It is also suitable to estimate the means and variances for continuous cases in three-arms study without the normality assumption.

Two arms study

1. Let *I* and *II* denote the responses of stratum one and two, respectively.
2. Subject is randomly assigned to either stratum *I* or *II* and then randomly assigned to treatment A or B. Let $c = 1$ if the subject is assigned to the treatment A and $c = 0$ for the treatment B.
3. In stratum *I*, subject is randomly allocated to treatment group *A* with probability θ and to treatment group *B* with probability $1 - \theta$. In stratum *II*, subject is randomly allocated to treatment group *A* with probability $1 - \theta$ and to treatment group *B* with probability θ.
4. Without breaking the blind, the response X and Y for treatment A and B, respectively, are not observable, but the responses for strata I and II are observable.

Moment generating function

1. Assuming that the treatment response and treatment sequence are independent.
2. $M_I(t) = \theta M_x(t) + (1-\theta) M_y(t)$
3. $M_{II}(t) = (1-\theta) M_x(t) + \theta M_y(t)$

Blind treatment Means

Taking rth derivative with respect to t and setting $t = 0$, we have rth moments for each stratum. That is,
$E(I^r) = \theta E(x^r) + (1-\theta)E(y^r)$ and
$E(II^r) = (1-\theta)E(x^r) + \theta E(y^r)$. If $r = 1$,

$$\mu_I = E(I) = \theta\mu_x + (1-\theta)\mu_y$$

$$\mu_{II} = E(II) = \theta\mu_y + (1-\theta)\mu_x.$$

Solving the equations for μ_x and μ_y, we have

$$\mu_x = \frac{\theta\mu_I - (1-\theta)\mu_{II}}{2\theta - 1}$$

$$\mu_y = \frac{\theta\mu_{II} - (1-\theta)\mu_I}{2\theta - 1}.$$

Dichotomized responses

Note that this model is equivalent to the model provided by Shih and Zhao (1997) if X and Y are dichotomized responses.

Stratum Variances

If $r = 2$,

$$E(I^2) = \theta E(x^2) + (1-\theta)E(y^2)$$

$$E(II^2) = (1-\theta)E(x^2) + \theta E(y^2).$$

The variances of the stratum responses is given by

$$\sigma_I^2 \equiv V(I) = \theta\sigma_x^2 + (1-\theta)\sigma_y^2 + \theta(1-\theta)(\mu_x - \mu_y)^2. \quad (1)$$

$$\sigma_{II}^2 \equiv V(II) = \theta\sigma_y^2 + (1-\theta)\sigma_x^2 + \theta(1-\theta)(\mu_x - \mu_y)^2. \quad (2)$$

Treatment Variances

Solving the equations for σ_x^2 and σ_y^2, we have

$$\sigma_x^2 = \frac{(1-\theta)\sigma_{II}^2 - \theta\sigma_I^2}{1-2\theta} - \theta(1-\theta)(\mu_x - \mu_y)^2$$
$$\sigma_y^2 = \frac{(1-\theta)\sigma_I^2 - \theta\sigma_{II}^2}{1-2\theta} - \theta(1-\theta)(\mu_x - \mu_y)^2.$$

It can be expressed as

$$\sigma_x^2 = \frac{(1-\theta)\sigma_{II}^2 - \theta\sigma_I^2}{1-2\theta} - \theta(1-\theta)(\frac{\mu_I - \mu_{II}}{2\theta-1})^2$$
$$\sigma_y^2 = \frac{(1-\theta)\sigma_I^2 - \theta\sigma_{II}^2}{1-2\theta} - \theta(1-\theta)(\frac{\mu_I - \mu_{II}}{2\theta-1})^2.$$

Unbiased estimators of population means of stratum I and II

- Suppose we have the responses $\{I_1, I_2, \cdots, I_{n_1}\}$ in stratum I and the responses $\{II_1, II_2, \cdots, II_{n_2}\}$ in stratum II.
- The unbiased estimators of population means of stratum I and II are given by $\hat{\mu}_I = \sum_{i=1}^{n_1} I_i/n_1$ and $\hat{\mu}_{II} = \sum_{i=1}^{n_2} II_i/n_2$, respectively.
- Let $\hat{\boldsymbol{\xi}} = (\hat{\mu}_I, \hat{\mu}_{II})'$ and $\boldsymbol{\xi} = (\mu_I, \mu_{II})'$.
- Since stratum I and II are mutually independent, the random vector $\hat{\boldsymbol{\xi}}$ follows a bivariate distribution with mean $\boldsymbol{\xi}$ and variance covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_I^2/n_1 & 0 \\ 0 & \sigma_{II}^2/n_2 \end{pmatrix}$.

