Empirical Size and Power of a Hybrid Statistic for Matched and Unmatched Designs in Interim Analysis of Clinical Trials: A Simulation Study

Yunfei Wang and Dejian Lai*

*Correspondence: dejian.lai@uth.tmc.edu

Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, 1200 Pressler St., Room 1008, Houston TX 77030, USA.

Abstract

A hybrid test statistics was proposed in the literature to analyze matched studies with non-normally distributed outcomes. In this article, we investigated and compared the hybrid statistic with the meta-analysis t test under commonly used interim analysis settings in clinical trials. We estimated the empirical powers and the empirical type I errors among the 10,000 simulated datasets with different sample sizes, different effect sizes, different correlation coefficients for matched pairs, and different data distributions, respectively, in the interim and final analysis with 4 different group sequential methods. Results from our simulation study show that, compared to the meta-analysis t-test commonly used for data with normally distributed observations, the hybrid statistic almost keeps the power for data observed from normally distributed random variables and generally achieves greater power for log-normally, and multinomially distributed random variables with matched and unmatched subjects as well as with outliers. Powers rose with the increase in sample size, effect size, and correlation coefficient for the matched pairs. In addition, lower type I errors were observed by using the hybrid statistic in most of the cases studied, which indicates that this test is also conservative for data with outliers in the interim analysis of clinical trials.

Keywords: Clinical Trials, Empirical Power, Empirical Size, Interim Analysis

Introduction

In clinical trials, an interim analysis is commonly applied to obtain an evidence of a significant difference of the study interference. It could save patient resources and shorten the drug development and approval time if a trial is terminated with high confidence at an earlier time than planned. Group sequential designs are the most commonly used methods in the interim analysis.

Pocock (1977) and O’Brien and Fleming (1979) proposed their group sequential designs of interim analysis [1,2]. Both of their designs assume a fixed maximum number of interim analyses (K) with an equal scale space. But the Pocock design needs a higher number of a maximum sample size while the O’Brien-Fleming design requires a higher number of an expected sample size if the study continues to the final analysis [3].

Wang and Tsiatis (1987) proposed a family of two sided tests which provide boundaries of different shapes with different values of parameter delta [4]. The parameter could be varied to put more emphasis on a low maximum sample size or a low expected sample size in a group sequential analysis. Typically the delta ranges from 0 (O’Brien-Fleming design) to 0.5 (Pocock design) and a delta of 0.25 was selected as the middle point in our study. Haybittle (1971) and Peto et al. (1976) also suggested a simple form of sequential monitoring in which \( H_0 \) is rejected in the kth interim analysis \((k<K)\) if \(|Z_k| \geq 3\) [5,6].

Data collections in clinical trials are varied in types. Sometimes both matched and unmatched subjects could be included in atrial and it is important to properly include both of them in the interim analysis. The meta-analysis t-test with an inverse variance-weighted method is a commonly used to combine results of matched and unmatched subjects for data with normal distribution [7-11]. In practice, there are many more
different types of outcome data. For example, two eyes would be randomized to treatment or control groups as a matched pair if both eyes of a patient are sick in ophthalmology clinical trials. However, if some of patients have onesick and one normal eye and then only the sick eye would be randomized in the trial. In addition, the study outcome such as visual acuity is usually tested based on the distances and sizes of letters and recorded as 20/x in vision measurement, where x denotes as the distance the patient would see in 20 feet and these measurements are neither normally nor dichotomously distributed. Therefore, this study will include both matched and unmatched observations with neither normally distributed nor dichotomously distributed values.

A hybrid statistic is reported by investigators of Early Treatment of Retinopathy of Prematurity for data analysis [12,13]. They applied this statistic effectively to combine results of from matched and unmatched portions in their research. In our study, we performed a simulation study of the hybrid statistic to assess its empirical sizes and powers under the setting of interim analysis. We also compared our results of the hybrid statistic to those with the meta-analysis t-test.

Meta-analysis t-test and hybrid statistic
The meta-analysis t-test with the inverse variance-weighted method is usually applied in a two-treatment comparison for normally distributed data with paired and unpaired subjects. Briefly, let \( T_{\text{paired}} \) and \( T_{\text{unpaired}} \) be the observed mean difference and variance for paired and unpaired subjects, respectively, a weighted mean difference of the meta-analysis t-test at \( k \)th interim analysis is defined as

\[
T_{\text{combined}}^k = \frac{1}{\text{var}(T_{\text{paired}}^k)} T_{\text{paired}}^k + \frac{1}{\text{var}(T_{\text{unpaired}}^k)} T_{\text{unpaired}}^k
\]

and

\[
\text{var}(T_{\text{combined}}^k) = \frac{1}{\text{var}(T_{\text{paired}}^k)} + \frac{1}{\text{var}(T_{\text{unpaired}}^k)}
\]

In above formulations, the \( T \) statistic for the combined and uncombined data is the paired two sample T statistic to calculate \( T \) value at the \( k \)th interim analysis.

Asymptotically, assuming that \( T_{\text{combined}}^k \) has a normal distribution and the mean difference under the null hypothesis of \( H_0 \) is zero, then, we have the following normally distributed test statistic evaluated at the \( k \)th interim analysis:

\[
Z_k = \frac{T_{\text{combined}}^k - E(T_{\text{combined}}^k)}{\sqrt{\text{var}(T_{\text{combined}}^k)}} = \frac{T_{\text{combined}}^k - 0}{\sqrt{\text{var}(T_{\text{combined}}^k)}}
\]

For the hybrid statistic defined by Byun et al, we can also evaluate it at the \( k \)th interim analysis [13]. Briefly, at interim analysis \( k \), let \( \hat{\theta}_{\text{paired}}^k \) be the median of the Walsh averages associate with the Wilcoxon signed rank test defined as median \( \left\{ \frac{(z_i + z_j)}{2}, i \leq j = 1, \ldots, n_{1,k} \right\} \), where \( \frac{(z_i + z_j)}{2} \) is a Walsh average and \( z_i = y_i - x_i, i = 1, \ldots, n_{1,k} \) and \( \text{var} \left( \hat{\theta}_{\text{paired}}^k \right) \) its variance. Let \( \hat{\theta}_{\text{unpaired}}^k \) be the median of the cross differences of the unmatched two samples defined as median \( \left\{ y_j - x_j, j = 1, \ldots, n_{2,k}; l = 1, \ldots, n_{3,k} \right\} \), where \( n_{1,k} \) is the number of matched observations at stage \( k \) and \( n_{2,k} \) and \( n_{3,k} \) are the number of observations of the unmatched samples, respectively.

