Efficient Clinical Trials

John H. Powers, MD
Senior Medical Scientist
SAIC in support of
Clinical Collaborative Research Branch
NIAID/NIH
Introduction

- Definitions – what is a “small” clinical trial?
- Review of ethical considerations in designing trials
- Review of basic “ingredients” of sample size estimation and concept of “power”
- Ways to increase efficiency of trials – pros and cons for various methods
  - Focused and relevant research question
  - Changing error rates (not suggested)
  - Enhancing effect sizes
  - Decreasing variability
Definitions

- **Clinical trial** – a controlled prospective study enrolling human subjects often used to evaluate the effectiveness and/or harms of interventions in treatment, prevention or diagnosis of disease.

- **Efficiency** –
  - (in physics) ratio of useful work to the energy supplied to it
  - In clinical trials, getting valid and reliable answers to important questions with the least amount of resources
    - Does not mean putting patients at risk because of less valid data
    - Lower sample size does not mean less work = MORE planning
Lower Sample Size = More Planning

“Clinical trials with small numbers of participants, however, must address broad sets of issues different from those that must be addressed in trials with large numbers of participants. It is in those circumstances of trials with small sample sizes that approaches to optimization of the study design and data interpretation pose greater challenges.”

- Small Clinical Trials: Issues and Challenges, Institute of Medicine, 2001 http://www.nap.edu/catalog/10078.html p. ix
Definitions

- What is a “large” or “small” clinical trial?

- IOM defines a “large” trial as one that has adequate sample size to answer the primary research question = “large enough”

- A trial with very few participants may still have adequate statistical power e.g. if effect size is large

- Balance between exposing research subjects to potential harms of experimental interventions with obtaining valid answers
The Continuing Unethical Conduct of Underpowered Clinical Trials

Scott D. Halpern, MSCE
Jason H. T. Karlawish, MD
Jesse A. Berlin, ScD

More than 20 years have passed since investigators first described the ethical problems of conducting randomized controlled trials (RCTs) with insufficient statistical power. Because such studies may not adequately test the underlying hypotheses, they have been considered “scientifically useless” and therefore unethical in their exposure of participants to the risks and burdens of human research. Despite this long-standing challenge, many clinical investigators continue to conduct underpowered studies and fail to calculate or report appropriate (a priori) power analyses. Not only do these scientific and ethical errors persist in the general medical literature, but 3 recent reports also highlight the alarming prevalence of these problems in more specialized fields.

Patients and healthy volunteers thus continue to participate in research that
Underpowered Studies and Ethics

- “A proposed study that cannot answer the question being asked because the necessary sample size cannot be attained should not be conducted on ethical grounds. That is, it is unacceptable to expose patients or research participants to harms, even inconveniences, if there is no prospect that useful and potentially generalizable information will result from the study.”

- Small Clinical Trials: Issues and Challenges, Institute of Medicine, 2001
  http://www.nap.edu/catalog/10078.html p. 14
“The importance of conducting small clinical trials only when there are no alternatives cannot be overemphasized. The committee is not encouraging the use of small clinical trials, but, rather provides advice on strategies that should be considered in the design and analysis of small clinical trials when the opportunity to perform a randomized clinical trial with adequate statistical power is not possible. In doing so, it recognizes that small clinical trials frequently need to be viewed as part of a continuing process of data collection. Thus, for some trials it might be impossible to definitively answer a research question with a high degree of confidence. In those cases, perhaps the best that one can do is assess the next set of questions to be asked.”

Concerns About Small Clinical Trials

- Small numbers increase variability and leave much to chance
- Statistically significant outcomes may not be generalizable (only apply to circumstances in trial)
- Too many variables to assess cause and effect
- Only able to discern gross effects and limited ability to analyze covariates
- Incapable of identifying adverse events

Small Clinical Trials: Issues and Challenges, Institute of Medicine, 2001 [http://www.nap.edu/catalog/10078.html](http://www.nap.edu/catalog/10078.html) p. 15
Situations Where Smaller Clinical Trials Justifiable

- Rare diseases
- Unique study populations (e.g. astronauts)
- Individually tailored therapies
- Environments that are isolated
- Emergency situations
- Public health urgencies

Small vs Efficient

- While small clinical trials are a last resort, efficient clinical trials are always justifiable.