Estimators of blind treatment means

The population treatment means can be simplified in a linear model as

$$\begin{pmatrix} \mu_I \\ \mu_{II} \end{pmatrix} = \begin{pmatrix} \theta & 1-\theta \\ 1-\theta & \theta \end{pmatrix} \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

or $\boldsymbol{\xi} = \boldsymbol{\theta}\boldsymbol{\mu}$. If $\boldsymbol{\theta}$ is a nonsingular matrix, we have $\boldsymbol{\mu} = \boldsymbol{\theta}^{-1}\boldsymbol{\xi}$, where

$$\boldsymbol{\theta}^{-1} = \frac{1}{2\theta - 1} \begin{pmatrix} \theta & -1+\theta \\ -1+\theta & \theta \end{pmatrix}.$$

Hence, the blind estimator $\hat{\boldsymbol{\mu}} = \boldsymbol{\theta}^{-1}\hat{\boldsymbol{\xi}}$ follows a bivariate distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\theta}^{-1}\boldsymbol{\Sigma}\boldsymbol{\theta}^{-1}$.

Estimator of unblind treatment mean

Let $\bar{x}_I$ and $\bar{x}_{II}$ denote the sample means of stratum I and II, respectively, for treatment A. The sample mean of treatment A, namely the unblind estimator of μ_x, is given by

$$\bar{x} = \frac{\bar{x}_I n_1 \theta + \bar{x}_{II} n_2 (1-\theta)}{\theta n_1 + (1-\theta) n_2}.$$

Since the samples between strata are mutually independent, the variance of treatment A is

$$\begin{aligned} V(\bar{x}) &= \frac{\sigma_x^2}{(\theta n_1 + (1-\theta) n_2)^2} (n_1 \theta + n_2 (1-\theta)) \\ &= \frac{\sigma_x^2}{\theta n_1 + (1-\theta) n_2}. \end{aligned}$$

Blind vs unblind estimators of treatment means

Since

$$\lim_{\theta \to 1^-} V(\hat{\mu}_x) = \lim_{\theta \to 1^-} V(\bar{x}) = \sigma_x^2/n_1$$

and

$$\lim_{\theta \to 0^+} V(\hat{\mu}_x) = \lim_{\theta \to 0^+} V(\bar{x}) = \sigma_x^2/n_2$$

, the variances of $\hat{\mu}_x$ and $\bar{x}$ both approach to σ_x^2/n_1 as θ approaches 1 and the variances of $\hat{\mu}_x$ and $\bar{x}$ both approach to σ_x^2/n_2 as θ approaches 0.

Unbiased estimators of stratum variances

The unbiased estimators of population variances of stratum I and II are given by $\hat{\sigma}_I^2 = \sum_{i=1}^{n_1}(I_i - \bar{I})^2/(n_1 - 1)$ and $\hat{\sigma}_{II}^2 = \sum_{i=1}^{n_2}(II_i - \bar{II})^2/(n_2 - 1)$, respectively, where $\bar{I} = \sum_{i=1}^{n_1} I_i/n_1$ and $\bar{II} = \sum_{i=1}^{n_2} II_i/n_2$. If the stratum response I and II follow normal distributions, then $\hat{\sigma}_I^2(n_1 - 1)/\sigma_I^2$ and $\hat{\sigma}_{II}^2(n_2 - 1)/\sigma_{II}^2$ follow the chi-square distribution with degree of freedom $n_1 - 1$ and $n_2 - 1$, respectively. Let the random vector $\hat{\boldsymbol{\zeta}} = (\hat{\sigma}_I^2, \hat{\sigma}_{II}^2)'$ and $\boldsymbol{\zeta} = (\sigma_I^2, \sigma_{II}^2)'$. Since $\hat{\sigma}_I^2$ and $\hat{\sigma}_{II}^2$ are independent, the multivariate central limit theorem (Rao 1973, p. 128) implies

$$\hat{\boldsymbol{\zeta}} \xrightarrow{d} N(\boldsymbol{\zeta}, \boldsymbol{\Sigma}_1)$$

, where the covariance matrix

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 2\sigma_I^4/(n_1 - 1) & 0 \\ 0 & 2\sigma_{II}^4/(n_2 - 1) \end{pmatrix}.$$