The median of these cross difference is the Hodges-Lehman estimate based on the Wilcoxon rank sum test, \( \text{var} \left( \hat{\theta}_{\text{unpaired}}^k \right) \) is its variance [16]. The hybrid test statistic evaluated at the \( k \)th interim analysis is calculated as following,

\[
T_{\text{combined}}^k = \frac{1}{\text{var}(\hat{\theta}_{\text{paired}}^k)} \frac{\hat{\theta}_{\text{paired}}^k}{\text{var}(\hat{\theta}_{\text{unpaired}}^k)} + \frac{1}{\text{var}(\hat{\theta}_{\text{unpaired}}^k)} \frac{\hat{\theta}_{\text{unpaired}}^k}{\text{var}(\hat{\theta}_{\text{paired}}^k)}
\]

\[
Z_k = \frac{T_{\text{combined}}^k - E(T_{\text{combined}}^k)}{\sqrt{\text{var}(T_{\text{combined}}^k)}} = \frac{T_{\text{combined}}^k - 0}{\sqrt{\text{var}(T_{\text{combined}}^k)}}
\]

The null hypothesis of \( H_0 \) in the \( k \)th interim analysis was rejected if \( \left| Z_k \right| \geq \hat{C}_v \) where \( \hat{C}_v \) equals to \( C_v(K, \alpha) \) for Pocock design, to \( C_{\text{HF}}(K, \alpha) \times \frac{K}{k} \) for O’Brien-Fleming design, to \( C_{\text{HP}}(K, \alpha, \Delta) \times \frac{(k/k)!^{1/2}}{k} \) for Wang-Tsiatis design, and to 3 (if \( k < K \)) or to \( C_{\text{HP}}(K, \alpha) \) (if \( k = K \)) for Haybittle-Peto design.

Design and Results of Simulation Studies of Empirical Sizes
Based on the \( Z \) values defined in last section, for the Pocock design, the nominal significance level is \( \frac{n}{2} = 2 \times [1 - \Phi(C_v(K, \alpha))] \) with the constant of \( C_v(K, \alpha) \) at \( K \) groups. For O’Brien-Fleming design, the nominal significance level is \( \alpha_{13} = 2 \times [1 - \Phi(C_v(K, \alpha) / \sqrt{K/k})] \) with the constant of \( C_v(K, \alpha) \) at \( K \) groups for O’Brien-Fleming design. The null hypothesis of \( H_0 \) will be rejected and the study would be terminated if \( \left| Z_k \right| \geq \hat{C}_v(K, \alpha) \) for Pocock design or \( \left| Z_{13} \right| \geq \hat{C}_v(K, \alpha) \) for O’Brien-Fleming design.
For a dataset with 20% outliers, 80% of subjects were from previous simulated dataset, and 20% of subjects were simulated from an effect size of $\mu _2 = \mu _3 = 10$ with a variance of 20 for normally distributed random variables. For log-normally distributed datasets, the effect size of $\ln(\mu _2) = \ln(\mu _3) = 2$ with a logarithm variance of 2 were used for 20% subjects with outliers.

Datasets observed from multinomially distributed random variables were also simulated but not showed in this paper due to the similar results with log-normally distributed random variables.

The $z$-values from different sample sizes at 5 significance levels with Pocock, O’Brien-Fleming, Wang-Tsiatis, and Haybittle-Peto designs were summarized in Table 1. Compared to the boundary of a standard normal distribution, all empirical boundaries estimated using both the meta-analysis t-test and the hybrid statistic for data with matched and unmatched subjects were far and wide from the standard Z-test when the sample size was equal to 20. However, they were gradually narrowed and closed to the standard boundary with the increase of the sample size. Compared to the meta-analysis t-test, empirical boundaries estimated by using the hybrid statistic were closer to the standard boundary line and especially when $n \geq 40$. (Figures 1 and 2).

Figure 1. O’Brien-Fleming boundaries estimated by hybrid statistic.

Results from above indicate that both the meta-analysis t-test and the hybrid statistic have large type I errors for normally distributed data with matched and unmatched subjects when the sample size is small ($n \leq 20$). But with the number increasing from 20 to 100, type I errors were decreased, especially by using the hybrid statistic.

Table 2 showed the type I errors among those four different designs with sample size 100. The largest empirical type I error from the Pocock design was observed in the first interim analysis ($k=1$). But type I error increased with the $k$ increasing for the O’Brien-Fleming design (Figure 3). The trend of empirical

![Diagram of O'Brien-Fleming boundaries estimated by hybrid statistic.](image)
Table 1. Simulated z-values at 5 different points of significance levels with different sample sizes estimated by using the meta-analysis t-test and the hybrid statistic.

| Significance Level | Standard z-value when n=20 | z-value when n=40 | z-value when n=60 | z-value when n=80 | z-value when n=100 |
|--------------------|---------------------------|------------------|------------------|------------------|-------------------|
|                    | t-test                    | H-test           | t-test           | H-test           | t-test            |
| Pocock             |                            |                  |                  |                  |                  |
| $a_1$ = 0.0158     | 2.413                     | 4.011            | 2.914            | 2.941            | 2.440             |
| $a_5$ = 0.0413     | 2.040                     | 3.000            | 2.188            | 2.379            | 1.967             |
| O'Brien-Fleming    |                            |                  |                  |                  |                  |
| $a_1 = 0.000005$   | 4.562                     | 20.876           | 16.918           | 8.660            | 6.933             |
| $a_2 = 0.0013$     | 3.226                     | 8.000            | 5.778            | 4.445            | 3.679             |
| $a_3 = 0.0084$     | 2.634                     | 4.809            | 3.465            | 3.329            | 2.750             |
| $a_4 = 0.0225$     | 2.281                     | 3.621            | 2.618            | 2.730            | 2.264             |
| $a_5 = 0.0413$     | 2.040                     | 3.000            | 2.188            | 2.379            | 1.967             |
| Wang-Tsiatis ($\Delta = 0.25$) |                       |                  |                  |                  |                  |
| $a_1 = 0.0014$     | 3.194                     | 7.784            | 5.659            | 4.406            | 3.635             |
| $a_2 = 0.0072$     | 2.686                     | 5.020            | 3.614            | 3.429            | 2.828             |
| $a_3 = 0.015$      | 2.427                     | 4.076            | 2.945            | 2.965            | 2.459             |
| $a_4 = 0.024$      | 2.259                     | 3.554            | 2.577            | 2.696            | 2.236             |
| $a_5 = 0.033$      | 2.136                     | 3.237            | 2.349            | 2.512            | 2.076             |
| Haybittle-Peto     |                            |                  |                  |                  |                  |
| $a_1, a_2 = 0.0027$| 3                        | 6.526            | 4.685            | 4.029            | 3.299             |
| $a_5 = 0.047$      | 1.99                      | 2.893            | 2.112            | 2.304            | 1.909             |

Figure 2. O'Brien-Fleming boundaries estimated by meta-analysis t-test.

Figure 3. Type I errors for normal distribution (N=100).

Type I errors for the Wang-Tsiatis design was between that of the Pocock and O'Brien-Fleming designs, and the trend for the Haybittle-Peto design was similar to that of the Pocock design in the interim analyses for $k_i\leq K$ and then to that of the O'Brien-Fleming design in the final analysis.