- Different methods to improve efficiency are useful (or not) depending on disease under study and research question/setting.
Review of Sample Size Considerations
Normal Distribution of a Sample

95% of data will be within 1.96 standard deviations of sample mean for large samples (>30)
Sample Size – Superiority Trials

- Clinical trials compare average effects in groups of subjects administered intervention to those not administered intervention
- Examine if populations differ by more than chance (for superiority trials)
- Example: Two groups with point estimate for means of 50 and zero, sample size of 12 per group, SD = 60
In this example

- Blue area represents power (in this case 0.497)

- “critical region” of 0.05 test represented by dashed lines and red area

- To show difference due to greater than chance, want mean of one curve to be outside of red area
Sample Size

- Sample size in this example is 24 per group

- As sample size increases, overlap between curves decreases (assuming there really is a difference to show)

- Blue area increases = power is 0.80

- Mean of one group now outside of “critical area”

- Notice still a good deal of overlap – only mean value is outside critical area
Sample Size

- Sample size increased to 48 per group
- Distribution of data narrower and more precise
- Power = 0.98
- Mean well outside of “critical area”
Sample Size

- Sample size increased to 96 per group
- Data still more precise
- Power = 1.000 since no overlap of curves
- Mean well outside of “critical area”

http://www.tufts.edu/~Egdallal/sizenotes.htm
Sample Size

- Want to select sample size large enough to show a difference if there is one to detect, but not too large
  - Do not want to expose subjects unnecessarily to harm since this is an experiment evaluating interventions with unknown harm/benefits
  - Use of resources - time, effort and money

- Sample size based on four parameters ("ingredients")
  - Type 1 error - usually specified as 0.05 two sided (0.025 on either side of curve)
  - Type 2 error (1 - type 2 error is power) usually specified as 0.10 to 0.20 (power of 80%-90%)
  - Standard deviation of data (variability)
  - Treatment difference – Difference between point estimate of effect for intervention and point estimate for effect with control (delta)
Pictorial Representation of Sample Size

Critical Value

$H_0 : \mu_0 - \mu_1 = 0$

$H_1 : \mu_0 - \mu_1 = \delta$

$S.E. = \sigma \sqrt{\frac{2}{n}}$

$\alpha/2$

$0 + z_{1-\alpha/2} \sigma \sqrt{\frac{2}{n}}$

$\delta - z_{1-\beta} \sigma \sqrt{\frac{2}{n}}$

Power $= 1 - \beta$

$\bar{y}_0 - \bar{y}_1$
Pictorial Representation of Sample Size

**treatment effect size**

\[ H_0 : \mu_0 - \mu_1 = 0 \quad H_1 : \mu_0 - \mu_1 = \delta \]

**Critical Value**

**variability**

\[ S.E. = \sigma \sqrt{\frac{2}{n}} \]

\[ 0 \pm z_{1-\alpha/2} \sigma \sqrt{\frac{2}{n}} \quad \delta - z_{1-\beta} \sigma \sqrt{\frac{2}{n}} \]
Efficient Clinical Trials

Ways to decrease sample size
1. Focused and relevant research question
2. Changing error rates (not suggested)
3. Enhancing effect sizes
   - More homogenous populations
   - Choosing populations in whom effect size is larger
   - Optimizing exposure
   - Continuous instead of dichotomous outcomes
   - More common or composite outcomes
4. Decreasing variability
   - More sensitive/specific measures
   - Assuring follow-up of enrolled subjects
   - Study designs
     - Cross-over
     - N=1 studies
     - Sequential trial designs (e.g. dose-escalation studies)
Efficient Clinical Trials

