DESIGN PRINCIPLES FOR SMALL CLINICAL TRIALS

Christopher S. Coffey
Professor, Department of Biostatistics
Director, Clinical Trials Statistical and Data Management Center

November 27, 2012

The University of Iowa
Clinical Trials Statistical and Data Management Center
In this lecture, we will:

I. Discuss the importance of adequate study planning for small clinical trials

II. Describe some analytical approaches that have merit with small clinical trials

III. Describe several proposed designs for small clinical trials
The wonderful land of Asymptopia:
QUESTION: “What is a small clinical trial?”

ANSWER: Depends on the context.

- A stroke researcher may think of a ‘small clinical trial’ as an early phase trial to develop a new compound.

- An ALS researcher may think of a ‘small clinical trial’ as a confirmatory phase III clinical trial that is limited in size.

We will address both types of studies.
There is no magic – we want the “right” answer

Small study ≠ little version of large study.

We must know what we are sacrificing:

- Perhaps less precision than in large trials
- Perhaps we use a less definitive outcome
Before addressing some possible designs of interest, it is useful to review some key recommendations from the Executive Summary in the National Academy of Sciences document.
Recommendation #1: Define the research question.

Before undertaking a small clinical trial it is particularly important that the research question be well defined and that the outcomes and conditions to be evaluated be selected in a manner that will most likely help clinicians make therapeutic decisions.
Recommendation #2: Tailor the design.

Careful consideration of alternative statistical design and analysis methods should occur at all stages in the multistep process of planning a clinical trial.

When designing a small clinical trial, it is particularly important that the statistical design and analysis methods be customized to address the clinical research question and study population.
Recommendation #3:
Clarify methods of reporting of results of clinical trials.

In reporting the results of a small clinical trial, with its inherent limitations, it is particularly important to carefully describe all sample characteristics and methods of data collection and analysis for synthesis of the data from the research.

Recommendation #4:
Perform corroborative statistical analyses.

Given the greater uncertainties inherent in small clinical trials, several alternative statistical analyses should be performed to evaluate the consistency and robustness of the results of a small clinical trial.
Recommendation #5:
Exercise caution in interpretation.

One should exercise caution in the interpretation of the results of small clinical trials before attempting to extrapolate or generalize those results.
Recommendation #6: More research on alternative designs is needed. Appropriate federal agencies should increase support for expanded theoretical and empirical research on the performances of alternative study designs and analysis methods that can be applied to small studies.

Areas worthy of more study may include theory development, simulated and actual testing including comparison of existing and newly developed or modified alternative designs and methods of analysis, simulation models, study of limitations of trials with different sample sizes, and modification of a trial during its conduct.

Summary:

1) Define your question
2) Tailor your design
3) Report clearly
4) Do corroborative analysis
5) Interpret cautiously
6) Do more research on other designs
Summary:
1) Define your question
2) Tailor your design
3) Report clearly
4) Do corroborative analysis
5) Interpret cautiously
6) Do more research on other designs

So, why is this any different from other trials?
Three basic requirements for any clinical trial:

1) Trial should examine an important research question

2) Trial should use a rigorous methodology that can answer the question of interest

3) Trial must be based on ethical considerations and assure that risks to subjects are minimized
These three requirements should also apply to all small clinical trials.

But, the second requirement (validly addressing a research question) is often problematic in small trials.

Small clinical trials are more prone to variability and may only be adequately powered to detect large intervention effects.

As a consequence, the importance of adequate study planning is magnified in small clinical trials.
Two general approaches:

- Use a methodological approach that enhances the efficiency of standard statistical properties
- Use an alternate/innovative design
Use efficient outcome measures & measure precisely.