The results of empirical sizes for log-normally distributed random variables are tabulated in Table 3. Lower type I errors were observed by using the hybrid statistic, which indicates that this test is also conservative for data with log-normally distributed outcomes. Similar results were also observed when outliers were presented in the interim analysis of clinical trials.

Empirical type I errors were affected by sample sizes and correlation coefficient for O'Brien-Fleming design are shown in Figure 3. They were greater than standard type I error when $N=100$, but similar with a maximum sample size of 500 for data without outliers (Figure 4). For data with outliers, empirical type I errors were lower than standard normal type I error when $N=100$ and $k_i \geq 3$ (Figure 5), but were similar to standard one when $N=500$ (Figure 6). Meanwhile, lower type I errors, especially for data with outliers and with low sample sizes,
were observed by using the hybrid statistic in most cases of the study. These results suggest that using the hybrid statistic was less likely to commit type I error when there were outliers.

The largest empirical type I error of the Pocock design for data observed from log-normally distributed random variables with and without outliers was also observed in the first interim analysis. Type I errors increased with the increase of the sample size for the O’Brien-Fleming design, similar to the data observed from normally distributed random variables for the Wang-Tsiatis design and the Haybittle-Peto design. Results for data observed from log-normally distributed random variables suggest that both tests have lower type I errors, especially for data with outliers. Moreover, compared to results estimated by using the meta-analysis t-test, a lower type I error was also observed by using the hybrid statistic.

We also observed lower type I errors by using hybrid statistics than that of using meta t statistic for multinomially distributed random variables with and without outliers. These results are similar to those observed from log-normally distributed random variables.

### Table 2. Type I errors (α) for data observed from normally distributed random variables with matched and unmatched subjects (Maximum sample size=100).

| Group Sequential Method | Group (k) | Significance Level (α) | Sample Size (n) | Type I Error (α), d=0, σ^2=1 | Type I Error (α), d=0, σ^2=1 |
|------------------------|-----------|------------------------|-----------------|-------------------------------|-------------------------------|
|                        |           |                        |                 | t-test H-test t-test H-test    | t-test H-test t-test H-test    |
|                        |           |                        |                 | p=0.2                        | p=0.5                        |
|                        |           |                        |                 | p=0.8                        | p=0.8                        |
|                        |           |                        |                 | p=0.2                        | p=0.5                        |
|                        |           |                        |                 | p=0.8                        | p=0.8                        |

- **Pocock**
  - 1: 0.0158
  - 2: 0.0158
  - 3: 0.0158
  - 4: 0.0158
  - 5: 0.0158

- **O’Brien-Fleming**
  - 1: 0.000005
  - 2: 0.0013
  - 3: 0.0084
  - 4: 0.0225
  - 5: 0.0413

- **Wang-Tsiatis**
  - 1: 0.0014
  - 2: 0.0072
  - 3: 0.015
  - 4: 0.024
  - 5: 0.033

- **Haybittle-Peto**
  - 1: 0.0027
  - 2: 0.0027
  - 3: 0.0027
  - 4: 0.0027
  - 5: 0.0027

Note: 1. k_i=1, n_i=20 [10 matched subjects (5 pairs) and 10 unmatched subjects (5/5 in treatment/control regimens)], k_i = 2, n_i=2xn_i; k_i = 3, n_i = 3xn_i; k_i = 4, n_i=4xn_i; k_i = 5, n_i = 5xn_i.
2. d = (μ_1 - μ_2), where μ_1 = μ_0 = 0; d_2 = (μ_2 - μ_3), where μ_2 = μ_3 = 10; μ_0 = 3, σ_0^2 = 1, respectively. The correlation coefficients of 0.2, 0.5, and 0.8 were used as the data simulation.
3. t-test: meta-analysis t-test; H-test: Hybrid statistic.

Data observed from normally and log-normally distributed random variables were simulated with effects of μ_1-μ_2 = 0.2, 0.3, 0.6 (where μ_1=0,σ_0^2=1), and effects of ln(σ_0^2)=0, ln(μ_1)=0, ln(σ_0^2)=0, respectively. The correlation coefficients of 0.2, 0.5, and 0.8, respectively, for matched pairs were used as the data simulation.

In order to simulate datasets with 20% outliers, 80% of subjects were from above datasets and the other 20% were simulated with effect sizes of μ_1-μ_2 = 0.2, 0.3, 0.6 (where μ_2=10 and σ_2^2=20), for normally distributed random variables, or with an effect size of ln(σ_1^2) = ln(μ_1)=0, ln(σ_2^2)=0, respectively. The correlation coefficients of 0.2, 0.5, and 0.8 were used as the data simulation.

A dataset observed from multinomially distributed random variables was obtained from a dataset simulated based on log-normally distributed random variables, and continuous variables were transformed to discrete variables in the data transformation. For example, values ≤0.1 were transformed into 0, and values between 0.1-0.3 were changed into 0.2, etc.

Each dataset was simulated 10,000 times, and both the meta-analysis t-test and the hybrid statistic were applied for 4 interim analyses and a final analysis. Four group sequen-
Table 3. Type I errors (α) for data observed from normally distributed random variables with matched and unmatched subjects (Maximum sample size=100).