Blind estimators of treatment variances

Since $\hat{\mu}_I$, $\hat{\mu}_{II}$, $\hat{\sigma}_I^2$ and $\hat{\sigma}_{II}^2$ are mutually independent, the random vector $(\hat{\mu}_I, \hat{\mu}_{II}, \hat{\sigma}_I^2, \hat{\sigma}_{II}^2)'$ follows a quadvariate normal distribution with mean $(\mu_I, \mu_{II}, \sigma_I^2, \sigma_{II}^2)'$ and variance-covariance matrix $\mathbf{\Sigma} = diag(\sigma_I^2/n_1, \sigma_{II}^2/n_2, 2\sigma_I^4/(n_1-1), 2\sigma_{II}^4/(n_2-1))$ asymptotically. Estimating Eqs. (9) and (10), the blind estimators of σ_x^2 and σ_y^2 are given by

$$\hat{\sigma}_x^2 = \frac{(1-\theta)\hat{\sigma}_{II}^2 - \theta\hat{\sigma}_I^2}{1-2\theta} - \theta(1-\theta)(\frac{\hat{\mu}_I - \hat{\mu}_{II}}{2\theta-1})^2$$

$$\hat{\sigma}_y^2 = \frac{(1-\theta)\hat{\sigma}_I^2 - \theta\hat{\sigma}_{II}^2}{1-2\theta} - \theta(1-\theta)(\frac{\hat{\mu}_I - \hat{\mu}_{II}}{2\theta-1})^2.$$

Distribution of Blind estimators of treatment variances

By the delta method, $\hat{\sigma}_x^2$ has distribution similar to the normal with mean σ_x^2 and variance $4\frac{\theta^2(1-\theta)^2}{(2\theta-1)^4}(\mu_I-\mu_{II})^2(\frac{\sigma_I^2}{n_1}+\frac{\sigma_{II}^2}{n_2})+2(\frac{\theta}{1-2\theta})^2\frac{\sigma_I^4}{n_1-1}+2(\frac{1-\theta}{1-2\theta})^2\frac{\sigma_{II}^4}{n_2-1}$, and $\hat{\sigma}_y^2$ has distribution similar to the normal with mean σ_y^2 and variance $4\frac{\theta^2(1-\theta)^2}{(2\theta-1)^4}(\mu_I-\mu_{II})^2(\frac{\sigma_I^2}{n_1}+\frac{\sigma_{II}^2}{n_2})+2(\frac{1-\theta}{1-2\theta})^2\frac{\sigma_I^4}{n_1-1}+2(\frac{\theta}{1-2\theta})^2\frac{\sigma_{II}^4}{n_2-1}$, for large samples. The accuracies of $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ depend on θ, μ_I, μ_{II}, σ_I^2, σ_{II}^2 and n.

Accuracies of Blind estimators of treatment variances depends on θ

$$\hat{\sigma}_x^2 \xrightarrow{d} N(\sigma_x^2, 2\sigma_x^4/(n_1 - 1))$$

and

$$\hat{\sigma}_y^2 \xrightarrow{d} N(\sigma_y^2, 2\sigma_y^4/(n_2 - 1))$$

as $\theta \to 1^-$. Similarly,

$$\hat{\sigma}_x^2 \xrightarrow{d} N(\sigma_x^2, 2\sigma_x^4/(n_2 - 1))$$

and

$$\hat{\sigma}_y^2 \xrightarrow{d} N(\sigma_y^2, 2\sigma_y^4/(n_1 - 1))$$

as $\theta \to 0^+$.

Unlbind estimators of treatment variances

The unblind estimators of σ_x^2 and σ_y^2 are given by

$$S_x^2 = \frac{\sum_{i=1}^{\theta n_1}(x_i - \bar{x})^2 + \sum_{i=1}^{(1-\theta)n_2}(x_i - \bar{x})^2}{\theta n_1 + (1-\theta)n_2 - 1}$$

and

$$S_y^2 = \frac{\sum_{i=1}^{(1-\theta)n_1}(y_i - \bar{y})^2 + \sum_{i=1}^{\theta n_2}(y_i - \bar{y})^2}{(1-\theta)n_1 + \theta n_2 - 1}$$

, respectively.

