Clinical Trials: Statistical Considerations

Theresa A Scott, MS
Vanderbilt University
Department of Biostatistics
theresa.scott@vanderbilt.edu
http://biostat.mc.vanderbilt.edu/TheresaScott

NOTE: this lecture covers only some of the many statistical considerations surrounding clinical trials.

Outline

- Design:
  - Randomization
  - Blinding
  - Sample size calculation
- Analysis:
  - Baseline assessment
  - Intention-to-treat analysis.
  - Kaplan-Meier Estimator and Comparison of survival curves
  - Cox Proportional Hazards Model
- Reporting:
  - CONSORT Statement
- References
Randomization

- The process by which each subject has the same chance of being assigned to either the intervention arm or the control arm (i.e., each treatment group).

- Goals of randomization:
  - To produce groups that are comparable (i.e., balanced) with respect to known or unknown risk factors (i.e., prognostic baseline characteristics that could confound an observed association).
  - To remove bias (selection bias and accidental bias).
  - To guarantee the validity of statistical tests.
  - To balance treatment groups, stratification factors, or both.

- When to randomize?
  - After determining eligibility.
  - As close to treatment time as possible (to avoid death or withdrawal before treatment start).
Randomization, cont’d

- Randomization ‘Don’ts’:
  - Every other patient.
  - Days of the week.
  - Odd/even schemes using last digit of MRN, SSN, or date of birth.

- Randomization ‘Do’s’:
  - Formal
  - Secure
  - Reproducible
  - Unpredictable

- Two most important features:
  - (1) that the procedure truly allocates treatments randomly; and
  - (2) that the assignments are tamperproof (ie, neither intentional nor unintentional factors can influence the randomization).

Randomization methods

- Basic:
  - Simple randomization ←
  - Replacement randomization
  - Random permuted blocks (aka, Permuted block randomization) ←
  - Biased coin

- Methods for treatment balance over prognostic factors and institution:
  - Stratified permuted block randomization ←
  - Minimization
  - Stratifying by institution

- Other:
  - Pre-randomization
  - Response-adaptive randomization
  - Unequal randomization
Simple randomization

- Possible methods:
  - Toss a coin: if heads, randomize to treatment A; if tails, randomize to treatment B.
  - Generate a list of random digits: can use available tables or a computer program.
    - Example for two treatments arms (A and B),
      - Random digits 0 to 4 → A; 5 to 9 → B.
    - Example for three treatment arms (A, B, and C),
      - Random digits 1 to 3 → A; 4 to 6 → B; 7 to 9 → C; ignore if 0.
  - Do not use an alternating assignment (ie, ABABAB...).
    - No random component to this method; next assignment will be known.
  - Pro: Easy to implement.
  - Con: At any point in time, there may be imbalance in the number of subjects assigned to each treatment arm.
    - Balance improves as number of subjects increases.
  - In general, it is desirable to restrict the randomization in order to ensure similar treatment numbers (ie, balance) throughout the trial.

Permuted block randomization

- Randomization is performed in ‘blocks’ of predetermined size.
  - Ensures the number of subjects assigned to each treatment arm is not far out of balance (ie, number of subjects in treatment A = number of subjects in treatment B).
- Most common method:
  - Write down all permutations for a given block size b and the number of treatment arms.
    - Example: For b = 4 and 2 treatment arms (A and B), there are 6 permutations: AABB, ABAB, BAAB, BABA, BBAA, and ABBA.
  - For each block, randomly choose one of the permutations to assign the treatments.
  - Achieve balance between treatment arms at the end of each block of assigned subjects.
- Considerations:
  - Number of subjects in each treatment arm will never differ by more than b/2 for any b.
  - Keep b unknown to investigators and keep b fairly small (ie, number of permutations increases rapidly).
  - Do not use blocks of size 2 – easy to guess next treatment assignment.
  - Block size can be varied over time, even randomly.
Stratified permuted block randomization

- Guarantees treatment balance within prognostic factors.
  - Especially important in small trials.

