Sequential Monitoring of Conditional Randomization Tests

William F. Rosenberger and Victoria Plamadeala

Department of Statistics
George Mason University
Fairfax, VA
and
Precision Therapeutics
Pittsburgh, PA

IWSM&A 2012, Rouen, France
June 2012
1 Background on Randomization Tests

2 Monte Carlo Methods for Computing Conditional Tests

3 Sequential Randomization Tests

4 Conclusions
SEQUENTIAL MONITORING WITH CONDITIONAL RANDOMIZATION TESTS

BY VICTORIA PLAMADEALA AND WILLIAM F. ROSENBERGER

Precision Therapeutics and George Mason University

Sequential monitoring in clinical trials is often employed to allow for early stopping and other interim decisions, while maintaining the type I error rate. However, sequential monitoring is typically described only in the context of a population model. We describe a computational method to implement sequential monitoring in a randomization-based context. In particular, we discuss a new technique for the computation of approximate conditional tests following restricted randomization procedures and then apply this technique to approximate the joint distribution of sequentially computed conditional randomization tests. We also describe the computation of a randomization-based analog of the information fraction. We apply these techniques to a restricted randomization procedure, Efron’s [Biometrika 58 (1971) 403–417] biased coin design. These techniques require derivation of certain conditional probabilities and conditional covariances of the randomization procedure. We employ combinatoric techniques to derive these for the biased coin design.

1. Introduction. Sequential monitoring refers to analyzing data periodically during the course of a clinical trial, with the purpose of detecting early evidence in support of or against a hypothesis. A desirable feature of such a monitoring plan would be flexible inspections of the data that can occur at arbitrary time points. At the same time, sequentially tested hypotheses must maintain the overall probability of type I error at the prespecified level, since repeated testing is known to inflate it. The Lan and DeMets (1983) error spending approach for sequential monitoring allows this. The approach makes use of a type I error spending function, which depends on the amount of “statistical information” available at the time of the interim inspection. In the context of sequential monitoring, the statistical information is a measure of how far a trial has progressed. Under a population model, the amount of interim information—the information fraction—is defined as the proportion of Fisher’s information observed thus far in the trial. The type I error spending function ration the amount of type I error that may be spent at each look commensurate to the information fraction. The critical value associated with the allowable probability of type I error at a certain interim look is obtained and

Received July 2011; revised November 2011.

1Supported by NSF Grant DMS-09-04253 under the 2009 American Reinvestment and Recovery Act.


Key words and phrases. Biased coin design, conditional reference set, random walk, restricted randomization, sequential analysis.
Everyone knows that clinical trials do not follow a population model with random sampling. In fact the only random mechanism is the randomization itself.

For this reason, the FDA often requires a randomization-based inference analysis by applying a randomization test (which they often call “re-randomization test”) to the primary outcome analysis. Such analyses are typically done by using Monte Carlo simulation to regenerate randomization procedures and then compute the test statistics $p$-value as the proportion of times the simulated statistic exceeds the observed.

However, this is an “unconditional” test, which may not be desirable. It also ignores the possibility that there was sequential monitoring in the trial. In this talk we describe how to compute conditional tests and incorporate them in the sequential monitoring plan.
$n$ subjects are randomized

Group A

$\ n_1 \$

Group B

$n - n_1$
$n$ subjects are randomized

$n_1$

$n - n_1$

Group A

Data due to group A

observe some difference

Group B

Data due to group B

William F. Rosenberger and Victoria Plamadeala

Randomization Tests
Test $H_0: A$ and $B$ are not different.
$n$ subjects are randomized

$n_1$  $n - n_1$

Group A  Data due to group A

Data due to one group C

Group B  Data due to group B

Test $H_0: A$ and $B$ are not different.
Any observed difference is entirely due to the randomization

Test $H_0: A$ and $B$ are not different.
The $p$–value of the randomization test depends on the randomization procedure, $T_n = (T_1, ..., T_n)'$, used:

- **complete randomization:**

  $$
  \phi_{j+1} = P(T_{j+1} = 1) = P(T_{j+1} = 0) = 1/2, \ j = 0, ..., n - 1.\)

Typically we use some form of restricted randomization to promote balance between treatment assignments both at the end and throughout the middle of the trial:

- **permuted block design**
- **biased coin design** and its generalizations and extensions
Restricted randomization procedures:
Efron’s (1971) biased coin design (BCD)

\[ \phi_{j+1}(m_j) = P(T_{j+1} = 1 | \sum_{i=1}^{j} T_i = m_j) = \begin{cases} 
1/2, & \text{if } m_j = j/2 \\
p, & \text{if } m_j < j/2 \\
1 - p, & \text{if } m_j > j/2,
\end{cases} \]

$j = 0, \ldots, n-1$, $m_j$ is the number assigned to treatment 1 when $j$ subjects have already been randomized.
The $p$–value of the randomization test depends on

- the type of reference set used:
  - $Ω_u$ – unconditional set with cardinality $2^n$;
  - $Ω_c$ – conditional set with cardinality $\binom{n}{n_1}$, where $n_1$ is the final number assigned to treatment 1.

