CLINICAL TRIAL DESIGN BASICS

Suzanne Dahlberg, PhD
Dept. of Biostatistics & Computational Biology
December 11, 2018

A clinical trial is a scientific (*statistical*) experiment involving human subjects in which treatment is initiated specifically for therapy evaluation.

It is not an idealized experiment.

3

Bias and Random Error

Goals of study design are to

- control bias (systematic error)
- minimize variability (random error)

Random error averages over many observations, but bias cannot be eliminated.

4

Sources of Bias

- Selection bias
- Assignment bias

Specifically

- Post-enrollment exclusions
- Losses to follow-up
- Retroactive definitions

Minimize bias in different ways
(for example, with randomization, blinding)

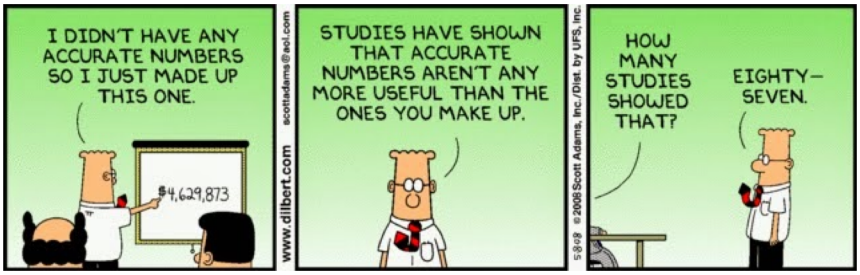
5

How many patients do I need?

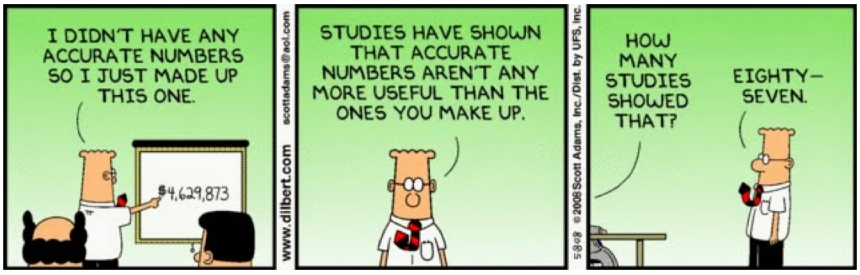
Depends on many things, for example:

- Study phase
- Superiority or non-inferiority
- Endpoint (proportion, time to event)
- Effect size
- Type I and II errors

IT'S NOT THAT SIMPLE



IT'S NOT THAT SIMPLE



Mixture of coursework/training and experience over many years.

Practical considerations

- Accrual
- Equipoise
- Strength of preliminary data
- Feasibility
- Patient population (adjuvant? metastatic?) and motivation to register
- Adverse event profile
- Funding
- Competitors
- And more…

On being a "collaborative" statistician

"More than ever before [we] require statisticians who have the insight to get at the real root causes of issues; … a broad-base of statistical AND subject material knowledge;… In short statistical knowledge is necessary but far from sufficient… the statistician must become a deeply immersed and proactively committed team member… that is getting out of one's "box" to learn about other disciplines and often performing tasks that do not directly relate to one's narrow specialty."

Hahn G, Hoerl R, "Key challenges for statisticians in business and industry." 1998 Technometrics 40:195-200.

Converting a scientific question into a trial

- There is often no single correct answer
- Context is important
  - Disease/population
  - Treatment (choice of therapy, line of therapy)
  - Enrollment potential
  - Correlative endpoints
  - Logistic support
- Ask lots of questions to avoid providing the right answer to the wrong question
  - The design usually cannot be fixed in hindsight

How many patients do I need?

- The primary objective is the only objective that determines the sample size
- The design is not just what it looks like or feels like, design is how it works (S. Jobs)
- Investigators often expect too much from trials
  - But conclusions cannot be drawn from every endpoint
  - Subset analyses (planned or unplanned) have limitations
- Arriving at the final design, after considering the interests of all parties (industry, NCI, study team) can take years
- Be peripherally aware of what you want the publication to look like

Accrual

- Sample size is partly determined by accrual rate
- Thereby impacts types of designs available to you
- Sometimes accrual is unknown (rare tumor types defined by mutations not yet studied in trials)
- Multi-center trial?
- This is not a “we think…” type of parameter
  - What has enrollment to prior trials been?
  - Why might enrollment to the proposed trial differ?

Endpoints

- Must be specified in the protocol for both clinical and laboratory studies
- Type of data collected for the evaluation of stated endpoint must be consistent with the planned analysis methodology
  - Continuous vs. categorical
  - Censored vs. uncensored
  - Binomial vs. multinomial
- Not restricted by phase of development
- Search literature for use and definition of endpoints in similar context as your trial
  - I am generally not a fan of made-up (“novel”) endpoints.

Examples of time-to-event endpoints

- Overall Survival
- Progression-free survival
- Disease-free survival
- Event-free survival
- Relapse-free survival
- Failure-free survival

Define the EVENT as well as the CENSORING time.

Examples of categorical endpoints

- Best objective response rate per RECIST 1.1
  - Modified RECIST or other criteria
  - Read the RECIST 1.1 paper!
  - Disease control rate (cytostatic agents)
- CR rate
- Toxicity rate (CTCAE version 5.0)
- Proportion alive and progression-free at 6 months
  - Consider scan interval and define what happens to censored patients

Which to choose as primary?