| Group Sequential Method | Group (k) | Significance Level (αi) | Sample Size (ni) | Type I Error (α), d=0, ln(σ²)=1 t-test | Type I Error (α), d=0, ln(σ²)=1 H-test |
|------------------------|-----------|------------------------|-----------------|---------------------------------------|---------------------------------------|
|                        |           |                        |                 | p=0.2 t-test                         | p=0.2 H-test                          |
|                        |           |                        |                 | p=0.5 t-test                         | p=0.5 H-test                          |
|                        |           |                        |                 | p=0.8 t-test                         | p=0.8 H-test                          |
| Pocock                 | 1         | 0.0158                 | 20              | 0.0467 0.0213 0.0447 0.0115          | 0.0439 0.0119 0.0244 0.0044           |
|                        | 2         | 0.0158                 | 40              | 0.0250 0.0068 0.0251 0.0070          | 0.0246 0.0057 0.0069 0.0028           |
|                        | 3         | 0.0158                 | 60              | 0.0218 0.0084 0.0209 0.0080          | 0.0201 0.0068 0.0051 0.0015           |
|                        | 4         | 0.0158                 | 80              | 0.0215 0.0074 0.0204 0.0066          | 0.0179 0.0047 0.0045 0.0014           |
|                        | 5         | 0.0158                 | 100             | 0.0200 0.0063 0.0180 0.0057          | 0.0181 0.0056 0.0043 0.0014           |
| O’Brien-Fleming        | 1         | 0.000005               | 20              | 0.0036 0.0013 0.0036 0.0015          | 0.0036 0.0013 0.0011 0.0001           |
|                        | 2         | 0.0013                 | 40              | 0.0033 0.0012 0.0033 0.0009          | 0.0030 0.0009 0.0011 0.0002           |
|                        | 3         | 0.0084                 | 60              | 0.0120 0.0051 0.0118 0.0046          | 0.0104 0.0035 0.0026 0.0005           |
|                        | 4         | 0.0225                 | 80              | 0.0306 0.0109 0.0295 0.0097          | 0.0278 0.0078 0.0077 0.0025           |
|                        | 5         | 0.0413                 | 100             | 0.0525 0.0208 0.0524 0.0207          | 0.0488 0.0182 0.0213 0.0067           |
| Wang-Tsiatis           | 1         | 0.0014                 | 20              | 0.0136 0.0043 0.0132 0.0040          | 0.0131 0.0035 0.0054 0.0011           |
|                        | 2         | 0.0072                 | 40              | 0.0125 0.0039 0.0128 0.0037          | 0.0126 0.0030 0.0039 0.0018           |
|                        | 3         | 0.015                 | 60              | 0.0209 0.0082 0.0204 0.0077          | 0.0192 0.0065 0.0046 0.0015           |
|                        | 4         | 0.024                 | 80              | 0.0327 0.0116 0.0315 0.0110          | 0.0295 0.0083 0.0085 0.0027           |
|                        | 5         | 0.033                 | 100             | 0.0409 0.0156 0.0405 0.0145          | 0.0382 0.0129 0.0149 0.0136           |
| Haybittle-Peto         | 1         | 0.0027                 | 20              | 0.0177 0.0050 0.0181 0.0050          | 0.0165 0.0046 0.0070 0.0015           |
|                        | 2         | 0.0027                 | 40              | 0.0062 0.0016 0.0062 0.0013          | 0.0059 0.0011 0.0020 0.0005           |
|                        | 3         | 0.0027                 | 60              | 0.0043 0.0019 0.0037 0.0017          | 0.0033 0.0011 0.0007 0.0004           |
|                        | 4         | 0.0027                 | 80              | 0.0033 0.0012 0.0029 0.0008          | 0.0027 0.0008 0.0006 0.0002           |
|                        | 5         | 0.047                 | 100             | 0.0585 0.0252 0.0572 0.0232          | 0.0555 0.0208 0.0262 0.0080           |

Note: 1. k1 = 1, n1 = 20 (10 matched subjects, 5 pairs) and 10 unmatched subjects (5/5 in treatment/control regimens), k2 = 2, n2 = 2xn1; k3 = 3, n3 = 3xn1; k4 = 4, n4 = 4xn1; k5 = 5, n5 = 5xn1
2. d = [ln(μ1) - ln(μ0)], where ln(μ0) = ln(μ1) = 0; d2 = [ln(μ3) - ln(μ2)], where ln(μ2) = ln(μ3) = 10
3. ρ: correlation coefficient for matched pairs
4. t-test: meta-analysis t-test
5. H-test: Hybrid statistic

Figure 4. Type I errors for normal distribution (N=500).

Figure 5. Type I errors for normal distribution with 20% outliers (N=100).

Note: Empirical powers for data observed from normally distributed random variables with matched and unmatched subjects were applied to estimate the empirical powers for each dataset.
Lai et al., *Journal of Medical Statistics and Informatics* 2020, [doi: 10.7243/2053-7662-8-5](http://www.hoajonline.com/journals/pdf/2053-7662-8-5.pdf)

Figure 6. Type I errors for normal distribution with 20% outliers (N=500).

Table 4. Empirical powers for data observed from normally distributed random variables with matched and unmatched subjects (Maximum sample size=300).

| Group Sequential Method | Group Sample Size (ni) | Power (%), di=0.3, σi=1 | Power (%), di=0.6, σi=1 |
|-------------------------|------------------------|--------------------------|--------------------------|
|                         |                        | t-test | H-test | t-test | H-test | t-test | H-test | t-test | H-test | t-test | H-test |
| Pocock                  | 1  60                 | 14.9   | 10.7   | 19.5   | 13.7   | 38.0   | 27.2   | 55.1   | 44.1   | 68.6   | 56.6   | 94.1   | 87.7   |
|                         | 2  120                | 26.7   | 22.8   | 35.9   | 30.7   | 68.0   | 61.5   | 85.9   | 81.3   | 94.2   | 91.3   | 99.9   | 99.8   |
|                         | 3  180                | 39.9   | 35.8   | 53.2   | 48.3   | 85.6   | 81.7   | 96.4   | 95.0   | 99.2   | 98.7   | 100    | 100    |
|                         | 4  240                | 52.5   | 48.5   | 66.8   | 62.3   | 94.1   | 92.5   | 99.4   | 99.0   | 99.9   | 99.9   | 100    | 100    |
|                         | 5  300                | 64.1   | 59.9   | 78.1   | 74.8   | 98.0   | 97.2   | 99.8   | 99.7   | 100    | 100    | 100    | 100    |
| O'Brien-Fleming         | 1  60                 | 0.4    | 0.2    | 0.6    | 0.3    | 2.5    | 1.2    | 4.9    | 2.3    | 9.3    | 4.6    | 37.0   | 21.1   |
|                         | 2  120                | 8.5    | 6.8    | 13.4   | 10.1   | 37.6   | 30.0   | 61.6   | 53.6   | 79.4   | 71.9   | 99.0   | 97.7   |
|                         | 3  180                | 31.5   | 27.9   | 44.6   | 39.6   | 80.2   | 75.1   | 94.5   | 92.2   | 98.6   | 97.9   | 100    | 100    |
|                         | 4  240                | 57.6   | 53.5   | 71.0   | 67.4   | 95.5   | 93.9   | 99.6   | 99.3   | 99.9   | 99.9   | 100    | 100    |
|                         | 5  300                | 76.5   | 73.8   | 87.2   | 84.9   | 99.1   | 98.8   | 100    | 99.9   | 100    | 100    | 100    | 100    |
| Wang-Tsiatis            | 1  60                 | 4.7    | 2.6    | 6.5    | 3.8    | 16.6   | 10.0   | 28.2   | 18.9   | 40.9   | 28.2   | 79.2   | 64.7   |
|                         | 2  120                | 12.5   | 15.5   | 27.1   | 22.3   | 58.2   | 50.2   | 79.3   | 73.3   | 90.8   | 86.1   | 99.8   | 99.5   |
|                         | 3  180                | 39.4   | 35.2   | 52.8   | 47.7   | 85.3   | 81.3   | 96.3   | 94.8   | 99.2   | 98.7   | 100    | 100    |
|                         | 4  240                | 58.5   | 54.2   | 71.7   | 68.2   | 95.6   | 94.2   | 99.6   | 99.4   | 100    | 99.9   | 100    | 100    |
|                         | 5  300                | 73.4   | 70.5   | 85.2   | 82.8   | 98.9   | 98.4   | 99.9   | 99.9   | 100    | 100    | 100    | 100    |
| Haybittle-Peto          | 1  60                 | 6.2    | 3.8    | 8.7    | 5.4    | 20.7   | 13.0   | 34.5   | 23.8   | 47.8   | 34.6   | 84.4   | 74.1   |
|                         | 2  120                | 12.1   | 9.8    | 18.2   | 14.2   | 45.6   | 38.1   | 70.0   | 62.6   | 84.5   | 78.7   | 99.5   | 98.7   |
|                         | 3  180                | 20.2   | 17.3   | 30.6   | 26.3   | 68.9   | 62.7   | 89.4   | 86.0   | 96.8   | 95.3   | 100    | 100    |
|                         | 4  240                | 30.7   | 27.3   | 44.7   | 40.0   | 84.8   | 80.2   | 97.0   | 95.2   | 99.6   | 99.2   | 100    | 100    |
|                         | 5  300                | 78.1   | 75.4   | 88.1   | 86.1   | 99.2   | 98.9   | 100    | 99.9   | 100    | 100    | 100    | 100    |

