RANDOMIZATION
Basic Ideas and Insights

Marvin Zelen
Harvard University

Graybill Conference
Ft.Collins, Colorado

June 11, 2008

Outline of Talk

- Introduction
- Randomized vs. Observational Studies
- Conditioning on sample size
- Events of low frequency, cluster randomization and the Poisson distribution
- Local vs. Global Inference : Restricted inference and increase in power.
1. Introduction

Randomization was first introduced by R.A. Fisher when he was at the Rothamstead Agricultural Station. Experimental units were plots of land to which treatments were assigned by randomization. Each plot of land had the same opportunity (chance) of receiving one of the treatments under study.

Most visible use of randomization today is the Phase III clinical trial

Experimental units are: patients

Definition: A randomized study is defined by each experimental unit having the same probability of being assigned one of the treatments under study. Implicitly a randomized study is one in which the investigator: (i) selects the treatments to study and (ii) the experimental units to be included in the study.

---

**Fundamental Principal In Comparing Groups**

- Groups must be alike in all important aspects and only differ in the treatment which each group receives.
- Otherwise difference in response between the groups may not be due to the treatments under study, but can be attributed to the particular characteristics of the group.
**Randomization**

Each patient has the same opportunity, as any other patient, of receiving any of the treatments under study. Allocation of treatments to patients is generally carried out using a chance mechanism so that neither the patient nor the physician know in advance which therapy will be assigned. **Randomized clinical trials are regarded as the best way to carry out clinical trials.**

---

**Advantages To Randomization**

- **Eliminates conscious bias**
  - Physician selection
  - Patient self selection

- **Balances unknown biases among treatment groups**
  - Supportive care
  - Patient management
  - Diagnosis and staging
  - Patient evaluation
  - Unknown factors affecting outcome

- **Groups are “alike on the average”**

- **The entire inference can only depend on the randomization.**
Disadvantages of Randomization

- Patient or physician may not care to participate in experiment involving a chance mechanism to decide treatment.

- May interfere with physician-patient relationship

2. Randomized vs. Observational Studies

Ideal observational study has well-defined populations

No control of treatments used in study

No control of how (experimental units) adopt or use treatments. Treatments may be drugs for a disease, diet, life-style characteristics — smoking, alcohol consumption, etc.

It is possible that the conclusions from observational studies and randomized studies may not agree. The principal reason is that observational studies may be biased, by the relationship between unknown factors and the selection of treatment.
Role of Unknown Prognostic Factors; Observational vs. Randomized Studies

\[ y : \text{outcome random variable} \]
\[ z : \text{indicator variable for treatment; } z = 1, 2, \ldots \]
\[ x : \text{prognostic variable which is unknown} \]

Comparisons are usually made comparing outcomes conditional on different treatments.

Let \( f(x, y, z) = \text{joint distribution} \). Then comparisons conditional on treatment correspond to \( f(y|z) \).

\[
f(x, y, z) = f(x)f(z|x)f(y|x, z)
\]

\[
f(y|z) = \sum_x f(x)f(z|x)f(y|x, z)/f(z)
\]

Suppose outcomes are independent of treatment; \( f(y|x, z) = f(y|x) \)

\[ f(y|z) = \sum_x f(x)f(z|x)f(y|x)/f(z) \]

Observational Study:

Randomized study \( f(z|x) = f(z) \)

(independent of \( x \))

Randomized Study:

\[
f(y|z) = \sum_x f(x)f(y|x)
\]

Note that even though outcomes are independent of treatment, the observational study will show dependence on treatment.
3. Conditioning on Sample Size

**Aim:** Randomize experimental units to two treatments \((A, B)\).

**Notation:**

\[
N = 2n \text{ number of experimental units evaluable}
\]

\[
Y_i : \text{ outcome for } i^{th} \text{ experimental unit} \\
\quad \quad \quad \quad \quad \quad \text{(not random)}
\]

\[
\delta_i = \begin{cases} 
1 & \text{if } i^{th} \text{ unit assigned to } A \\
0 & \text{otherwise}
\end{cases}
\]

---

**Two Methods of Randomization**

- **Method I:** Place two balls labeled A and B in urn. Sample balls with replacement.