- Duration of study until maturity can impact this (power driven by events not the number of patients)
- Depends on mechanism of action
  - Targeted therapy studies often implement response rate in single arm setting
  - PFS is best studied in randomized setting
  - DFS is similar but evaluated in early stage disease
  - OS is the gold standard for most phase III settings

Correlative endpoints

- Often not standardized
- Timing of return of results (for stratification)
- Can require understanding of the biology of the drug-marker pairing
- Sample size justification expected even though these are not usually the primary endpoint
- May need to account for timing of sample collection
- Assay failure
- Bias re: consent for participation

Feasibility endpoints

- Apply to phase I or II studies
- Can you deliver the drug as planned
  - % pts receiving at least 80% of planned doses
  - Adherence/compliance
- If therapy cannot be delivered as defined in the protocol, you will be putting the experimental arm at a disadvantage in future studies

"Pilot" (Piantadosi 2008)

- There is no definition for this term in the literature that codifies it as a class of design.
- The term is used mainly when investigators refuse to think creatively, quantitatively, and descriptively about the requirements of their study design.
- A main purpose of the term is to deflect criticism.

Phases of development

- Phase I – first in human or first time combining therapies
- "Expansion cohorts"
- Phase II – searching for early signal of efficacy
- Phase III – hoping to change practice

- Design parameters for each of these scenarios differs, thus impacting your sample size

Design terminology and parameters

- Null hypothesis vs. alternative hypothesis
- Effect size
  - Target hazard ratio
  - Difference in proportions
- Type I error: probability rejecting the null if null hypothesis is true
- Power: probability rejecting the null if alternative hypothesis is true
- Confidence interval: range of plausible values consistent with the observed result
- Phase II vs. Phase III parameters

Phase II designs

- Simon 2-stage design
  - Single arm, binary outcome, early stopping for futility
  - Tend to have to choose a very large effect size
  - How certain are you that you would NOT have observed the result in the absence of the novel therapy?
  - Response not a very good surrogate for OS, measurements imprecise
- Randomized comparative (includes a control arm)
  - The control arm is the current standard of care
- Randomized selection design (all arms are experimental)
- A brief word about phase II/III designs: sometimes these are impractical

Phase III designs

- Randomized comparative trials
- Interim analyses
- Very well controlled type I error
- Adjustments for multiple comparisons
  - Eg, 3-arm trial
- Co-primary endpoints
- Hierarchical testing
- Gatekeeping strategy

Monitoring

- Is there a DSMB/DSMC for your study?
- Safety & efficacy monitoring specified in study design
- Group sequential methods
- Early stopping
  - Not ethical to continue enrolling or treating patients on ineffective therapy
  - Limited resources
  - Most studies are negative
  - Stopping rules for toxicity, harm, inefficacy, futility, efficacy
  - Is there a DSMB for your study?

Hypothesis testing

- New treatment/combination vs. SOC (or active control)
  - Superiority trials (generally the goal for phase II and most phase III)
    - Hope that new treatment will demonstrate improvement in your primary endpoint relative to standard of care
  - Non-inferiority trials (phase III)
    - New treatment is 'not worse' than standard by some pre-specified margin
  - Equivalency trials (phase III, rare in oncology)
    - New treatment and standard are 'similar', eg, neither better or worse

Incorporating biomarkers

- Integral biomarkers vs. integrated biomarkers
- Integral: Tests inherent in the design from the onset and must be performed in real time for the conduct of the trial
- Single marker/treatment: enrichment design
- Multiple markers/treatments:
  - Umbrella trial
  - Basket trial

Platform trials (umbrella & basket)

Assuming there's a validated assay
that can reliably assess your marker(s):

Benefits

- Screening and recruitment (low prevalence markers)
- Lower cost with shared infrastructure
- Flexibility in design
  - Expansion cohorts or add/expand # protocols
- Innovative methods for learning
- Increase speed of drug development?

Challenges

- Study development and timelines (agreement among stakeholders)
- Operational burden
- Drug / marker performance (prevalence of markers, patient willingness to participate)
- Specific statistical considerations
  - No error-control with multiple comparisons
  - Reliable endpoints

Stratification

- Randomization helps to balance impact/bias from known and unknown factors
  - Larger study / more than double the number of patients
  - False sense of "unbiasedness"
  - Likely to be over-interpreted
- Stratification is a randomization tool
- Balance treatment assignment within each prognostic subgroup
- Avoid over-stratification
- Examples: Age, PS, Stage, Histology

Prognostic vs. Predictive Factors

Prognostic Factor: Any measurement that is associated with clinical outcome in the absence of therapy, or with the application of a standard therapy that all patients are likely to receive (a predictor of the natural history of the tumor).

Predictive Factor: Any measurement associated with response or lack of response to a particular therapy, where response can be defined using any of the clinical endpoints commonly used in clinical trials (eg, EGFR mutation in NSCLC).

Correlatives: the tissue is the issue.

- Are biopsies feasible? Optional? Mandatory?
- How many biopsies?
- Plasma collection? To be paired with results from tumor?
- What do these data look like?
- Plan for some % unevaluable/missing

Other complicating factors

- Slow accrual
- Drop-out
- Non-compliance with randomization
- Evolving treatment paradigms

So, how many patients do I need?

- With experience, answering this question gets easier though it's never easy
- Sometimes the statistical design is reverse-engineered
  - Back into calculation due to constraints in resources
  - With X patients, we are able to detect a HR = {something very ambitious}
- Writing the statistical plan and protocol is an entirely different topic

33

Summary

- Designing clinical trials takes time, careful planning and thought.
- It requires good science, adherence to the scientific method, and knowledge of protocol design.
- Consult with your biostatistician early and often to preserve study integrity.