Unlbind estimators of treatment variances

- Under the normality assumptions, $S_x^2(\theta n_1 + (1-\theta)n_2 - 1)/\sigma_x^2$ and $S_y^2((1-\theta)n_1 + \theta n_2 - 1)/\sigma_y^2$ follow the chi-squared distribution with degrees of freedom $\theta n_1 + (1-\theta)n_2 - 1$ and $(1-\theta)n_1 + \theta n_2 - 1$, respectively.
- Hence, S_x^2 has mean σ_x^2 and variance $2\sigma_x^4/(\theta n_1 + (1-\theta)n_2 - 1)$, and S_y^2 has mean σ_y^2 and variance $2\sigma_y^4/((1-\theta)n_1 + \theta n_2 - 1)$.
- When θ approaches 1, $V(S_x^2) = 2\sigma_x^4/(n_1 - 1)$ and $V(S_y^2) = 2\sigma_y^4/(n_2 - 1)$.
- When θ approaches to 0, $V(S_x^2) = 2\sigma_x^4/(n_2 - 1)$ and $V(S_y^2) = 2\sigma_y^4/(n_1 - 1)$.

Blind vs. unlbind estimators of treatment variances

In conclusion, the mean and variance of blind estimators for σ_x^2 and σ_y^2 are comparable to that of unblind estimators as $\theta \to 1$ or $\theta \to 0$ for large sample size.

Three arms study

- Let *I*, *II* and *III* stand for the responses of stratum one, two and three, respectively.
- Let the indicator $c_1 = 1$ if the subject is assigned to the treatment A, $c_2 = 1$ for treatment B, and $c_3 = 1$ for treatment C, where $\sum_{i=1}^{3} c_i = 1$.
- In stratum I, $P(c_1 = 1) = \theta_1$, $P(c_2 = 1) = \theta_2$ and $P(c_3 = 1) = \theta_3$. In stratum II, $P(c_1 = 1) = \theta_2$, $P(c_2 = 1) = \theta_3$ and $P(c_3 = 1) = \theta_1$. In stratum III, $P(c_1 = 1) = \theta_3$, $P(c_2 = 1) = \theta_1$ and $P(c_3 = 1) = \theta_2$. Let $\sum_{i=1}^{3} \theta_i = 1$.
- Without breaking the blind, the response X, Y and Z for treatment A, B and C, respectively, are not observable, but the responses for strata I, II and III are observable.

Moment generating functions

- Assuming that the stratum response and treatment response are independent.
- $M_I(t) = \theta_1 M_x(t) + \theta_2 M_y(t) + \theta_3 M_z(t)$
- $M_{II}(t) = \theta_2 M_x(t) + \theta_3 M_y(t) + \theta_1 M_z(t)$
- $M_{III}(t) = \theta_3 M_x(t) + \theta_1 M_y(t) + \theta_2 M_z(t)$

rth moments

Taking rth derivative with respect to t and setting $t = 0$ on the equations, we have rth moments for each period. That is,
$E(I^r) = \theta_1 E(x^r) + \theta_2 E(y^r) + \theta_3 E(z^r)$,
$E(II^r) = \theta_2 E(x^r) + \theta_3 E(y^r) + \theta_1 E(z^r)$ and
$E(III^r) = \theta_3 E(x^r) + \theta_1 E(y^r) + \theta_2 E(z^r)$. This can be expressed as

$$\begin{pmatrix} E(I^r) \\ E(II^r) \\ E(III^r) \end{pmatrix} = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_2 & \theta_3 & \theta_1 \\ \theta_3 & \theta_1 & \theta_2 \end{pmatrix} \begin{pmatrix} E(x^r) \\ E(y^r) \\ E(z^r) \end{pmatrix} \tag{3}$$

or $\boldsymbol{\xi^r = \theta \mu^r}$. If $\boldsymbol{\theta}$ is a nonsingular matrix, $\boldsymbol{\mu^r = \theta^{-1} \xi^r}$.