- **Ways to decrease sample size**
  1. Focused and relevant research question
  2. Changing error rates (not suggested)
  3. Enhancing effect sizes
     - More homogenous populations
     - Choosing populations in whom effect size is larger
     - Optimizing exposure
     - Continuous instead of dichotomous outcomes
     - More common or composite outcomes
  4. Decreasing variability
     - More sensitive/specific measures
     - Assuring follow-up of enrolled subjects
     - Study designs
        - Cross-over
        - N=1 studies
        - Sequential trial designs (e.g. dose-escalation studies)
1. Focusing the Research Question

- Research question needs to be one worth answering and of public health importance

- Need to focus question – more questions mean greater sample size or less clear answers = simplify

- “Many trials include measurements to try to figure out why the trial didn’t work after it has failed” – the post-mortem on “what the experiment died of”
1. Focusing the Research Question

- Sample size is calculated AFTER one decides on a research question.

- Starting out with a sample size and working back to “what can I get for this” is not justifiable in terms of choosing unrealistic or clinically meaningless/unachievable effect sizes.
Efficient Clinical Trials

Ways to decrease sample size
1. Focused and relevant research question
2. Changing error rates (not suggested)
3. Enhancing effect sizes
   - More homogenous populations
   - Choosing populations in whom effect size is larger
   - Optimizing exposure
   - Continuous instead of dichotomous outcomes
   - More common or composite outcomes
4. Decreasing variability
   - More sensitive/specific measures
   - Assuring follow-up of enrolled subjects
   - Study designs
     - Cross-over
     - N=1 studies
     - Sequential trial designs (e.g. dose-escalation studies)
2. Changing Error Rates

- Error rates of type 1 = 0.05 and type 2 = 0.10 or 0.20 are by convention

- But...false positive error rate of 1 in 20 trials is actually a low level of evidence; need justification to deviate from this

- Increasing type 2 error rate increases likelihood of false negative conclusions = spending resources for unclear answers
Sample Size and Types of Error/Power
2. Changing Error Rates

- Some discussion in medical literature of using one-sided rather than two-sided hypothesis testing = needs justification
  - But….changes level of evidence to 1 in 10 false positive rate – what justifies this?
  - Assumes we are only interested in only one side of the story – how much better is the test intervention, but don’t we care how much harm might be done?
  - Often statements of “this intervention can’t possibly be harmful” are not justified by evidence

- So…don’t change error rates unless scientifically justified and even then most researchers want to see one sided 0.025 error rates
Efficient Clinical Trials

• **Ways to decrease sample size**
  1. Focused and relevant research question
  2. Changing error rates (not suggested)
  3. Enhancing effect sizes
     - More homogenous populations
     - Choosing populations in whom effect size is larger
     - Optimizing exposure
     - Continuous instead of dichotomous outcomes
     - More common or composite outcomes
  4. Decreasing variability
     - More sensitive/specific measures
     - Assuring follow-up of enrolled subjects
     - Study designs
        - Cross-over
        - N=1 studies
        - Sequential trial designs (e.g. dose-escalation studies)
Enhancing Effect Sizes

- Trials measure average effects in groups of subjects – “one size fits all” approach may not be correct

- More homogenous populations can both decrease variability and increase effect sizes in presence of effect modification

- Effect modification is presence of quantitatively different effect sizes of an intervention based on a baseline variable

- Requires knowledge of natural history of disease and some evidence from prior trials

- Choosing population in which effect size is larger decrease sample size
Confounding and Effect Modification

Outcome

Baseline variable

No confounding

No effect modification

test

control

\{\text{effect size}\}

Confounding

No effect modification

No confounding

Effect modification

Confounding

Effect modification
Examples

- **Enrolling subjects in trial in whom effect is expected to be zero**
  - Dilutes effect size
  - Ethical issues of exposure to harm for no benefit

- **Trastuzumab (Herceptin) in breast cancer**
  - Mechanism of action by binding to HER2 proteins in person with specific genetic mutation
  - 20% to 30% of person with breast cancer have this mutation
  - Potentially harmful in those without the mutation
Efficient Clinical Trials