- Process:
  - Define the strata (ie, the prognostic factor(s) that are most likely important).
    - Commonly used: clinical center, age, and gender.
  - Randomize within each stratum.
    - That is, for each subgroup (ie, combo of strata), perform a separate permuted block randomization.
    - Example:

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt; 40</td>
<td>ABBA, BAAB, …</td>
<td>BABA, BAAB, …</td>
</tr>
<tr>
<td>Age 41 to 60</td>
<td>BBAA, ABAB, …</td>
<td>ABAB, BBAA, …</td>
</tr>
<tr>
<td>Age ≥ 60</td>
<td>AABB, ABBA, …</td>
<td>BAAB, ABAB, …</td>
</tr>
</tbody>
</table>

- Considerations:
  - The block size should be small ($b = 2$ or $4$) to maintain balance in small strata and to ensure that the overall imbalance is not too great.
  - Increased number of stratification variables or increased number of levels with strata leads to fewer patients per stratum.
    - In small samples, sparse data in many cells defeats the purpose of stratification.
  - Stratification variables should be used in the statistical analysis.
    - Otherwise, the tests (ie, p-values) will be too conservative.
  - Very large trials (ie, >500 subjects) may not require stratification.
Implementation of randomization

- A set of standard operating procedures (SOPs) for the generation, implementation, and administration of the randomization is required to ensure the integrity of a clinical trial.

- Clinician(s) and biostatistician(s) need to discuss the selection of the randomization method and necessary information needed to generate the ‘randomization list’.

- The randomization method employed for the study should be described in detail in the study protocol without disclosure of the block size, if permuted block randomization is used.

- A formal request for the ‘randomization list’ cannot be sent unless the study protocol has been obtained and approved by all necessary IRBs.

Design > Randomization

Design

- Blinding -
Blinding

- Keeping the identity of treatment assignments masked for:
  - Subject
  - Investigator (treatment team / evaluator)
  - Monitoring committee (sponsor)

- Three possible types:
  - Single-blind: subject does not know his/her treatment assignment.
  - Double-blind: subject and investigator (treatment team / evaluator) do not know treatment assignments.
  - Triple-blind: subject, investigator (treatment team / evaluator), nor monitoring committee (sponsor) do not know treatment assignments.

- Purpose: bias reduction.
  - Each group blinded eliminates a different source of bias.

Reasons for blinding

- **Reasons for subject blinding:** if the treatment is known to the subject:
  - Those on ‘no treatment’ or standard treatment may be discouraged and drop out of the study.
  - Those on the new drug may exhibit a placebo effect (i.e., the new drug may appear better when it actually is not).
  - Subject reporting and cooperation may be biased depending on how the subject feels about the treatment.

- **Reasons for treatment team blinding:** treatment can be biased by knowledge of the treatment, especially if the treatment team has preconceived ideas about either treatment through:
  - Dose modifications,
  - Intensity of patient examinations,
  - Need for additional treatments,
  - Influence on patient attitude through enthusiasm (or not) shown regarding the treatment.
Reasons for blinding, cont’d

- Reasons for evaluator blinding:
  - If the endpoint is subjective, evaluator bias will lead to recording more favorable responses on the preferred treatment.
  - Even supposedly ‘hard’ endpoints (eg, blood pressure, MI) often require clinical judgment.

- Reasons for monitoring committee blinding:
  - Treatments can be objectively evaluated.
  - Recommendations to stop the trial for ‘ethical’ reasons will not be based on personal biases.
  - NOTE: Triple-blind studies are hard to justify for reasons of safety and ethics.

  → Although blinded trials require extra effort, sometimes they are the only way to get an objective answer to a clinical question.

Feasibility of blinding

- Ethics:
  - The double-blind procedure should not result in any harm or undue risk to a patient.
    - eg, may be unethical to give ‘simulated’ treatments to a control group.

- Practicality:
  - May be impossible to blind some treatments (eg, radiation therapy equipment is usually in constant use).
  - Requiring a ‘sham’ treatment might be a poor use of resources.

- Avoidance of bias:
  - Blinded studies require extra effort (eg, manufacturing look-alike pills, setting up coding systems, etc).
  - Consider the sources of bias to decide if the bias reduction is worth the extra effort.

- Compromise:
  - Sometimes partial blinding (eg, independent blinded evaluators) can be sufficient to reduce bias in treatment comparison.
Design

-Sample size calculation-

Preparing to calculate sample size

1. What is the main purpose of the trial?
   - The question on which sample size is based.
   - Most likely interested in assessing treatment differences.

2. What is the principal measure of patient outcome (i.e., endpoint)?
   - Continuous? Categorical? Time-to-event? Are some time-to-event values censored?