- **Method II:** Place \(n\) balls in urn labeled A and \(n\) balls in urn labeled B. Sample without replacement.

**Note:** Method I may have unequal sample sizes for A and B. Method II has equal sample sizes.
Two Methods of Randomization

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>{δ_i}</td>
<td>independent</td>
<td>dependent</td>
</tr>
<tr>
<td>(P[\delta_i = 1])</td>
<td>(\frac{1}{2})</td>
<td>(\frac{1}{2})</td>
</tr>
<tr>
<td>(\text{var} \ \delta_i)</td>
<td>(\frac{1}{4})</td>
<td>(\frac{1}{4})</td>
</tr>
<tr>
<td>(\text{cov} (\delta_i, \delta_j), i \neq j)</td>
<td>0</td>
<td>(-\frac{1}{4(N-1)})</td>
</tr>
<tr>
<td>(E(S_a))</td>
<td>(\frac{N}{4} \bar{Y} / 2)</td>
<td>(\frac{N}{4} \bar{Y} / 2)</td>
</tr>
<tr>
<td>(V(S_a))</td>
<td>(\frac{1}{4} \sum_{i=1}^{N} Y_i^2)</td>
<td>(\frac{1}{4} \sum_{i=1}^{N} (Y_i - \bar{Y})^2 / (N-1))</td>
</tr>
</tbody>
</table>

\[\frac{N}{4} \sum_{i=1}^{N} \frac{(Y_i - \bar{Y})^2}{N-1} \leq \frac{1}{4} \sum_{i=1}^{N} Y_i^2\] for nearly all \(N\)

Analysis: Sample sizes fixed or random?

\[S_a = \sum_{i=1}^{N} \delta_i Y_i, S_b + S_a = S\]

\[n_a = \sum_{i=1}^{N} \delta_i, n_a + n_b = N\]

Since \(S_a + S_b = S\), comparison of \(\bar{Y}_a = S_a / n_a\) with \(\bar{y}_b = S_b / n_b\) is equivalent to comparing \(S_a\) with its expected value.

Although the sample size in Method I is random, the analysis should condition on the actual sample size observed as this will result in a smaller variance for comparing A with B.

\[\text{Var } S_a : \ \Sigma Y_i^2 / 4 \text{ (Method I)}\]

\[\text{Var } S_b : \ N \ pq \ \Sigma (Y_i - \bar{Y})^2 / (N - 1) \text{ (Method II)}\]

Where \(p = \frac{n_a}{N}, \ q = 1-p\)
Cluster Sampling

Up to now we have considered randomization in which an individual experimental unit (patients) are randomized to treatments.

In some situations it is more feasible to randomize treatments to groups of experimental units; i.e. groups may be patients in a hospital, families, geographic regions. Cluster randomization or cluster sampling is used to describe this kind of random assignment.

We will show that cluster and individual randomization are equivalent if the outcomes have low probabilities of occurring.

4. Events of Low Frequency, the Poisson Distribution and Cluster Sampling

Suppose $Y_j = \begin{cases} 1 & \text{if event observed} \\ 0 & \text{otherwise} \end{cases}$ and event is rare; e.g. incidence or death due to particular disease; incidence of breast cancer is 80-110/100,000 per year.

\[
E(S_u|n_u) = pNY.
\]

\[
V(S_u|n_u) = pq \frac{N}{N-1} \sum_{j=1}^{N} (Y_j - \bar{Y})^2
\]

\[
\sum_{j} (Y_j - \bar{Y})^2 = \sum_{j} Y_j^2 - NY^2 \sim \sum_{j} Y_j = NY,
\]

neglecting $\bar{Y}^2$
Is asymptotic
\[ N(0,1) \]

Where \( p = \frac{n_a}{N}, \quad q = 1 - p, \quad S = \sum Y_i \)

Alternatively, suppose \( S_a, S_b \) are independent Poisson random variables with

\[ E(S_a) = N_a \lambda, \quad E(S_b) = N_b \lambda, \]

Then the distribution of \( S_a \), conditional on \( S = S_a + S_b \) is binomial with

\[ E(S_a | S) = pS, \quad V(S_a | S) = pqS \]

**Conclusion**: Cluster Sampling from a Poisson distribution is equivalent to individual randomization.

---

**Local vs Global Inference**

- **Local Inference**: Conclusions apply only to the patients entered on trial.