Note: 1. k1 = 1, n1 = 60 (30 matched subjects (15 pairs) and 30 unmatched subjects (15/15 in treatment/control regimens)). k2 = 2, n2 = 2xni; k3 = 3, n3 = 3xni; k4 = 4, n4 = 4xni; k5 = 5, n5 = 5xni
2. di = (μ1 - μ0), where μ0 = 0, μ1 = 0.3 and/or μ1 = 0.6
3. ρi: correlation coefficient for matched pairs
4. t-test: meta-analysis t-test
5. H-test: Hybrid statistic

Subjects and with maximum sample sizes of 100, 300, and 500 were simulated using both the meta-analysis t-test and the hybrid statistic. Partial results for maximum sample size of 300 without or with outliers are summarized in Tables 4 and 5. Overall, all empirical powers increased with the increasing of sample size, number of interim analyses (k), treatment effect size (di), and correlation coefficient (ρi) for matched pairs using both the meta-analysis t-test and the hybrid statistic. Due to the similar trend of powers for 4 different group sequential designs listed in the tables, empirical powers of the O’Brien-Fleming design were used for graphs.

Table 4 shows the empirical powers of the hybrid test statistic are very close to that of the meta-analysis t-test under normal without outliers (Table 4 and Figures 7 and 8). However, the hybrid test is much more powerful when there are outliers (Table 5 and Figures 9 and 10). For a closer look when the alternative is log-normal with and without outlier, we summarize the results in Tables 6 and 7.

From Table 5, we can see empirical powers were less than 60% for di=0.3 except for k=5 and ρ=0.8 (Figure 11). Empir-
Table 5. Empirical powers for data observed from normally distributed random variables with matched and unmatched subjects and with 20% outliers (Maximum sample size=300).

| Group Sequential Method | Group (k) | Sample Size (n) | Power (%), $d_i=0.3, \sigma^2=1$ (20% outlines with $d_i=0.3, \sigma^2=20$) | Power (%), $d_i=0.6, \sigma^2=1$ (20% outlines with $d_i=0.6, \sigma^2=20$) |
|-------------------------|-----------|----------------|-----------------------------------------------------------------|-----------------------------------------------------------------|
|                         |           |                | $p_i=0.2$ t-test H-test                                          | $p_i=0.5$ t-test H-test                                          |
|                         |           |                | $p_i=0.8$ t-test H-test                                          | $p_i=0.8$ t-test H-test                                          |
| Pocock                  | 1         | 60             | 3.2 2.4 3.3 2.7 3.6 3.9 11.3 11.0 14.1 13.1 22.6               | 5.2 5.6 6.5 7.9 8.4 8.8 13.4 14.2 19.0 18.2 30.2               |
|                         | 2         | 120            | 4.4 8.6 4.7 11.2 5.2 7.5 18.4 22.0 29.1 23.0 39.1             | 5.7 11.2 6.7 13.4 7.9 10.2 21.6 27.0 36.1 27.0 45.0             |
|                         | 3         | 180            | 5.2 15.2 5.7 20.5 6.2 10.7 18.4 22.0 29.1 23.0 39.1            | 6.4 11.2 6.7 13.4 7.9 10.2 21.6 27.0 36.1 27.0 45.0            |
|                         | 4         | 240            | 5.9 21.1 6.5 28.9 7.0 15.1 18.4 22.0 29.1 23.0 39.1            | 6.4 11.2 6.7 13.4 7.9 10.2 21.6 27.0 36.1 27.0 45.0            |
|                         | 5         | 300            | 7.1 27.8 7.6 37.6 8.5 20.3 26.0 31.0 39.5 29.0 53.0            | 6.4 11.2 6.7 13.4 7.9 10.2 21.6 27.0 36.1 27.0 45.0            |
| O’Brien-Fleming         | 1         | 60             | 0.0 0.0 0.0 0.0 0.0 0.1 2.0 2.1 3.1 3.0 5.0                 | 0.0 0.0 0.0 0.0 0.0 0.0 0.1 0.1 0.2 0.1 0.2              |
|                         | 2         | 120            | 0.5 1.5 0.5 1.9 0.5 3.6 15.6 23.6 23.6 6.0 5.0                | 0.5 1.5 0.5 1.9 0.5 3.6 15.6 23.6 23.6 6.0 5.0                |
|                         | 3         | 180            | 3.1 10.4 3.4 14.5 3.6 32.7 23.6 53.6 23.6 48.6 23.6          | 3.1 10.4 3.4 14.5 3.6 32.7 23.6 53.6 23.6 48.6 23.6          |
|                         | 4         | 240            | 7.7 25.3 8.1 33.6 9.1 62.7 26.9 84.8 29.2 73.0 29.2          | 7.7 25.3 8.1 33.6 9.1 62.7 26.9 84.8 29.2 73.0 29.2          |
|                         | 5         | 300            | 13.5 42.3 14.4 53.1 15.8 81.3 41.5 95.5 48.6 46.9 100       | 13.5 42.3 14.4 53.1 15.8 81.3 41.5 95.5 48.6 46.9 100       |
| Wang-Tsiatis            | 1         | 60             | 0.6 0.4 0.6 0.5 0.5 0.9 2.3 2.8 2.7 3.6 3.5                | 0.6 0.4 0.6 0.5 0.5 0.9 2.3 2.8 2.7 3.6 3.5                |
|                         | 2         | 120            | 2.4 4.9 2.5 6.6 2.7 14.6 9.0 31.8 10.2 43.2 12.0             | 2.4 4.9 2.5 6.6 2.7 14.6 9.0 31.8 10.2 43.2 12.0             |
|                         | 3         | 180            | 5.0 14.8 5.5 20.1 6.0 40.2 17.9 65.6 19.9 78.7 21.9           | 5.0 14.8 5.5 20.1 6.0 40.2 17.9 65.6 19.9 78.7 21.9           |
|                         | 4         | 240            | 8.0 25.9 8.6 34.6 9.5 63.6 27.7 85.5 29.9 93.3 32.3           | 8.0 25.9 8.6 34.6 9.5 63.6 27.7 85.5 29.9 93.3 32.3           |
|                         | 5         | 300            | 11.7 38.4 12.5 49.3 13.7 87.8 41.5 95.5 48.6 46.9 100      | 11.7 38.4 12.5 49.3 13.7 87.8 41.5 95.5 48.6 46.9 100      |
| Haybittle-Peto          | 1         | 60             | 0.9 0.6 0.8 0.7 0.7 1.2 3.2 4.0 3.7 5.2 5.0                | 0.9 0.6 0.8 0.7 0.7 1.2 3.2 4.0 3.7 5.2 5.0                |
|                         | 2         | 120            | 1.0 2.3 1.0 3.3 1.0 8.4 5.2 21.5 5.9 31.2 7.1                | 1.0 2.3 1.0 3.3 1.0 8.4 5.2 21.5 5.9 31.2 7.1                |
|                         | 3         | 180            | 1.4 5.1 1.4 7.6 1.6 20.5 7.4 42.6 8.4 58.6 9.8                | 1.4 5.1 1.4 7.6 1.6 20.5 7.4 42.6 8.4 58.6 9.8                |
|                         | 4         | 240            | 1.7 8.1 1.9 12.6 1.9 33.4 9.9 61.8 11.5 77.9 13.2           | 1.7 8.1 1.9 12.6 1.9 33.4 9.9 61.8 11.5 77.9 13.2           |
|                         | 5         | 300            | 14.7 44.4 15.5 55.1 16.8 82.6 43.4 95.9 46.0 98.8 48.8      | 14.7 44.4 15.5 55.1 16.8 82.6 43.4 95.9 46.0 98.8 48.8      |