3. What statistical test will be used to assess treatment differences?
   - Eg, t-test, chi-square, log-rank?
   - What α-level are you assuming? Is your alternative hypothesis one-tailed or two-tailed?

4. What result is anticipated with the standard treatment?

5. How small a treatment difference is it important to detect and with what degree of certainty (i.e., power)?
Example scenarios...

- Desire for both is to compare drug A (standard) to drug B (new).

  **Dichotomous outcome:**
  - Want to test $H_0$: $p_{\text{standard}} = p_{\text{new}}$ vs $H_a$: $p_{\text{standard}} \neq p_{\text{new}}$, where $p_{\text{standard}}$ = proportion of failures expected on drug A and $p_{\text{new}}$ = proportion of failures on drug B.
  - If an ‘event’ is a failure, want a reduced proportion on the new drug.
  - Use PS > ‘Dichotomous’ tab.

- Continuous outcome:
  - Want to test $H_0$: mean_{standard} – mean_{new} = 0 vs $H_a$: mean_{standard} – mean_{new} ≠ 0.
  - Eq, a 10 mg/dl difference (reduction) in cholesterol on the new drug.
  - Additionally need an estimate of SD.
  - Assume all observations are known completely (ie, no censoring).
  - Assume data to be approximately normally distributed.
  - Use PS > ’t-test’ tab.

---

Example scenarios..., cont’d

- Time-to-event outcome:
  - Want to test $H_0$: median survival time on standard drug = median survival time on new drug vs $H_a$: median survival time on standard drug ≠ median survival time on new drug.
  - If an ‘event’ is a failure, want those on new drug to survive longer (ie, have large median survival time).
  - Similarly, want to test $H_0$: hazard ratio of standard drug to new drug = 1 vs $H_a$: hazard ratio of standard drug to new drug ≠ 1.
  - If an ‘event’ is a failure, want those on new drug to have a smaller hazard of having the event → hazard ratio of new drug to standard drug > 1.
  - Additionally need estimates of
    - (1) accrual time during which patients are recruited; and
    - (2) additional follow-up time after the end of recruitment.
  - General rule: accrual throughout the study period requires more patients that if all start at the beginning of the study.
  - Use PS > ‘Survival’ tab.
Additional consideration

- Adjustment for noncompliance (crossovers):
  - If assume a new treatment is being compared with a standard treatment,
    - Dropouts: those who refuse the new treatment some time after randomization and revert to the standard treatment.
    - Drop-ins: those who received the new treatment some time after initial randomization to the standard treatment.
  - Both generally dilute the treatment effect.
  - Example: (dichotomous outcome)
    - Suppose the true values are $p_{\text{drug}} = 0.6$ and $p_{\text{placebo}} = 0.4 \rightarrow \Delta = 0.6 - 0.4 = 0.2$.
    - Enroll $N = 100$ in each treatment group.
    - Suppose 25% in drug group drop out and 10% in placebo group drop in.
    - So, we actually observe $p_{\text{drug}} = (75/100)*0.6 + (25/100)*0.4 = 0.55$ and $p_{\text{placebo}} = (90/100)*0.4 + (10/100)*0.6 = 0.42 \rightarrow \Delta = 0.55 - 0.42 = 0.13$.
  - The power of the study will be less than intended, or else the sample size must be increased to compensate for the dilution effect.

Design > Sample size calculation

Analysis

- Baseline Assessment -
Baseline Assessment

- Baseline data: collected before the start of the treatment (before or after randomization).
  - Used to describe the population studied (ie, ‘Table 1’).
  - Used to stratify, or at least check for balance over prognostic factors, demographic and socioeconomic characteristics, and medical history data.
    - NOTE: While randomization on average produces balance between groups, it does not guarantee balance in any specific trial.

- Checking for treatment balance:
  - Compare the baseline data for the subjects randomized to each treatment arm.
  - Differences can be tested, but one should not conclude ‘no imbalance’ if p-values are not significant (ie, > 0.5).
    - Small sample sizes can lead to not rejecting H0 because of low power, even when imbalance is present.
  - Unless sample sizes are very large, rejecting H0 implies an imbalance problem that should be addressed in the analysis.

Analysis > Baseline Assessment

Analysis

- Intent-to-treat Analysis -
Intent-to-treat analysis

- Under the intent-to-treat (ITT) principle, all randomized subjects should be included in the (primary) analysis, in their assigned treatment groups, regardless of compliance (ie, dropout or drop-in) with the assigned treatment.