- the metric of the treatment effect.
Example

- Suppose BCD\((p = 3/4)\) was used. The observed responses are \(\{2.3, 1.9, 2.2, 2.1, 2.0\}\) on treatments \(\{1, 0, 0, 1, 0\}\).

- The analog of the Wilcoxon rank sum test gives:

\[
\sum_{i=1}^{5} a_{jn} T_i = (2 - 2 1 0 - 1)(1 0 0 1 0)' = 2.
\]

- Choose the type of reference set:
  - Conditional: use only \(\binom{5}{2}\) sequences that satisfy \(n_1 = 2\).
  - Obtain the randomization distribution by computing all possible inner products on the reference set.
In the earlier example:

The unconditional set $\Omega_u$ contains $2^5$ sequences:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0,0,0,0,0</td>
<td>0,1,0,0,0</td>
<td>1,0,0,0,0</td>
<td>1,1,0,0,0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,0,0,0,1</td>
<td>0,1,0,0,1</td>
<td>1,0,0,0,1</td>
<td>1,1,0,0,1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,0,0,1,0</td>
<td>0,1,0,1,0</td>
<td>1,0,0,1,0</td>
<td>1,1,0,1,0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,0,0,1,1</td>
<td>0,1,0,1,1</td>
<td>1,0,0,1,1</td>
<td>1,1,0,1,1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,0,1,0,0</td>
<td>0,1,1,0,0</td>
<td>1,0,1,0,0</td>
<td>1,1,1,0,0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,0,1,0,1</td>
<td>0,1,1,0,1</td>
<td>1,0,1,0,1</td>
<td>1,1,1,0,1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,0,1,1,0</td>
<td>0,1,1,1,0</td>
<td>1,0,1,1,0</td>
<td>1,1,1,1,0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,0,1,1,1</td>
<td>0,1,1,1,1</td>
<td>1,0,1,1,1</td>
<td>1,1,1,1,1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In the earlier example:

The conditional set $\Omega_c$ that satisfies $\{\sum_{i=1}^{5} T_i = N_1(5) = 2\}$:

- $0, 0, 0, 1, 1$
- $0, 0, 0, 1, 0$
- $0, 0, 1, 0, 1$
- $0, 0, 1, 1, 0$
- $1, 1, 0, 0, 0$
- $1, 0, 0, 0, 1$
- $1, 0, 0, 1, 0$
- $1, 0, 1, 0, 0$
In the earlier example:

The conditional distribution of \( S(T) = (2, -2, 1, 0, -1)T \) under the BCD\((p)\):

\[
\begin{bmatrix}
0 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
2 \\
-2 \\
1 \\
0 \\
-1
\end{bmatrix}
= \begin{bmatrix}
-1 \\
0 \\
1 \\
-3 \\
-2 \\
-1 \\
1 \\
2 \\
3 \\
0
\end{bmatrix}
\]

with prob. \[
\begin{bmatrix}
0.5(1 - p)^2 p^2 / C \\
0.5(1 - p)^2 p^2 / C \\
0.5^2(1 - p)p^2 / C \\
0.5^2(1 - p)p^2 / C \\
0.5^3 p^2 / C \\
0.5^3 p^2 / C \\
0.5^2(1 - p) p^2 / C \\
0.5^3 p^2 / C \\
0.5^3 p^2 / C
\end{bmatrix}
\]

where \( C = p^2(2.5 - 3p + p^2) \).

William F. Rosenberger and Victoria Plamadeala  Randomization Tests
Conditional tests under the BCD\((p)\): computational techniques

- Exact tests (Hollander and Peña, 1988; Mehta, Patel and Wei, 1988a) for modern clinical trials are computationally infeasible.

- No asymptotic approximation exists for the randomization test under Efron’s BCD.
Conditional Tests Comparing Two Groups
Method 1: Sampling from $\Omega_u$

$\Omega_u$, sample $K$ sequences

$\Omega_c$, retain $N_c$ sequences that satisfy $N_1(n) = n_1$
Method 1: Sampling from $\Omega_u$

- Use the randomization rule $\phi_{j+1}$ to generate these sequences.