Note: 1. $k_1 = 1, n_1 = 60$ (30 matched subjects (15 pairs) and 30 unmatched subjects (15/15 in treatment/control regimens))

2. $d_i = (\mu_1 - \mu_0)$, where $\mu_0 = 0, \mu_1 = 0.3$ and/or $\mu_1 = 0.6$; $d_i^2 = (\mu_3 - \mu_2)$, where $\mu_2 = 10, \mu_3 = 10.3$ and/or $\mu_3 = 10.6$

3. $\rho_i$: correlation coefficient for matched pairs

4. t-test: meta-analysis t-test

5. H-test: Hybrid statistic

Figure 7. Normal distribution (N=300, d=0.3).

Figure 8. Normal distribution (N=300, d=0.6).
Table 6. Empirical powers for data observed from log-normally distributed random variables with matched and unmatched subjects (Maximum sample size=300).

| Group Sequential Method | Group (k<sub>i</sub>) | Sample Size (n<sub>i</sub>) | Power (%) | di = 0.3, ln(σ²)=1 | Power (%) | di = 0.6, ln(σ²)=1 |
|------------------------|----------------------|-----------------------------|-----------|---------------------|-----------|---------------------|
|                        |                      |                             | ρ<sub>i</sub> = 0.2 | H-test | t-test | ρ<sub>i</sub> = 0.5 | H-test | t-test | ρ<sub>i</sub> = 0.8 | H-test | t-test | ρ<sub>i</sub> = 0.2 | H-test | t-test | ρ<sub>i</sub> = 0.5 | H-test | t-test | ρ<sub>i</sub> = 0.8 | H-test | t-test |
| Pocock                 | 1                    | 60                          | 9.2       | 3.7     | 11.4 | 4.6       | 20.1     | 8.1   | 34.9 | 18.3 | 43.1 | 23.2 | 66.9 | 40.3 |
|                        | 2                    | 120                         | 16.0      | 10.7   | 20.2 | 14.0     | 37.4     | 29.2  | 63.4 | 57.9 | 73.5 | 69.3 | 91.5 | 89.4 |
|                        | 3                    | 180                         | 23.1      | 20.7   | 29.4 | 28.4     | 53.4     | 54.7  | 81.3 | 85.0 | 88.6 | 92.2 | 97.6 | 98.8 |
|                        | 4                    | 240                         | 32.0      | 33.3   | 39.8 | 43.3     | 66.1     | 74.9  | 90.9 | 95.7 | 95.4 | 98.6 | 99.3 | 99.9 |
|                        | 5                    | 300                         | 39.8      | 44.7   | 49.0 | 57.1     | 75.8     | 87.3  | 95.9 | 99.0 | 98.3 | 99.7 | 99.7 | 100 |
| O'Brien-Fleming        | 1                    | 60                          | 0.1       | 0      | 0.1   | 0        | 0.1      | 0.1   | 0.5 | 0.2   | 0.9 | 0.3 | 3.6 | 0.8 |
|                        | 2                    | 120                         | 3.0       | 1.6    | 4.4  | 2.5      | 11.4     | 6.9   | 30.0 | 21.4 | 40.9 | 30.7 | 70.7 | 60.2 |
|                        | 3                    | 180                         | 16.8      | 14.1   | 21.9 | 19.8     | 44.4     | 43.8  | 74.7 | 77.9 | 84.1 | 87.7 | 96.2 | 97.9 |
|                        | 4                    | 240                         | 36.8      | 38.7   | 45.4 | 49.3     | 71.1     | 80.1  | 92.8 | 97.2 | 96.5 | 99.1 | 99.5 | 99.9 |
|                        | 5                    | 300                         | 54.7      | 61.7   | 63.7 | 73.4     | 85.8     | 94.4  | 98.3 | 99.6 | 99.2 | 100  | 99.9 | 100 |
| Wang-Tsiatis           | 1                    | 60                          | 1.5       | 0.6    | 2.1  | 0.6      | 5.0      | 1.3   | 11.3 | 4.3  | 15.8 | 6.0  | 33.9 | 13.7 |
|                        | 2                    | 120                         | 10.3      | 6.5    | 13.0 | 8.5      | 26.9     | 19.0  | 52.2 | 44.2 | 63.5 | 56.3 | 86.7 | 82.4 |
|                        | 3                    | 180                         | 22.6      | 20.3   | 28.9 | 27.7     | 52.8     | 53.9  | 80.9 | 84.6 | 88.4 | 92.0 | 97.5 | 98.8 |
|                        | 4                    | 240                         | 37.8      | 39.6   | 46.3 | 50.5     | 71.9     | 80.8  | 93.2 | 97.4 | 96.7 | 99.1 | 99.5 | 99.9 |
|                        | 5                    | 300                         | 50.7      | 57.4   | 59.9 | 69.6     | 83.6     | 93.0  | 97.8 | 99.5 | 99.0 | 99.9 | 99.9 | 100 |
| Haybittle-Peto         | 1                    | 60                          | 2.6       | 0.9    | 3.4  | 1.1      | 7.2      | 2.3   | 16.0 | 6.4  | 21.2 | 8.7  | 41.6 | 18.5 |
|                        | 2                    | 120                         | 5.2       | 3.0    | 7.3  | 4.3      | 17.1     | 10.7  | 39.0 | 29.6 | 50.5 | 40.6 | 78.1 | 70.4 |
|                        | 3                    | 180                         | 9.0       | 6.9    | 12.5 | 10.3     | 29.2     | 27.7  | 61.6 | 62.5 | 73.2 | 75.9 | 92.5 | 94.7 |
|                        | 4                    | 240                         | 13.7      | 13.3   | 19.2 | 20.0     | 42.8     | 47.7  | 77.5 | 84.2 | 86.4 | 92.8 | 97.4 | 99.3 |
|                        | 5                    | 300                         | 56.6      | 63.8   | 65.6 | 75.1     | 86.7     | 94.9  | 98.5 | 99.7 | 99.4 | 100  | 99.9 | 100 |