- **Ways to decrease sample size**
  1. Focused and relevant research question
  2. Changing error rates (not suggested)
  3. Enhancing effect sizes
     - More homogenous populations
     - Choosing populations in whom effect size is larger
     - Optimizing exposure
     - Continuous instead of dichotomous outcomes
     - More common or composite outcomes
  4. Decreasing variability
     - More sensitive/specific measures
     - Assuring follow-up of enrolled subjects
     - Study designs
       - Cross-over
       - N=1 studies
       - Sequential trial designs (e.g. dose-escalation studies)
Optimizing Exposure

- Many trials include only one dose of an intervention with wide inter-individual variability in exposure

- Optimize dose based on pre-clinical and early clinical studies – pharmacokinetics and pharmacodynamics
  - Forms a hypothesis to test
  - Not a substitute for clinical trials

- Standardize exposure of interventions - need unblinded third party to do this to maintain blinding in trial
Efficient Clinical Trials

- **Ways to decrease sample size**
  1. Focused and relevant research question
  2. Changing error rates (not suggested)
  3. Enhancing effect sizes
     - More homogenous populations
     - Choosing populations in whom effect size is larger
     - Optimizing exposure
     - Continuous instead of dichotomous outcomes
     - More common or composite outcomes
  4. Decreasing variability
     - More sensitive/specific measures
     - Assuring follow-up of enrolled subjects
     - Study designs
        - Cross-over
        - N=1 studies
        - Sequential trial designs (e.g. dose-escalation studies)
Continuous vs Dichotomous Outcomes

- Continuous outcomes have more power to detect differences since uses all the data.

- Dichotomous outcomes require:
  1. Categorization – assumes all data in a single category are similarly important which may not be true.
  2. Choosing correct time point to evaluate – if you’re wrong, you miss it.

- Requires more frequent data capture – patient diaries or phone contact collected in systematic way.
Time to Event - Cholera

Time to Event - Influenza

Figure 2. Time to Alleviation of All Symptoms in Influenza-Infected Patients

Efficient Clinical Trials

- **Ways to decrease sample size**
  1. Focused and relevant research question
  2. Changing error rates (not suggested)
  3. Enhancing effect sizes
    - More homogenous populations
    - Choosing populations in whom effect size is larger
    - Optimizing exposure
    - Continuous instead of dichotomous outcomes
    - More common or composite outcomes
  4. Decreasing variability
    - More sensitive-specific measures
    - Assuring follow-up of enrolled subjects
    - Study designs
      - Cross-over
      - N=1 studies
      - Sequential trial designs (e.g., dose-escalation studies)
Common or Composite Outcomes

- Choosing more common events or composite outcomes increases number of events and increases power to detect differences.

- Some problems with interpretation:
  - Only use when outcomes measured are of similar importance to patients – if driven by less important outcomes may mask inferiority on more important outcomes.
  - Does not necessarily imply beneficial effect on all part of a composite.
Composite Endpoints - Issues

Table I. Structure of data on death and hospitalization (hypothetical data).

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Active</td>
</tr>
<tr>
<td>Died but never hospitalized during follow-up (%)</td>
<td>15</td>
</tr>
<tr>
<td>Hospitalized and died during follow-up (%)</td>
<td>5</td>
</tr>
<tr>
<td>Hospitalized, alive at the end of follow-up (%)</td>
<td>20</td>
</tr>
<tr>
<td>None of the above (%)</td>
<td>60</td>
</tr>
</tbody>
</table>
Efficient Clinical Trials

- **Ways to decrease sample size**
  1. Focused and relevant research question
  2. Changing error rates (not suggested)
  3. Enhancing effect sizes
     - More homogenous populations
     - Choosing populations in whom effect size is larger
     - Optimizing exposure
     - Continuous instead of dichotomous outcomes
     - More common or composite outcomes
  4. Decreasing variability
     - More sensitive/specific measures
     - Assuring follow-up of enrolled subjects
     - Study designs
        - Cross-over
        - N=1 studies
        - Sequential trial designs (e.g. dose-escalation studies)
More Sensitive/Specific Measures

- All measures composed of the true measure plus some associated error

- Error can be of two types:
  - Random – by chance alone
  - Systematic bias – based on individual biases and interpretations

- Decreasing error (“noise”) in relation to true measure (“signal”) allows smaller sample size
Non-Standardized Measures and Error

Non-Standardized Measures and Error


- 16 oncologists asked to measure 12 simulated tumor masses, two pairs of which were identical in size.