Treatment means and variances

- If $r = 1$, we have the mean vector

$$\mu = \theta^{-1}\xi. \tag{4}$$

- The variance vector is given by

$$\begin{pmatrix} \sigma_x^2 \\ \sigma_y^2 \\ \sigma_z^2 \end{pmatrix} = \theta^{-1} \begin{pmatrix} \sigma_I^2 \\ \sigma_{II}^2 \\ \sigma_{III}^2 \end{pmatrix} + \theta^{-1} \begin{pmatrix} \mu_I^2 \\ \mu_{II}^2 \\ \mu_{III}^2 \end{pmatrix} - \begin{pmatrix} \mu_x^2 \\ \mu_y^2 \\ \mu_z^2 \end{pmatrix}. \tag{5}$$

Unbiased estimators of blind treatment means

We observe the sample of strata I, II and III, which is denoted by $\{I_1, I_2, \cdots, I_{n_1}\}$, $\{II_1, II_2, \cdots, II_{n_2}\}$ and $\{III_1, III_2, \cdots, III_{n_3}\}$, respectively. The unbiased estimator of mean vector for strata I, II and III is given by

$$\begin{aligned} \hat{\boldsymbol{\xi}} &= \begin{pmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \\ \hat{\xi}_3 \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^{n_1} I_i/n_1 \\ \sum_{i=1}^{n_2} II_i/n_2 \\ \sum_{i=1}^{n_3} III_i/n_3 \end{pmatrix}. \end{aligned}$$

The unbiased estimator of treatment mean vector, namely the blind estimator, is given by

$$\hat{\mu} = \theta^{-1}\hat{\xi}$$

Covariance matrix of blind estimator

Since the samples of strata are mutually independent, the covariance matrix of blind estimator can be expressed as

$$\begin{aligned} Cov(\hat{\boldsymbol{\mu}}) &= Cov(\boldsymbol{\theta}^{-1}\hat{\boldsymbol{\xi}}) \\ &= \boldsymbol{\theta}^{-1} Cov(\hat{\boldsymbol{\xi}}) \boldsymbol{\theta}^{-1} \\ &= \boldsymbol{\theta}^{-1} \begin{pmatrix} \sigma_I^2/n_1 & 0 & 0 \\ 0 & \sigma_{II}^2/n_2 & 0 \\ 0 & 0 & \sigma_{III}^2/n_3 \end{pmatrix} \boldsymbol{\theta}^{-1} \end{aligned}$$

, where σ_I^2, σ_{II}^2, σ_{III}^2 stand for the variances of stratum I, II and III, respectively.

Covariance matrix of blind estimator

Since

$$\lim_{\theta_1 \to 1^-} \boldsymbol{\theta}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

$$\lim_{\theta_1 \to 1^-} \begin{pmatrix} \sigma_I^2 \\ \sigma_{III}^2 \\ \sigma_{II}^2 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 \\ \sigma_y^2 \\ \sigma_z^2 \end{pmatrix}.$$

$$\lim_{\theta_1 \to 1^-} Cov(\hat{\boldsymbol{\mu}}) = \lim_{\theta_1 \to 1^-} \begin{pmatrix} \sigma_I^2/n_1 & 0 & 0 \\ 0 & \sigma_{III}^2/n_3 & 0 \\ 0 & 0 & \sigma_{II}^2/n_2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_x^2/n_1 & 0 & 0 \\ 0 & \sigma_y^2/n_3 & 0 \\ 0 & 0 & \sigma_z^2/n_2 \end{pmatrix}.$$

Unblind estimators of treatment means

Let $\bar{x}_I$, $\bar{x}_{II}$ and $\bar{x}_{III}$ denote the sample means of stratum I, II and III, respectively, for treatment A, $\bar{y}_I$, $\bar{y}_{II}$ and $\bar{y}_{III}$ denote the sample means of stratum I, II and III, respectively, for treatment B, and $\bar{z}_I$, $\bar{z}_{II}$ and $\bar{z}_{III}$ denote the sample means of stratum I, II and III, respectively, for treatment C. The sample means of treatment A, B and C, namely the unblind estimators of μ_x, μ_y and μ_z, respectively, are given by

$$\begin{aligned}
\bar{x} &= \frac{\bar{x}_I n_1\theta_1 + \bar{x}_{II} n_2\theta_2 + \bar{x}_{III} n_3\theta_3}{\theta_1 n_1 + \theta_2 n_2 + \theta_3 n_3} \\
\bar{y} &= \frac{\bar{y}_I n_1\theta_2 + \bar{y}_{II} n_2\theta_3 + \bar{y}_{III} n_3\theta_1}{\theta_2 n_1 + \theta_3 n_2 + \theta_1 n_3} \\
\bar{z} &= \frac{\bar{z}_I n_1\theta_3 + \bar{z}_{II} n_2\theta_1 + \bar{z}_{III} n_3\theta_2}{\theta_3 n_1 + \theta_1 n_2 + \theta_2 n_3}.
\end{aligned}$$