Note: 1. k<sub>i</sub> = 1, n<sub>i</sub> = 60 [30 matched subjects (15 pairs) and 30 unmatched subjects (15/15 in treatment/control regimens)]; k<sub>i</sub> = 2, n<sub>i</sub> = 2xn<sub>i</sub>; k<sub>i</sub> = 3, n<sub>i</sub> = 3xn<sub>i</sub>; k<sub>i</sub> = 4, n<sub>i</sub> = 4xn<sub>i</sub>; k<sub>i</sub> = 5, n<sub>i</sub> = 5xn<sub>i</sub> 1. 2. di = (µ<sub>1</sub> - µ<sub>0</sub>), where µ<sub>0</sub> = 0, µ<sub>1</sub> = 0.3 and/or µ<sub>1</sub> = 0.6 3. ρ<sub>i</sub>: correlation coefficient for matched pairs 4. t-test: meta-analysis t-test 5. H-test: Hybrid statistic

Empirical powers were increased for d=0.6, and all were above 60% for k=4 and 80% for final analysis (k=5) (Figure 12). Empirical powers increased as the increasing of sample size. At N=300, it reached 80% when d=0.3, ρ<sub>i</sub>=0.8 at k=3, and d=0.6, ρ<sub>i</sub>=0.8 at k=2 or at k=3 for each ρ<sub>i</sub> for data without outliers. When there are outliers in the data, results in Table 7 clearly indicated that the hybrid test statistic is much powerful (Figures 13 and 14). For example, as shown in Table 7, the power for O’Brien-Fleming
Table 7. Empirical powers for data observed from log-normally distributed random variables with matched and unmatched subjects and with 20% outliers (Maximum sample size=300).

| Group Sequential Method | Group Sample Size (k, n) | Power (%), $d_i = 0.3, \ln(\sigma_i)=1$ (20% outlines with $d_j = 0.3, \ln(\sigma_j)=2$) | Power (%), $d_i = 0.6, \ln(\sigma_i)=1$ (20% outlines with $d_j = 0.6, \ln(\sigma_j)=2$) |
|-------------------------|-------------------------|-------------------------------------------------|-------------------------------------------------|
|                         |                         | $\rho = 0.2$ t-test H-test                         | $\rho = 0.2$ t-test H-test                         |
|                         |                         | $\rho = 0.5$ t-test H-test                         | $\rho = 0.5$ t-test H-test                         |
|                         |                         | $\rho = 0.8$ t-test H-test                         | $\rho = 0.8$ t-test H-test                         |
|                         |                         | $\rho = 0.2$ t-test H-test                         | $\rho = 0.2$ t-test H-test                         |
|                         |                         | $\rho = 0.5$ t-test H-test                         | $\rho = 0.5$ t-test H-test                         |
|                         |                         | $\rho = 0.8$ t-test H-test                         | $\rho = 0.8$ t-test H-test                         |
| Pocock                  | 1 60                    | 1.5 0.9 1.6 1.0 2.1 1.1 5 4.3 5.9 5.2 8.0 7.9 | 10.7243/2053-7662-8-5 |
|                         | 2 120                   | 1.8 2.9 2.0 3.5 2.4 6.4 6.6 22.5 8.0 28.5 10.6 46.4 |
|                         | 3 180                   | 1.9 7.4 2.2 9.3 2.7 18.1 8.9 52.9 10.6 62.7 13.4 81.4 |
|                         | 4 240                   | 2.3 13.7 2.6 18.1 0.3 34.5 11.5 75.7 13.3 84.2 16.3 95.0 |
|                         | 5 300                   | 3.1 21.2 3.5 28.0 4.1 51.1 14.1 89.4 16.4 94.4 19.8 98.8 |
| O’Brien-Fleming         | 1 60                    | 0 0 0 0 0 0 0 0 0 0 0 0 0                  | 0 0 0 0 0 0 0 0 0 0 0 0 0 0                  |
|                         | 2 120                   | 0.1 0.2 0.1 0.1 0.3 0.1 0.6 0.7 3.8 0.9 5.4 1.5 12.0 |
|                         | 3 180                   | 1.0 4.1 1.1 5.5 1.3 11.5 5.2 40.3 6.3 50.7 8.6 72.6 |
|                         | 4 240                   | 3.5 18.0 4.0 22.8 4.8 41.5 14.9 80.9 17.2 87.9 20.6 96.6 |
|                         | 5 300                   | 8.5 37.8 9.2 46.3 10.4 70.4 27.2 95.9 30.3 98.2 34.6 99.7 |
| Wang-Tsiatis            | 1 60                    | 0.1 0.1 0.1 0.1 0.3 0.2 0.7 0.6 0.9 0.7 1.5 1.3     | 0.1 0.1 0.1 0.1 0.3 0.2 0.7 0.6 0.9 0.7 1.5 1.3 |
|                         | 2 120                   | 0.7 1.3 0.8 1.5 1.0 2.9 3.5 13.7 4.3 17.8 5.9 31.9 |
|                         | 3 180                   | 1.8 7.2 2.1 9.1 2.6 17.6 8.7 52.3 10.3 62.1 13.1 80.9 |
|                         | 4 240                   | 3.7 18.6 4.2 23.6 5.2 42.7 15.7 81.9 17.7 88.5 21.4 96.8 |
|                         | 5 300                   | 6.7 33.0 7.4 41.2 8.5 65.8 23.4 94.6 26.3 97.4 30.4 99.6 |
| Haybittle-Peto          | 1 60                    | 0.2 0.2 0.3 0.2 0.3 0.4 0.2 0.1 1.0 1.0 1.3 1.2 2.4 2.1 |
|                         | 2 120                   | 0.2 0.5 0.3 0.6 0.3 1.2 1.4 6.8 1.8 9.4 2.8 18.9 |
|                         | 3 180                   | 0.3 1.2 0.3 1.9 0.3 4.3 1.6 22.8 2.3 31.0 3.5 54.4 |
|                         | 4 240                   | 0.3 2.9 0.4 4.4 0.6 11.2 2.5 46.2 3.3 58.4 4.8 81.2 |
|                         | 5 300                   | 9.6 40.2 10.4 48.8 11.9 72.7 29.4 96.6 32.7 98.5 37.1 99.8 |