- Allowed assessment of 64 measurements by same investigator and 1920 comparisons by different investigators.

- 25% “reduction” as “response” for identical size masses = 19% response rate by same investigator, 25% between investigators (measurement error alone).
NMEs with PROs in approved labeling, by therapeutic category, ‘97–’01 (Willke et al, CCT 2004)

Note: Only therapeutic classes with at least 9 approvals are included.
Surrogate Endpoints

- Researchers often suggest biomarkers as “surrogate endpoints” in clinical trials – NOT direct measures of patient benefit

- Idea originally was to decrease follow-up time in chronic diseases – keep following subjects to validate biomarker (e.g. viral load in HIV/AIDS, cholesterol in stroke/MI prevention)

- Why use a surrogate in an acute disease when one can measure actual clinical outcomes?

- Surrogate as part of composite outcomes drive the entire outcome since more common
Limitations of Surrogate Endpoints

1. Surrogate endpoints may not be a true predictor of clinical outcomes
   - Interventions may be clinically useful with no effect on surrogate
   - Interventions may have effect on surrogate but no measurable clinical effects (or actual harms for patients)
   - Still must measure clinical events that occur during trial and these supersede effect on surrogate
   - May measure short term benefits that do not persist with longer followup

2. Surrogate endpoints may not yield a quantifiable clinical benefit that may be measured directly against adverse effects
   - ICH E-9 document, p.7

3. Surrogate endpoints may hinder development of new types of therapies by insisting on only one mechanism of action by which intervention may affect disease
Reasons why surrogate may not accurately predict clinical outcomes:

• unmeasured harms caused by intervention
• unmeasured benefits caused by intervention
• other mechanisms of disease other than those affected by intervention
• issues with measuring surrogate
• issues with measuring clinical outcomes
Use More Sensitive and Specific Measures

Comparison of pain scales in 124 patients:

NSAID free baseline vs. 4 wks oxaprozin
Bellamy et al., Curr Med Res Opin. 1999;15:121-7

<table>
<thead>
<tr>
<th>Scale</th>
<th>Improvement/SD</th>
<th>N for identical power</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAS</td>
<td>1.08</td>
<td>100</td>
</tr>
<tr>
<td>0-10 numerical</td>
<td>1.08</td>
<td>100</td>
</tr>
<tr>
<td>5 point pain category</td>
<td>0.97</td>
<td>120</td>
</tr>
<tr>
<td>Pain faces (Champion)</td>
<td>0.90</td>
<td>144</td>
</tr>
<tr>
<td>McGill (total)</td>
<td>0.68</td>
<td>256</td>
</tr>
</tbody>
</table>

From Dr Max’s talk IPPCR 2008
Efficient Clinical Trials

- Ways to decrease sample size
  1. Focused and relevant research question
  2. Changing error rates (not suggested)
  3. Enhancing effect sizes
     - More homogenous populations
     - Choosing populations in whom effect size is larger
     - Optimizing exposure
     - Continuous instead of dichotomous outcomes
     - More common or composite outcomes
  4. Decreasing variability
     - More sensitive/specific measures
     - Assuring follow-up of enrolled subjects
     - Study designs
       - Cross-over
       - N=1 studies
       - Sequential trial designs (e.g. dose-escalation studies)
Follow-Up and Missing Data

- Enrolling subjects and then losing them to follow-up or missing data (failure to collect or losing it) results in effort for no gain

- Requires planning on part of researchers and work during the trial to make it as easy as possible for research subjects to return
  - Phone calls and reminders
  - Transportation
  - Home visits