Variances of treatment means

Since the samples among strata are mutually independent, the variances of treatment A, B and C are

$$
\begin{aligned}
V(\bar{x}) &= \frac{\sigma_x^2}{\theta_1 n_1 + \theta_2 n_2 + \theta_3 n_3} \\
V(\bar{y}) &= \frac{\sigma_y^2}{\theta_2 n_1 + \theta_3 n_2 + \theta_1 n_3} \\
V(\bar{z}) &= \frac{\sigma_z^2}{\theta_3 n_1 + \theta_1 n_2 + \theta_2 n_3}.
\end{aligned}
$$

Blind treatment means vs. unblind treatment means

$$\lim_{\theta_1 \to 1^-} V(\hat{\mu}_x) = \lim_{\theta \to 1^-} V(\bar{x}) = \sigma_x^2/n_1$$
$$\lim_{\theta_1 \to 1^-} V(\hat{\mu}_y) = \lim_{\theta \to 1^-} V(\bar{y}) = \sigma_y^2/n_3$$
$$\lim_{\theta_1 \to 1^-} V(\hat{\mu}_z) = \lim_{\theta \to 1^-} V(\bar{z}) = \sigma_z^2/n_2.$$

$$\lim_{\theta_2 \to 1^-} V(\hat{\mu}_x) = \lim_{\theta_2 \to 1^-} V(\bar{x}) = \sigma_x^2/n_2$$
$$\lim_{\theta_2 \to 1^-} V(\hat{\mu}_y) = \lim_{\theta_2 \to 1^-} V(\bar{y}) = \sigma_y^2/n_1$$
$$\lim_{\theta_2 \to 1^-} V(\hat{\mu}_z) = \lim_{\theta_2 \to 1^-} V(\bar{z}) = \sigma_z^2/n_3$$

$$\lim_{\theta \to 1^-} V(\hat{\mu}_x) = \lim_{\theta \to 1^-} V(\bar{x}) = \sigma_x^2/n_3$$

Stratum variances

- To estimate the variances, the unbiased estimators of population variances of stratum I, II and III are given by $\hat{\sigma}_I^2 = \sum_{i=1}^{n_1}(I_i - \bar{I})^2/(n_1 - 1)$ and $\hat{\sigma}_{II}^2 = \sum_{i=1}^{n_2}(II_i - \overline{II})^2/(n_2 - 1)$ and and $\hat{\sigma}_{III}^2 = \sum_{i=1}^{n_3}(III_i - \overline{III})^2/(n_3 - 1)$, respectively, where $\bar{I} = \sum_{i=1}^{n_1} I_i/n_1$, $\overline{II} = \sum_{i=1}^{n_2} II_i/n_2$ and $\overline{III} = \sum_{i=1}^{n_3} III_i/n_3$.
- If the stratum response I, II and III follow normal distributions, then $\hat{\sigma}_I^2(n_1 - 1)/\sigma_I^2$, $\hat{\sigma}_{II}^2(n_2 - 1)/\sigma_{II}^2$ and $\hat{\sigma}_{III}^2(n_3 - 1)/\sigma_{III}^2$ follow the chi-square distribution with degree of freedom $n_1 - 1$, $n_2 - 1$ and $n_3 - 1$, respectively.

Stratum variances

- Let the random vector $\hat{\boldsymbol{\zeta}} = (\hat{\sigma}_I^2, \hat{\sigma}_{II}^2, \hat{\sigma}_{III}^2)'$ and $\boldsymbol{\zeta} = (\sigma_I^2, \sigma_{II}^2, \sigma_{III}^2)'$.
- Since $\hat{\sigma}_I^2$, $\hat{\sigma}_{II}^2$ and $\hat{\sigma}_{III}^2$ are independent, the multivariate central limit theorem (Rao 1973, p. 128) implies

$$\hat{\boldsymbol{\zeta}} \xrightarrow{d} N(\boldsymbol{\zeta}, \boldsymbol{\Sigma_1})$$

, where the covariance matrix

$$\boldsymbol{\Sigma_1} = \begin{pmatrix} 2\sigma_I^4/(n_1-1) & 0 & 0 \\ 0 & 2\sigma_{II}^4/(n_2-1) & 0 \\ 0 & 0 & 2\sigma_{III}^4/(n_3-1) \end{pmatrix}.$$