- Rationale:
  - Randomization ensures there are no systematic differences between treatment groups.
  - The exclusion of patients from the analysis on a systematic basis (eg, lack of compliance with assigned treatment) may introduce systematic difference between treatment groups, thereby biasing the comparison.

- Arguments made against ITT analysis:
  - In assessing the effect of a drug, it makes not sense to include in the treatment arm subject who didn’t get the drug, or got too little drug to have any effect.
  - The question of primary interest is whether the drugs works if used as intended.

- Food for thought: if in a randomized study an analysis is done which does not classify all patients to the groups to which they were randomized, the study can no longer be strictly interpreted as a randomized trial (ie, the randomization is ‘broken’).

Analysis

- Kaplan-Meier Estimator and Comparison of survival Curves -
Survival analysis

- Set of methods for analyzing data where the outcome is the time until the occurrence of an event of interest.
  - A.k.a. ‘time-to-event’ analysis.
  - Event often called ‘failure’ → a.k.a. analysis of failure time data.
- ‘Time-to-event’ outcome is common type of outcome in randomized clinical trials.
  - Examples: time to cardiovascular death; time to tumor recurrence; time to a response (10% decrease in BP).
- Distinguished by its emphasis on estimating the time course of events.
  - Three requirements to determine failure time precisely:
    - A time origin must be unambiguously defined.
    - A scale for measuring the passage of time must be agreed upon.
    - The meaning of failure (i.e., the definition of the event) must be entirely clear.
- Each subject’s outcome has two components:
  - (1) whether the subject experienced the event (No/Yes).
  - (2) length of time between the time origin and the occurrence of the event or censoring.

Censoring

- General idea: when the value of an observation is only partially known.
  - Subjects not followed long enough for the event to have occurred have their ‘event times’ censored at the time of last known follow-up.
- Types of censoring.
  - Right: the time to the event is known to be greater than some value.
  - Left: the time to the event is known to be less than some value.
  - Interval: the time to the event is known to be in a specified interval.
  - Type I: stop the RCT after a predetermined time (e.g., at 2 years follow-up).
  - Type II: stop the RCT after a predetermined number of events have occurred.
- Most statistical analyses assume that what causes a subject to be censored is independent of what would cause him/her to have an event.
  - If this is not the case, informative censoring is said to be present.
  - Example: if a subject is pulled off of a drug because of a treatment failure, the censoring time is indirectly reflecting a bad clinical outcome (resulting analysis will be biased).
Censoring, cont’d

- In addition to termination of the study, also result from:
  - Loss to follow-up.
  - Drop-out, which includes (for example) a patient who dies in an automobile accident before relapsing.
- Illustration of censoring:

![Illustration of censoring](image)

Kaplan-Meier (K-M) Estimator

- Procedure for estimating a survival curve (i.e., the survival function) and its standard error.
- Survival function: probability of being free of the event at a specified time.
  - Also can be thought as probability that a subject will ‘survive’ past a specified time.
  - Non-increasing, step-wise (i.e., not smooth) – a ‘step’ down for every event.
- Illustration of calculation:
  - First, order to follow-up times, either to event or censoring.
  - Second, denote by indicator variables whether the event (or censoring occurred) at each follow-up time.

<table>
<thead>
<tr>
<th>$t$</th>
<th>No. subjects at risk</th>
<th>Deaths</th>
<th>Censored</th>
<th>Cumulative survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
<td>1</td>
<td>0</td>
<td>0.962</td>
</tr>
<tr>
<td>10</td>
<td>49</td>
<td>1</td>
<td>0</td>
<td>0.962 $\times 0.962 = 0.923$</td>
</tr>
<tr>
<td>20</td>
<td>48</td>
<td>1</td>
<td>0</td>
<td>0.923 $\times 0.923 = 0.855$</td>
</tr>
<tr>
<td>30</td>
<td>47</td>
<td>1</td>
<td>0</td>
<td>0.855 $\times 0.855 = 0.846$</td>
</tr>
<tr>
<td>40</td>
<td>46</td>
<td>1</td>
<td>0</td>
<td>0.846 $\times 0.846 = 0.808$</td>
</tr>
<tr>
<td>50</td>
<td>45</td>
<td>1</td>
<td>0</td>
<td>0.808 $\times 0.808 = 0.760$</td>
</tr>
<tr>
<td>60</td>
<td>44</td>
<td>1</td>
<td>1</td>
<td>0.760 $\times 0.760 = 0.579$</td>
</tr>
<tr>
<td>70</td>
<td>43</td>
<td>0</td>
<td>1</td>
<td>0.579 $\times 0.709 = 0.401$</td>
</tr>
<tr>
<td>80</td>
<td>42</td>
<td>1</td>
<td>1</td>
<td>0.401 $\times 0.709 = 0.286$</td>
</tr>
<tr>
<td>90</td>
<td>41</td>
<td>1</td>
<td>1</td>
<td>0.286 $\times 0.709 = 0.203$</td>
</tr>
<tr>
<td>100</td>
<td>40</td>
<td>1</td>
<td>1</td>
<td>0.203 $\times 0.709 = 0.143$</td>
</tr>
</tbody>
</table>