- **Global Inference**: Conclusions of trial apply to the population with disease.

Under what circumstances do these different kinds of inferences apply?
Idealized Clinical Trial Process

Population with Disease

Random Sample of Patients

Randomization

Global Inference requires a random sample of patients
Random Sample of Patients ???

- Nearly all clinical trials do not have a random sample of patients.

- Only a local inference can be made unless additional assumptions are assumed.

- The randomization process can serve as the basis for making a local inference.

- In most situations power will be increased as the inference is more narrow.

Basic Philosophy and Practise of Analysis of Trials

**Philosophy**

- Analysis should be guided by the design of the study and account for significant factors affecting trial outcomes

**Practise**

Most analyses of multi center trials ignore the design of the study :e.g permuted blocks and center variation.
Investigations of Local Inference:

**Power**

- Outcome variables: continuous, binomial, survival
- Randomized multi center trials (two treatments)
- Methodology conditions on the number of patients randomized within a center (conditions on ancillary statistic)
- Study design is permuted blocks of size four:
  Sample size for each treatment is equalized for every fourth patient entered on trial.
- Group sequential and fixed sample size.

---

### Table 1: Power Comparison for Continuous and Binary Outcomes ($N = 4$)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>120</th>
<th>360</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Institutions</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Conditional test</td>
<td>.55</td>
<td>.48</td>
</tr>
<tr>
<td>Stratified t</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>Conditional test</td>
<td>.38</td>
<td>.36</td>
</tr>
<tr>
<td>Mantel-Haenszel</td>
<td>.30</td>
<td>.29</td>
</tr>
</tbody>
</table>

**NOTE:** $\alpha = .05$, two-tailed test.

Continuous outcomes: Data are generated from lognormal distribution. The difference between two group means is 1.07; Binary outcomes: Data are generated from Bernoulli distribution.

Success probabilities for two treatments are .5 and .7 respectively.
Table 2: Power Comparison for Survival Outcomes ($N = 4$)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>120</th>
<th>360</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Institutions</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Conditional test (Gehan)</td>
<td>.38</td>
<td>.33</td>
</tr>
<tr>
<td>Stratified Gehan</td>
<td>.19</td>
<td>.14</td>
</tr>
<tr>
<td>Conditional test (Logrank)</td>
<td>.37</td>
<td>.32</td>
</tr>
<tr>
<td>Stratified Logrank</td>
<td>.20</td>
<td>.14</td>
</tr>
</tbody>
</table>

NOTE: $\alpha = .05$, two-sided test.

Data are generated from exponential distribution. Ratio of two hazards is 2. Percentage of censored observations is 18.5%.

Table 3: Power Comparison in Group Sequential Testing

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Type of Test</th>
<th>No. of Institutions</th>
<th>10</th>
<th>20</th>
<th>40</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous</td>
<td>Conditional</td>
<td>.72</td>
<td>.70</td>
<td>.69</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unconditional</td>
<td>.51</td>
<td>.49</td>
<td>.48</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td>Binary</td>
<td>Conditional</td>
<td>.91</td>
<td>.91</td>
<td>.90</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unconditional</td>
<td>.76</td>
<td>.76</td>
<td>.77</td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>Survival*</td>
<td>Conditional (Logrank score)</td>
<td>.64</td>
<td>.63</td>
<td>.62</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uncond. Stratified Logrank</td>
<td>.57</td>
<td>.57</td>
<td>.57</td>
<td>.57</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Maximum sample sizes is 400 with 4 looks at equal information increments. One-sided test with $\alpha = .025$. Block size is 4. *percentage of censored outcome is 18.5%.

Stopping boundaries are calculated using O'Brien-Fleming rule.
Thank you for coming!