Basic formula for design:

\[ N = \frac{2\sigma^2(z_\alpha + z_\beta)^2}{\Delta^2} \Rightarrow N \propto \frac{\sigma^2}{\Delta^2} \]

\( N = \) sample size
\( \sigma^2 = \) variance
\( \Delta = \) clinically meaningful difference/effect to detect
  (or smallest biologically credible difference)
Note that this also implies….

\[ \Delta \propto \frac{\sigma}{\sqrt{N}} \]

If the population we are studying is big (e.g. cardiology, breast cancer, etc.):

- Just increase \( N \) to reduce \( \Delta \)
- And – a little sloppiness is not harmful
BUT: If our population is small (e.g., genetic disease, rare cancers):

- Cannot increase $N$
- Only solution is to decrease the variance
Type of Outcome Measure

- Different types of outcome measures exhibit different levels of accuracy
- Using outcomes that provide higher accuracy generally increases statistical power
Type of Outcome Measure

- Continuous outcomes most efficient
  - Beware statistically, not clinically, significant

- Binary outcomes are least efficient
  - Sometimes the only outcome of real interest
    (Elimination of disease, Restoration of function)

- Time-to-event may be more efficient than binary
Some examples:

- **Hypertension**
  - Continuous: Systolic Blood Pressure (SBP)
  - Binary: SBP < 130 mmHg
  - Time-to-event: Time until first SBP < 130 mmHg

- **Pain**
  - Continuous: Pain Score
  - Binary: Pain Score > 4
  - Time-to-Event: Time to pain relief
Parametric vs. Nonparametric Approaches:

- A nonparametric approach does not require any distributional assumptions
  - Generally more robust
- A parametric approach can lead to higher power, if the distributional results are satisfied

Thus, in a small trial, it is very important to know whether the distributional assumptions (i.e., normality) are satisfied.
How to increase power?

- Usual RCT – As model-free as possible:
  - Have large sample sizes
  - Do Intent-to-Treat Analysis
  - Don’t worry about noise
How to increase power?

- Usual RCT – As model-free as possible
- Small populations
  - Use models (but pre-specify)
  - Check EACH observation before you unblind
  - Carefully evaluate alternative designs
Historical Controls are useful when:

- Comparing a new treatment for a well studied area
- Data from published studies remains relevant
- Randomized controls are not feasible
Historical Controls:

- Advantages:
  - Inexpensive (…not always!)
  - All subjects get desired treatment
  - You often find a BIG difference

- Disadvantages:
  - Current & historical populations may be different
  - Current treatment may be different (even if there is no ‘therapy’)
Designs of interest in small clinical trials:

- Parallel group design
- Repeated measures design
- Crossover design
- N-of-1 design
- Futility design
- Ranking/Selection design
- Bayesian design
- Adaptive designs

Covered in later talks
Multiple observations or response variables are obtained for each subject.

- Repeated measurements over time (longitudinal)
- Multiple measurements on same subject

Allows both between-subject and within-subject comparisons.

Can reduce the required sample size needed to obtain a specific target power.
Suppose you are measuring over time:

- STANDARD: Final value – Baseline value
- BETTER: Final value, with baseline value as a covariate
Suppose you are measuring over time:

- STANDARD: Final value – Baseline value
- BETTER: Final value, with baseline value as a covariate
- STILL BETTER: Longitudinal
  - Differentiate “through” vs. “at”
  - Think about variance/covariance structure
  - Think how you want to model time
CROSSOVER DESIGN

Each subject exposed to all treatments

- Order of treatments randomized
- First may show better (or worse) effect

Prognostic factors balanced – self vs. self

Required sample size reduced considerably due to self vs. self comparisons

Each participant receives the active treatment at some point during the study
CROSSOVER DESIGN

Half look like this:  

<table>
<thead>
<tr>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Pbo</td>
</tr>
</tbody>
</table>

Half look like this:  

<table>
<thead>
<tr>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pbo</td>
<td>Active</td>
</tr>
</tbody>
</table>
Disadvantages:

- Disease needs to be long-term
- Treatment must be taken regularly over time
- Relevant outcomes must occur and be measured over time
- Not relevant for acute treatments
- Concerns due to a ‘carryover effect’
N-OF-1 DESIGN

A special case of a crossover/repeated measures design, that requires only a single subject.

Subject undergoes treatment for several pairs of periods.