Analysis > Kaplan-Meier Estimator and Comparison of Survival Curves
Calculated K-M estimates most often plotted, including the 95% CI.
- Hash marks (if shown) represent censoring.
- From the K-M estimates, can determine:
  - Median survival time (ie, time at which Prob(Survival) = 50%).
  - Probability of survival at a given time point (ie, 1-year survival rate).
- K-M estimates can also be calculated across a stratifying variable (eg, gender).
  - Estimate the survival function in each group of the stratifying variable (eg, a set of estimates for females and a set of estimates for males).

Comparison of survival curves

- (Non-parametric) Log-rank test: a statistical method for comparing two or more survival curves (ie, between (independent) groups).
  - $H_0$: the survival curves are equal vs $H_a$: the survival curves are not equal.
  - More formally, $H_0$: the probability of having an event in any group is equal (for all time points during the study).
    - Rejection of $H_0$ indicates that the event rates differ between groups at one or more time points during the study (ie, the probability of survival is different).
  - Falls short in the following situations:
    - When the survival curves cross at one or more points in time.
    - Stratifying variable is a continuous variable (eg, age).
    - Have more than one stratifying variable.
    - When you want to quantify the difference between the survival curves.
Comparison of survival curves, cont’d

- For plot on right, log-rank p-value = 0.303.
  - Conclusion: Failed to find a significant difference between the two survival curves.
- Plots can also include 95% CIs for each curve – can be too busy.
- Plots can also include the number of subjects at risk at specific time points – often placed in lower margin.

Analysis

- Cox Proportional Hazards Model -
Cox Proportional Hazards Model

- Widely used in the analysis of survival data to explain (ie, estimate) the effect of explanatory variables on survival times.

- Proportional Hazards (PH) Models in general:
  - Way of modeling the hazard function: the probability that the event of interest occurs at a specified time \( t \) given that it has not occurred prior to time \( t \).
    - Can be thought of as the instantaneous rate at which events occur.

- Back to the Cox PH model…
  - Widely used ‘semi-parametric’ approach that does not require the assumption of any particular hazard function or distribution of survival time data.
  - Proportional hazard assumption: requires that the ratio of hazards between any two fixed sets of covariates is constant.
    - That is, the event hazard rate may change over time, but the ratio of event hazards between two groups of individuals is constant.

Cox PH Model, cont’d

- Interpretation of model output:
  - NOTE: assumes the effect of any predictors is the same for all values of time.

  - The (raw estimated) regression coefficient for a covariate \( X_j \) is the increase/decrease in the log hazard at any time point \( t \) if \( X_j \) is increased by one unit and all other predictors are held constant.

  - The effect of increasing \( X_j \) by 1 unit is to increase/decrease the hazard of the event by a factor of \( \exp(\text{coefficient}) \) at all points in time.
    - \( \exp(\text{coefficient}) \) also interpreted as a hazard ratio.
    - Cox PH model output is most often reported in terms of estimated hazard ratios and their 95% CIs.
      - Do not report raw coefficients or p-values.
      - If 95% CI contains the value 1, then estimated hazard ration is not significant.
Cox PH Model, cont’d

- Example:
  - Estimating the effect of treatment, adjusted for age, on survival in a RCT comparing two treatments (Treatment B and Treatment A) for ovarian cancer.
  - Model output:

<table>
<thead>
<tr>
<th></th>
<th>Coeff</th>
<th>HR</th>
<th>95% CI of HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment B</td>
<td>-0.80</td>
<td>0.45</td>
<td>(0.13, 1.55)</td>
</tr>
<tr>
<td>Age</td>
<td>0.15</td>
<td>1.16</td>
<td>(1.06, 1.27)</td>
</tr>
</tbody>
</table>

- Interpretation:
  - The hazard of surviving for those who received Treatment B is 0.45 times that for those who received Treatment A.
  - The hazard of surviving is 1.16 times higher for each 1 year increase of age.
    - The hazard of surviving is 2.17 times higher (exp(5*0.15)) for each 5 year increase of age.
  - We failed to find a significant effect of both treatment and age.