- The main problem is deciding on $K$, the overall MC sample size.
Method 1: Sampling from $\Omega_u$

- $K \sim NB(P(N_1(n) = n_1), N_c)$

- One can compute $P(N_1(n) = n_1)$ exactly under the BCD($p$) (Markaryan and Rosenberger, 2010).
Method 1: Sampling from $\Omega_u$

Table: Approximate 95th percentile of $K$ for various $n$, $n_1$, $N_c = 2500$; BCD (2/3).

<table>
<thead>
<tr>
<th>$n$</th>
<th>$n_1 = 0.45n$</th>
<th>$n_1 = 0.48n$</th>
<th>$n_1 = 0.50n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>3,531,344</td>
<td>55,060</td>
<td>5,117</td>
</tr>
<tr>
<td>200</td>
<td>3,611,280,266</td>
<td>881,557</td>
<td>5,117</td>
</tr>
<tr>
<td>500</td>
<td>$3,877,310 \times 10^{12}$</td>
<td>3,611,026,232</td>
<td>5,117</td>
</tr>
</tbody>
</table>
Method 2: Sampling directly from $\Omega_c$

- Perhaps some computation efficiency could be gained if the $N_c$ sequences were sampled directly from $\Omega_c$.
- This requires adjusting the randomization rules $\phi_{j+1}$ to generate sequences directly from $\Omega_c$. 
Method 2: Sampling directly from $\Omega_c$

The appropriate sampling rule is

$$p_{j+1} = P \left( T_{j+1} = 1 | \sum_{i=1}^{j} T_i = m_j, \sum_{i=1}^{n} T_i = n_1 \right)$$

$$= \begin{cases} 
\phi_{j+1}(m_j) \frac{P(N_1(n) = n_1 | N_1(j+1) = m_j + 1)}{P(N_1(n) = n_1 | N_1(j) = m_j)}, & 1 \leq j \leq n - 1, \\
1/2 \frac{P(N_1(n) = n_1 | T_{j+1} = 1)}{P(N_1(n) = n_1)}, & j = 0.
\end{cases}$$
Method 2: Sampling directly from $\Omega_c$

Closed form solutions for $p_{j+1}$ were derived for Efron’s BCD($p$) using combinatorics.
1. Use $p_{j+1}$ to sample $N_c$ sequences directly from $\Omega_c$.

2. Evaluate the statistic for each sequence.

3. Use the sample proportion, $\hat{p}$, as the estimator of $p_c$, the conditional test $p$-value.

4. \[ \text{MSE}(\hat{p}) = p_c(1 - p_c)/N_c \leq 1/4N_c \leq \epsilon. \] For $\epsilon = 0.0001$, $N_c \geq 2500$.

5. For higher precision, one can use $P(|\hat{p} - p_c| \leq 0.1p_c) = 0.99$. If $p_c = 0.04$, $N_c \approx 16,000$, by the CLT.
Sequential Randomization Tests Comparing Two Groups
Suppose an $\alpha$ level test is planned for some clinical trial comparing two treatments after $n$ subjects have responded.

But want to allow for $L - 1$ interim inspections of the data after $1 \leq r_1 < r_2 < \ldots < r_{L-1} < r_L = n$ patients have responded.
A Type I error spending function (Lan and DeMets, 1983), a nondecreasing function

\[ \alpha^* : [0, 1] \mapsto [0, \alpha]. \]

Information fraction: \( 0 < t_1 < t_2 < \ldots < t_{L-1} < t_L = 1. \)
A typical sequential monitoring plan with $L$ interim looks:

- $P(V_{r1} > d_1) = \alpha^*(t_1)$
- $P(V_{r1} \leq d_1, V_{r2} > d_2) = \alpha^*(t_2) - \alpha^*(t_1)$
- $P(V_{r1} \leq d_1, V_{r2} \leq d_2, V_{r3} > d_3) = \alpha^*(t_3) - \alpha^*(t_2)$
- $\vdots$
- $P(V_{r1} \leq d_1, V_{r2} \leq d_2, \ldots, V_{rL} > d_L) = \alpha - \alpha^*(t_{L-1})$. 

William F. Rosenberger and Victoria Plamadeala
The randomization version of the sequential testing plan:

\[
\begin{align*}
P(V_{r_1} > d_1 | N_1(r_1) = n_{11}) &= \alpha^*(t_1), \\
P(V_{r_1} \leq d_1, V_{r_2} > d_2 | N_1(r_1) = n_{11}, N_1(r_2) = n_{12}) &= \alpha^*(t_2) - \alpha^*(t_1), \\
P(V_{r_1} \leq d_1, V_{r_2} \leq d_2, V_{r_3} > d_3 | \cap_{j=1}^{3} N_1(r_j) = n_{1j}) &= \alpha^*(t_3) - \alpha^*(t_2), \\
&\vdots \\
P(V_{r_1} \leq d_1, ..., V_{r_{L-1}} \leq d_{L-1}, V_n > d_L | \cap_{j=1}^{L} N_1(r_j) = n_{1j}) &= \alpha - \alpha^*(t_{L-1}).
\end{align*}
\]