For each pair:
- Subject receives experimental treatment for one part of each pair
- Subject receives alternative treatment for other pair
- Order of two treatments within each pair is randomized
The final outcome of the trial is a determination about the best treatment for the particular subject under study.

Results of a series of N-of-1 trials may be combined using meta-analysis.
A futility design is a screening tool to identify agents that should not be candidates for phase III trials while minimizing costs/sample size.

The main intent is to identify ineffective treatments, not to establish efficacy.

The origins lie in cancer research.

- Motivation: Translation of basic research findings to human trials is difficult and prone to failure.
To use a futility design, a researcher must define what would be considered futile.

For example, suppose that a 10% increase in favorable response rates is clinically meaningful.

A futility design would be set up to determine if one can rule out that the new treatment is at least 10% better than the standard treatment (or placebo).
### FUTILITY DESIGN

Statistical Hypotheses:

<table>
<thead>
<tr>
<th></th>
<th>Null Hypothesis ((H_0))</th>
<th>Alternative Hypothesis ((H_A))</th>
<th>Implication of Rejecting (H_0)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Usual Design</strong></td>
<td>(\mu_T = \mu_P)</td>
<td>(\mu_T \neq \mu_P)</td>
<td>New Treatment is Effective (Harmful)</td>
</tr>
<tr>
<td><strong>Futility Design</strong></td>
<td>(\mu_T - \mu_P \geq 0)</td>
<td>(\mu_T - \mu_P &lt; 0)</td>
<td>New Treatment is Futile</td>
</tr>
</tbody>
</table>
Type I and Type II Errors:

<table>
<thead>
<tr>
<th></th>
<th>Type I Error (α)</th>
<th>Type II Error (β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usual Design</td>
<td>Ineffective Therapy is Effective</td>
<td>Effective Therapy is Ineffective</td>
</tr>
<tr>
<td>Futility Design</td>
<td>Effective Therapy is Ineffective</td>
<td>Ineffective Therapy is Effective</td>
</tr>
</tbody>
</table>
Negative predictive values are high
Pr(Not Effective | Futile)

Positive predictive values are not so high
Pr(Effective | Not Futile)

Thus, futility designs are good at identifying ineffective agents, but not good at identifying effective agents.

However, this is an improvement over running underpowered efficacy trials in phase II or conducting phase III trials as the first rigorous test of efficacy for a new treatment.
Ranking and selection procedures are statistical techniques for comparing parameters of multiple (\(k\)) study populations.

Generally require smaller sample sizes than trials designed to estimate and test treatment effects.
Selection designs can be used to:

- Select the treatment with the best response out of \( k \) potential treatments
- Rank treatments in order of preference
- Rule out poor treatments for further study (Helpful with ‘pipeline’ problem)
An appropriate study design has sufficient sample size, adequate power, and proper control of bias to allow a meaningful interpretation of the results.

Although small clinical trials pose important limitations, the above issues cannot be ignored.

The majority of methods research for clinical trials is based on large sample theory.

Additional research into innovative designs for small clinical trials is needed.
1. Which of the following describes the N-of-1 Design?

a) A design to address whether there is clear evidence that some desired level of effect would not be achieved in a larger phase III trial.

b) A design that require only a single subject, who undergoes treatment for several pairs of periods.

c) A design where multiple response variables are observed for each subject.

d) A design that can be used to select the treatment with the best response out of a number of possible treatments

e) None of the above
2. Which of the following describes the Futility Design?

a) A design to address whether there is clear evidence that some desired level of effect would not be achieved in a larger phase III trial.

b) A design that require only a single subject, who undergoes treatment for several pairs of periods.

c) A design where multiple response variables are observed for each subject.

d) A design that can be used to select the treatment with the best response out of a number of possible treatments.

e) None of the above
3. Which of the following describes the Repeated Measures Design?

a) A design to address whether there is clear evidence that some desired level of effect would not be achieved in a larger phase III trial.

b) A design that require only a single subject, who undergoes treatment for several pairs of periods.

c) A design where multiple response variables are observed for each subject.

d) A design that can be used to select the treatment with the best response out of a number of possible treatments

e) None of the above