- Subjects who don’t follow protocol are not “missing” and should be included in Intent to Treat (ITT) analyses
Efficient Clinical Trials

Ways to decrease sample size
1. Focused and relevant research question
2. Changing error rates (not suggested)
3. Enhancing effect sizes
   - More homogenous populations
   - Choosing populations in whom effect size is larger
   - Optimizing exposure
   - Continuous instead of dichotomous outcomes
   - More common or composite outcomes
4. Decreasing variability
   - More sensitive/specific measures
   - Assuring follow-up of enrolled subjects
   - Study designs
     - Cross-over
     - N=1 studies
     - Sequential trial designs (e.g. dose-escalation studies)
Certain kinds of study designs can decrease variability and thereby decrease sample size.

Cross-over and n=1 trials both use subjects as their own controls.

Randomize subjects to receive one intervention or the other (sometimes with wash out period in between).
Study Designs

- Limitations of cross-over and n=1 trials
  - Most useful in chronic illnesses with stable course of disease
  - If effects carried over from one period treatment to the next then can bias study results
  - “Period effect” for instance in seasonal diseases
  - Most useful in diseases where treatment effect is rapid onset or rapid cessation when intervention stopped
Study Designs

- Sequential trial designs often used in Phase 1 dose escalation clinical trials
- Based on pre-defining what level of adverse events or not will allow progression to the next dose
- Makes decision based on data acquired during trial
Study Designs

- “Adaptive” designs - word that describes a variety of changes in trial design based on data accumulated during the trial

- Sequential trial dose escalation design is one form but more challenging when modifying other variables like the outcome measure

- Advantage = more subjects assigned to more successful treatment

- Disadvantage = heterogeneity of subjects based on important risk factors which change as trial progresses introduces bias over time
A caveat....

- All the modification we have discussed apply to SUPERIORITY trials

- In non-inferiority trials investigators are locked into the design that was used to demonstrate the effect of the control intervention compared to placebo
  - Disease definition
  - Population
  - Outcome definition and timing
  - Concomitant medications

- Biases which trend results toward no difference in a superiority trial (like too small a sample size) result in false positive conclusions in non-inferiority trials
Non-Inferiority Trials

- Need to know reliable and reproducible benefit of control compared to placebo – not “recommended in treatment guidelines” only

- Once this is known, need to evaluate methods used in the prior placebo controlled trials – just like doing an experiment in the lab where conditions are held constant

- Given numerous biases should be trial design of last resort –

- Reasons to use NI trials not consistent with reasons to do small clinical trials (presupposes an already effective intervention)
Non-inferiority trials are unethical because they disregard patients’ interests

Silvio Garattini, Vittorio Bertele

Equivalence trials have been widely used to assess new drugs, but have recently lost ground to a non-inferiority design. This type of trial is usually accepted by regulatory authorities for approval of new drugs or new indications, although the US Food and Drugs Administration has raised some concerns. In this paper, we argue that the scientific community should ban non-inferiority and equivalence trials because they are unethical, whatever measures are taken to prevent their methodological pitfalls and inappropriate interpretation of results. Exceptions might exist, but we could not identify a situation in which patients can justifiably be entered into a trial that will not provide them with any advantage. but not to the extent that it is recognised as such. For example, if the non-inferiority limit is set at 7.5%, an increase in the incidence of serious events or deaths—say 7% instead of the 5% currently established for the comparator—is not seen as large enough to mark a difference between the new and the control drug. The new drug will therefore be considered non-inferior to the old drug, even if in 1000 patients treated with the former, there could be 20 more deaths than with the latter.

These arguments also apply to equivalence trials, which aim to prove similarity of a new drug to the comparator, since true equivalence is theoretical and is difficult to
Conclusions

- Developing an efficient trial starts with planning and a good research question.

- Question comes first, sample size second.

- Various methods to increase effect sizes and decrease variability, when applied in the correct setting, can provide valid and reliable answers to important public health questions.

- For some disease, developing the tools (better outcome measures, better data on natural history) is a good start to better trials.