(Zhang and Rosenberger, 2008)
Equivalent to

\[
\begin{align*}
P \left( V_{r_1} > d_1 \mid N_1(r_1) = n_{11} \right) &= \alpha^* (t_1), \\
P \left( V_{r_2} > d_2 \mid V_{r_1} \leq d_1, \bigcap_{j=1}^{2} \{ N_1(r_j) = n_{1j} \} \right) &= \frac{\alpha^* (t_2) - \alpha^* (t_1)}{1 - \alpha^* (t_1)}, \\
P \left( V_{r_3} > d_3 \mid \bigcap_{j=1}^{2} \{ V_{r_j} \leq d_j \}, \bigcap_{j=1}^{3} \{ N_1(r_j) = n_{1j} \} \right) &= \frac{\alpha^* (t_3) - \alpha^* (t_2)}{1 - \alpha^* (t_2)}, \\
&\vdots \\
P \left( V_{n} > d_L \mid \bigcap_{j=1}^{L-1} \{ V_{r_j} \leq d_j \}, \bigcap_{j=1}^{L} \{ N_1(r_j) = n_{1j} \} \right) &= \frac{\alpha - \alpha^* (t_{L-1})}{1 - \alpha^* (t_{L-1})}.
\end{align*}
\]
1. At each look $l$, sample sequences that satisfy
   \[ \bigcap_{j=1}^{l} \{ N_1(r_j) = n_{1j} \} \].

2. Retain those that satisfy \[ \bigcap_{j=1}^{l-1} \{ V_{r_j} \leq d_j \} \].

3. Evaluate the current stage statistic $V_{r_l}$ for each retained sequence.

4. The resulting values approximate the required distribution.

5. Use this distribution to estimate $d_l$. 
The main problem is to sample sequences that satisfy
\[ \bigcap_{j=1}^{l} \{ N_1(r_j) = n_{1j} \}. \]
The randomization–based information fraction is defined as the ratio of the conditional variances (Rosenberger and Lachin, 2002):

\[ t_I = \frac{a'_{r_I} \Sigma_{|r_I} a_{r_I}}{a'_{n} \Sigma_{|n} a_{n}}. \]

where \( \Sigma_{|r_I} = \text{Var}(T^{(r_I)}|N_1(r_1) = n_{11}, \ldots, N_1(r_I) = n_{1I}) \), for which we have derived exact computational expressions under the BCD\( (p) \) and complete randomization.
We generate an allocation sequence under the BCD(3/4) and observations from two normal distributions to simulate the sequential analysis of the randomization test (3 inspections).

At each inspection the goal is to estimate the information by extrapolating the unknown data using three methods:

1. Sampling from two different normal distributions with estimated unknown parameters.
2. Sampling from one normal distribution with estimated unknown parameters.
3. Resampling (with replacement) the unknown data by groups.
Table: Mean (SD) of simulated $\alpha$ for an $\alpha = 0.05$ upper tail sequential test over a Monte Carlo sample size of 1000, $N_c = 2500$, interpolating the unknown observations by sampling with replacement.

<table>
<thead>
<tr>
<th>Look</th>
<th>$r_l$</th>
<th>$n_{1l}$</th>
<th>$t_l$</th>
<th>$\alpha_l^*$</th>
<th>$\hat{d}_l$</th>
<th>$\hat{\alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Look 1</td>
<td>250</td>
<td>126</td>
<td>0.3617</td>
<td>0.0011</td>
<td>1709</td>
<td></td>
</tr>
<tr>
<td>Look 2</td>
<td>300</td>
<td>148</td>
<td>0.6248</td>
<td>0.0121</td>
<td>1688</td>
<td></td>
</tr>
<tr>
<td>Look 3</td>
<td>350</td>
<td>174</td>
<td>1</td>
<td>0.0373</td>
<td>1501</td>
<td>0.0495 (0.0043)</td>
</tr>
</tbody>
</table>
Conclusions

1. This develops a sequential monitoring plan for conditional randomization tests under any restricted randomization procedure, provided we can compute exactly probabilities like $P(N_1(n) = n_1 | N_1(j + 1) = m_j + 1)$. For generalizations of the biased coin design, these are hard to derive.

2. We have answered a question posed by Zelen at a recent JSM: “How does one compute conditional randomization tests?” Sampling from the conditional reference set facilitated the approximation of the distribution of the conditional randomization test.

3. To our knowledge, no one has ever established a method for sequential monitoring of conditional randomization tests. We now have a format in which this can be done.