Blind treatment variances

Since $\hat{\mu}_I$, $\hat{\mu}_{II}$, $\hat{\mu}_{III}$, $\hat{\sigma}_I^2$, $\hat{\sigma}_{II}^2$ and $\hat{\sigma}_{III}^2$ are mutually independent, the random vector $(\hat{\mu}_I, \hat{\mu}_{II}, \hat{\mu}_{III}, \hat{\sigma}_I^2, \hat{\sigma}_{II}^2, \hat{\sigma}_{III}^2)'$ asymptotically follows a multivariate normal distribution with mean $(\mu_I, \mu_{II}, \mu_{III}, \sigma_I^2, \sigma_{II}^2, \sigma_{III}^2)'$ and variance-covariance matrix

$$\boldsymbol{\Sigma} = diag(\sigma_I^2/n_1, \sigma_{II}^2/n_2, \sigma_{III}^2/n_3, 2\sigma_I^4/(n_1-1), 2\sigma_{II}^4/(n_2-1), 2\sigma_{III}^4/(n_3-1))$$

Estimating Eq. (17), the blind estimators of variance vector is given by

$$\begin{pmatrix} \hat{\sigma}_x^2 \\ \hat{\sigma}_y^2 \\ \hat{\sigma}_z^2 \end{pmatrix} = \boldsymbol{\theta}^{-1} \begin{pmatrix} \hat{\sigma}_I^2 \\ \hat{\sigma}_{II}^2 \\ \hat{\sigma}_{III}^2 \end{pmatrix} + \boldsymbol{\theta}^{-1} \begin{pmatrix} \hat{\mu}_I^2 \\ \hat{\mu}_{II}^2 \\ \hat{\mu}_{III}^2 \end{pmatrix} - \begin{pmatrix} \hat{\mu}_x^2 \\ \hat{\mu}_y^2 \\ \hat{\mu}_z^2 \end{pmatrix} \qquad (6)$$

, where

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\theta}^{-1}\hat{\boldsymbol{\xi}}$$

and

Blind treatment variance

By the delta method for the vector function of random vector, the asymptotic distribution of the variance vector is given by

$$(\hat{\sigma}_x^2, \hat{\sigma}_y^2, \hat{\sigma}_z^2)' \xrightarrow{d} N((\sigma_x^2, \sigma_y^2, \sigma_z^2)', \phi \boldsymbol{\Sigma} \phi')$$

, where

$$\phi' = \frac{1}{det(\boldsymbol{\theta})} \begin{pmatrix} 2(\theta_2\theta_3 - \theta_1^3)(\mu_I - \mu_x) & -2(\theta_2^2 - \theta_1\theta_3)(\mu_I - \mu_x) & 2(\\ -2(\theta_2^2 - \theta_1\theta_3)(\mu_y - \mu_{II}) & 2(\theta_1\theta_2 - \theta_3^3)(\mu_y - \mu_{II}) & -2 \\ 2(\theta_1\theta_2 - \theta_3^2)(\mu_z - \mu_{III}) & -2(\theta_1^2 - \theta_2\theta_3)(\mu_z - \mu_{III}) & 2(\\ \theta_2\theta_3 - \theta_1^3 & -(\theta_2^2 - \theta_1\theta_3) & \\ -(\theta_2^2 - \theta_1\theta_3) & \theta_1\theta_2 - \theta_3^3 & \\ \theta_1\theta_2 - \theta_3^2 & -(\theta_1^2 - \theta_2\theta_3) & \end{pmatrix}$$

Blind treatment variances

In Eq. (16), $\mu_I \to \mu_x$, $\mu_{II} \to \mu_y$ and $\mu_{III} \to \mu_z$ as $\theta_1 \to 1^-$. Since, in Eq. (17),

$$\lim_{\theta_1 \to 1^-} (\sigma_x^2, \sigma_y^2, \sigma_z^2)' = (\sigma_I^2, \sigma_{II}^2, \sigma_{III}^2)',$$

we have

$$\lim_{\theta_1 \to 1^-} \phi \mathbf{\Sigma} \phi' = diag(2\sigma_x^4/(n_1 - 1), 2\sigma_y^4/(n_3 - 1), 2\sigma_z^4/(n_2 - 1)).$$

That is, $\lim_{\theta_1 \to 1^-} V(\hat{\sigma}_x^2) = 2\sigma_x^4/(n_1 - 1)$,
$\lim_{\theta_1 \to 1^-} V(\hat{\sigma}_y^2) = 2\sigma_y^4/(n_3 - 1)$ and
$\lim_{\theta_1 \to 1^-} V(\hat{\sigma}_z^2) = 2\sigma_z^4/(n_2 - 1)$.