Do we really need survival analysis?

- Why not use linear regression to model the survival time as a function of a set of predictor variables?
  - Time to event is restricted to be positive, which has a skewed distribution.
  - Change of interest – probability of surviving past a certain point in time.
  - Cannot effectively handle the censoring of the observations.

- OK, so why not use logistic regression to model the dichotomous event outcome (ie, model the probability of having an event)?
  - As before, interested in the probability of surviving past a certain time point.
  - As before, cannot effectively handle censoring.
  - Will have lower power - only considering whether each subject had the event, not the time until the subject possibly had the event.
    - Also, inherently assuming that the time until the event is the same for all subjects.
CONSORT Statement

- Product of CONSORT – Consolidated Standards of Reporting Trials.
  - http://www.consort-statement.org

- An evidence-based, minimum set of recommendations for reporting randomized clinical trials.
  - Offers a standard way for authors to prepare reports of trial findings, facilitating their complete and transparent reporting, and aiding their critical appraisal and interpretation.

- Comprised of a 25-item checklist and a flow diagram.
  - Checklist: focuses on reporting how the trial was designed, (conducted), analyzed, and interpreted.
    - Enables readers to assess the validity of the results.
  - Flow diagram: displays the progress of all participants through the trial.

The Checklist

- Title
  - Identification as a randomized trial in the title.
- Abstract
  - Structured summary of trial design, methods, results, and conclusions.
- Introduction
  - Scientific background and explanation of rationale.
  - Specific objectives or hypotheses.
- Methods
  - Description of trial design (such as parallel, factorial) including allocation ratio.
  - Eligibility criteria for participants.
  - Settings and locations where the data were collected (i.e., study settings).
  - The interventions for each group with sufficient details to allow replication, including how and when they were actually administered.
  - Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed.

The Checklist, cont’d

- Methods, cont’d
  - Any changes to trial outcomes after the trial commenced, with reasons.
  - How sample size was determined.
  - When applicable, explanation of any interim analyses and stopping guidelines.
  - Randomization
    - Method used to generate the random allocation sequence.
    - Type of randomization; details of any restriction (such as blocking and block size).
    - Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned.
    - Who generated the allocation sequence, who enrolled participants, and who assigned participants to interventions.
  - If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how.
  - If relevant, description of the similarity of interventions.
The Checklist, *cont’d*

- **Methods, *cont’d***
  - Statistical methods used to compare groups for primary and secondary outcomes.
  - Methods for additional analyses, such as subgroup analyses and adjusted analyses.

- **Results**
  - Participant flow diagram - for each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome (see accompanying handout).
  - For each group, losses and exclusions after randomization, together with reasons.
  - Dates defining the periods of recruitment and follow-up.
  - Why the trial ended or was stopped (ie, reason(s) for stopped trial).
  - A table showing baseline demographic and clinical characteristics for each group.
  - For each group, number of participants (denominator) included in each analysis (ie, numbers analyzed) and whether the analysis was an ‘intention-to-treat analysis’.
  - For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval).

The Checklist, *cont’d*

- **Results, *cont’d***
  - Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory.
  - All important harms or unintended effects in each group (ie, harms/adverse events).

- **Discussion**
  - Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses.
  - Generalizability (external validity, applicability) of the trial findings.
  - Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence.

- **Other Information**
  - Registration number and name of trial registry (clinicaltrials.gov).
  - Where the full trial protocol can be accessed, if available.
  - Sources of funding and other support (such as supply of drugs), role of funders.
References


• Vanderbilt’s ‘Clinical Trials’ MPH course.
  • Class notes provided by Yu Shyr, PhD.

• Additional lecture notes provided by Fei Ye, PhD and Zhigou (Alex) Zhao, MS.