Note: 1. $k_1 = 1, n_1 = 60$ (30 matched subjects (15 pairs) and 30 unmatched subjects (15/15 in treatment/control regimens)); $k_2 = 2, n_2 = 2x n_1$; $k_3 = 3, n_3 = 3x n_1$; $k_4 = 4, n_4 = 4x n_1$; $k_5 = 5, n_5 = 5x n_1$
2. $d_i = [\ln(\mu_i) – \ln(\mu_j)]$, where $\ln(\mu_0) = 0$, $\ln(\mu_1) = 0.3$ and/or $\ln(\mu_3) = 0.6$; $d_i^2 = [\ln(\mu_i) – \ln(\mu_j)]$, where $\ln(\mu_0) = 10$, $\ln(\mu_1) = 10.3$ and/or $\ln(\mu_3) = 10.6$
3. $\rho$: correlation coefficient for matched pairs
4. t-test: meta-analysis t-test
5. H-test: Hybrid statistic

Figure 11. Log-normal distribution (N=300, $d=0.3$).

Figure 12. Log-normal distribution (N=300, $d=0.6$).

design with 20% outlines with $d_j = 0.6, \ln(\sigma_j)=2$, the empirical power of the hybrid statistic achieved 95.9% whereas the meta t statistic reached only 27.2% at the final stage. In fact, almost all the empirical powers were larger than 90% at the final stage for the hybrid statistics, but the empirical power from meta t was mostly less than 30%.
Figure 13. Log-normal dis. 20% outliers (N=300, d=0.3).

Figure 14. Log-normal dis. 20% outliers (N=300, d=0.6).

Similar results were observed for multinomial outcomes as in for the log-normal distribution for the matched and unmatched subjects and with maximum sample sizes of 100, 300, and 500 using both the meta-analysis t-test and the hybrid statistic. These results are omitted here. The purpose of simulating multinomial outcomes was to validate that the hybrid statistic is not only powerful for data observed from continuous random variables with outliers but also for data observed from discrete variables with outliers in the interim analysis of clinical trials.

Discussions and Concluding Remarks
Meta-analysis t-test can be used for normally distributed random variables with matched and unmatched subjects. Byun, et al. reported a hybrid statistic for data from neither normally nor dichotomously distributed random variables under clinical trials without interim analysis [13]. In our study, we conducted simulation studies of an interim analysis of clinical trial with different sample size and with matched and unmatched subjects to compare the type I errors and powers using meta-analysis t-test and hybrid statistic. Our results showed empirical type I error rates were relatively large when sample sizes small for matched and unmatched subjects in the interim analysis by using both the meta-analysis t-test and the hybrid statistic. However, compared to the meta-analysis t-test, lower type I error rate was observed for the hybrid test statistic under the normal outcome setting. It was observed that with increase in sample sizes, empirical type I errors were gradually approached those for standard analysis (that is, clinical trials conducted without interim analysis). In general, hybrid test statistic had a lower type I error rate comparing to the meta-analysis t test.

The meta-analysis t-test is more powerful for ideal data observed from normally distributed outcomes with matched and unmatched subjects. However, the hybrid test is more robust and powerful than that of the meta-analysis when the outcome contained outliers in the interim analysis of clinical trials. All empirical powers rose with the increase in sample size, effect size, and correlation coefficient for matched pairs. Our results from the simulation study strongly support that the hybrid statistic has more power and low type I error for data with matched and unmatched subjects and with non-normal distribution, especially for those data with outliers.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions

| Authors’ contributions       | YW | DL |
|-----------------------------|----|----|
| Research concept and design | ✓  | ✓  |
| Collection and/or assembly of data | ✓  | -- |
| Data analysis and interpretation | ✓  | -- |
| Writing the article         | ✓  | -- |
| Critical revision of the article | ✓  | ✓  |
| Final approval of article   | ✓  | ✓  |
| Statistical analysis        | ✓  | -- |

Acknowledgement
This work was partially supported by Cancer Prevention Research Institute of Texas (RP170668).

Publication history
EIC: Jimmy Efird, East Carolina University, USA.
Received: 15-Mar-2020 Final Revised: 18-May-2020
Accepted: 12-Jun-2020 Published: 25-Jun-2020

References
1. Bellissant E, Benichou J and Chastang C. A microcomputer program for the design and analysis of phase II cancer clinical trials with two group sequential methods, the sequential probability ratio test, and the triangular test. *Comput Biomed Res*. 1994; 27:13-26. | Article | PubMed
2. O’Brien PC and Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979; 35:549-56. | PubMed
3. Jennison C and Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC. 2000.

4. Wang SK and Tsiatis AA. Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics*. 1987; 43:193-9. | PubMed

5. Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. *Br J Radiol*. 1971; 44:793-7. | Article | PubMed

6. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J and Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer*. 1976; 34:585-612. | Article | PubMed Abstract | PubMed FullText

7. Shadish WR and Haddock CK. Combining estimates of effect size. In: Cooper H, Hedges LV (editors). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation. 1994; 261-284.

8. Rosenthal R. Combining results of independent studies. *Psychological Bulletin*. 1978; 85:185-193.

9. Fleiss JL. *Measures of effect size for categorical data*. In: Cooper H, Hedges LV (editors). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation. 1994; 245-260.

10. Shutton AJ, Abrams KR, Jones DR, Sheldon TA and Song F. *Methods for Meta-Analysis in Medical Research*. New York: Wiley. 2000.

11. Rosenthal R. Combining results of independent studies. *Psychological Bulletin*. 1978; 85:185-193.

12. Early Treatment for Retinopathy of Prematurity Cooperative Group. Final Visual Acuity Results in the Early Treatment for Retinopathy of Prematurity Study. *Archives of Ophthalmology*. 2010; 128:1684-1701.

13. Byun J, Lai D, Luo S, Risser J, Tung B and Hardy RJ. A hybrid method in combining treatment effects from matched and unmatched studies. *Stat Med*. 2013; 32:4924-37. | Article | PubMed Abstract | PubMed FullText

14. Walsh JE. Some significance tests for the median which are valid under very general conditions. *Annals of Mathematical Statistics*. 1949; 20:64-81.

15. Wilcoxin F. Probability tables for individual comparisons by ranking methods. *Biometrics*. 1947; 3:119-22. | PubMed

16. Hodges JL and Lehmann EL. Estimates of location based on rank test. *Annals of Mathematical Statistics*. 1963; 34:598-611.