Unblind treatment variances

The unblind estimators of σ_x^2, σ_y^2 and σ_z^2 are given by

$$S_x^2 = \frac{\sum_{i=1}^{\theta_1 n_1}(x_i - \bar{x})^2 + \sum_{i=1}^{\theta_2 n_2}(x_i - \bar{x})^2 + \sum_{i=1}^{\theta_3 n_3}(x_i - \bar{x})^2}{\theta_1 n_1 + \theta_2 n_2 + \theta_3 n_3 - 1}$$

$$S_y^2 = \frac{\sum_{i=1}^{\theta_2 n_1}(y_i - \bar{y})^2 + \sum_{i=1}^{\theta_3 n_2}(y_i - \bar{y})^2 + \sum_{i=1}^{\theta_1 n_3}(y_i - \bar{y})^2}{\theta_2 n_1 + \theta_3 n_2 + \theta_1 n_3 - 1}$$

$$S_z^2 = \frac{\sum_{i=1}^{\theta_3 n_1}(z_i - \bar{z})^2 + \sum_{i=1}^{\theta_1 n_2}(z_i - \bar{z})^2 + \sum_{i=1}^{\theta_2 n_3}(z_i - \bar{z})^2}{\theta_3 n_1 + \theta_1 n_2 + \theta_2 n_3 - 1}$$

, respectively.

Unblind treatment variances

Under the normality assumptions, $S_x^2(\theta_1 n_1 + \theta_2 n_2 + \theta_3 n_3 - 1)/\sigma_x^2$, $S_y^2(\theta_2 n_1 + \theta_3 n_2 + \theta_1 n_3 - 1)/\sigma_y^2$ and $S_z^2(\theta_3 n_1 + \theta_1 n_2 + \theta_2 n_3 - 1)/\sigma_z^2$ follow the chi-squared distribution with degrees of freedom $\theta_1 n_1 + \theta_2 n_2 + \theta_3 n_3 - 1$, $\theta_2 n_1 + \theta_3 n_2 + \theta_1 n_3 - 1$ and $\theta_3 n_1 + \theta_1 n_2 + \theta_2 n_3 - 1$, respectively.

Blind treatment variances vs unblind treatment variances

- When θ_1 approaches to 1, $V(S_x^2) = 2\sigma_x^4/(n_1 - 1)$, $V(S_y^2) = 2\sigma_y^4/(n_3 - 1)$ and $V(S_z^2) = 2\sigma_z^4/(n_2 - 1)$, which are equivalent to the variances of the blind estimators of σ_x^2, σ_y^2 and σ_z^2.
- Similarly, the variances of the unblind estimators are the same as that of the blind estimators as $\theta_2 \to 1^-$ or $\theta_3 \to 1^-$.

Simulation results

- To study its accuracy, the simulation results to calculate the bias and the mean squares errors (MSE) of blind estimator and unblind estimator with 10,000 iterations are provided in Tables 1 and 2 for two-arm study and three-arm study, respectively.
- We generate the random numbers from normal distributions with particular parameters specified in each tables.
- In Table 1 and 2, for a given θ, the bias and MSE's are getting smaller as the sample size getting larger from 50 to 100 for both unblind and blind studies.
- The bias and MSE's of unblind estimators do not depend on θ when the sample size is fixed for both two-arms study and three-arms study.
- The unblind study always provides better results than the blind study in terms of bias and MSE in all cases.

Discussion

- Shih and Zhao (1997) provided a very nice strategy to execute the stratification in the clinical trial of a two-arm study for the dichotomized outcomes.
- The patients are first randomly assigned to either dummy stratum I or II and then randomly assign unequal number of patients into two treatment groups (A or B treatment) within each stratum.
- The proportion of patients of a particular group in one stratum is the opposite probability of the same group in the other stratum to preserve the equal number of the sample sizes of the two-arms study in total.
- The similar approach can be applied for the three-arms study to reach the balance observations among those arms.

Conclusion

- The proposed model is equivalent to the randomized response model proposed by Shih and Zhao (1997) if we have dichotomized outcomes in a two-arms study.
- For continuous response, the blind estimators and unblind estimators both are unbiased estimators for population means and variances, respectively.
- Eventually, the unblind estimators are uniformly minimum variance unbiased estimators of population means.
- The variances of the blind estimators become comparable to that of the unblinded estimators as the proportion of a particular treatment approaches 1 or 0.