STUDY DESIGNS:
FUNDAMENTALS AND INTERPRETATION

KEVIN M. SOWINSKI, PHARM.D., FCCP

PURDUE UNIVERSITY COLLEGE OF PHARMACY
INDIANA UNIVERSITY SCHOOL OF MEDICINE
WEST LAFAYETTE AND INDIANAPOLIS, INDIANA
Learning Objectives

1. Define, compare, and contrast the concepts of internal and external validity, bias, and confounding in clinical study design.
2. Identify potential sources of bias in clinical trials; select strategies to eliminate or control for bias.
3. Outline the hierarchy of evidence generated by various study designs.
4. Compare and contrast the advantages and disadvantages of various study designs (e.g., prospective; retrospective; case-control; cohort; cross-sectional; randomized controlled clinical trials; systematic review; meta-analysis). Delineate the difference between parallel and crossover study designs.
5. Select from various biostatistical measures to appropriately compare groups or their assessments from various study designs and use their findings/output to interpret results.
6. Define and evaluate odds, odds ratio, risk/incidence rate, relative risk (RR), and other risk estimates. Compute and evaluate number needed to treat and number needed to harm. Define and calculate terms such as point and period prevalence, incidence rate, prevalence rate, absolute risk difference, and RR difference.
7. Define and calculate terms such as true positive, false positive, true negative, false negative, sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio.

Self-Assessment Questions

Answers and explanations to these questions can be found at the end of the chapter.

Questions 1 and 2 pertain to the following case.
A recently released statin is associated with less myopathy than other currently available statins. After 2 years of use, a retrospective case-control study was undertaken by the manufacturer after 20 different reports of severe myopathy were sent to the U.S. Food and Drug Administration’s (FDA’s) MedWatch program. Risk factors for statin-induced myopathy were not assessed; however, both the cases and the controls of this study had identical diagnostic evaluations and were stratified according to the duration of statin use before the onset of myopathy.

1. Which type of bias is this study design most susceptible to?
   A. Confounding by indication
   B. Recall bias
   C. Diagnostic bias
   D. Misclassification

2. Which factor will be most affected by the type of bias likely to occur in this study?
   A. External validity
   B. Internal validity
   C. Assessment of exposure
   D. Number of patients needed for the study

3. When describing the results of a randomized controlled clinical trial, the investigators report using an intention-to-treat analysis to analyze their data. The results of their investigation comparing two diuretics for heart failure show no difference in the number of hospitalizations for decompensated heart failure between the treatment groups. Given their method of data analysis, which statement is most appropriate?
   A. May be susceptible to issues regarding recall bias
   B. Provides a good measure of effectiveness under typical clinical conditions
   C. Cannot provide an estimate of the method’s effectiveness
   D. May overestimate the actual treatment effect

4. The 95% confidence interval (CI) for the difference in recurrence rates between the two groups was −1.5% to 4.5%. Which conclusion is most appropriate?
   A. Twice-daily enoxaparin is superior to once daily.
   B. Superiority of twice-daily enoxaparin could not be established over once daily.
C. Once-daily enoxaparin is not inferior to twice daily.
D. No conclusion can be drawn because p-values are unavailable.

5. According to the data in the previous question and the result obtained, which best represents the number of patients who would need to be treated with twice-daily enoxaparin to prevent the recurrence of one VTE episode?
   A. Number needed to treat (NNT) would be 2
   B. NNT would be 67
   C. NNT would be 152
   D. NNT should not be calculated because the result was nonsignificant.

Questions 6 and 7 pertain to the following case.
A multicenter, double-blind, placebo-controlled trial randomly assigned 4837 patients to treatment with margarine supplemented with the omega-3 fatty acid α-linolenic acid (ALA) (margarine with ALA) or a placebo margarine. The primary combined end point was the rate of cardiovascular events, defined as fatal and nonfatal cardiovascular events and percutaneous coronary interventions. Data were analyzed according to intention-to-treat analysis with the use of a Cox proportional hazards model. The hazard ratio (HR) and 95% CI for the margarine with ALA group were 0.91 and 0.78–1.05, respectively. In the prespecified subgroup of women, margarine with ALA was associated with an HR of 0.73 (95% CI, 0.51–1.03).

6. Which statement is most appropriate?
   A. Margarine with ALA statistically significantly reduced the risk of cardiovascular events (p<0.05).
   B. Margarine with ALA statistically significantly reduced the risk of cardiovascular events (p<0.01).
   C. Margarine with ALA did not significantly reduce the risk of cardiovascular events (p>0.05).
   D. Without a p-value, it is not possible to determine whether margarine with ALA affected cardiovascular events.

7. When the study was being designed, which choice describes the outcome for which the study was most likely to have been powered?
   A. Differences in the rate of the composite outcome, cardiovascular events
   B. Differences in the rate of percutaneous coronary interventions
   C. Differences in the rate of the composite outcomes in women
   D. Differences in the rate of the composite outcomes in men

8. In a meta-analysis of studies examining the effects of several antihypertensive drugs, the odds ratio (OR) for treatment with low-dose diuretics compared with calcium channel blockers for cardiovascular disease events was 0.84 (95% CI, 0.75–0.95). Which statement is the most appropriate interpretation of these findings?
   A. Treatment of hypertension with low-dose diuretics was more effective in preventing cardiovascular disease events than treatment with calcium channel blockers.
   B. Treatment of hypertension with calcium channel blockers was more effective in preventing cardiovascular disease events than treatment with low doses of diuretics.
   C. The difference observed between treatment with calcium channel blockers and low doses of diuretics is not statistically significant.
   D. The odds of developing cardiovascular events when treating hypertension with low doses of diuretics are lower than when using calcium channel blockers.
I. INTRODUCTION

A. Why Do You Need to Know About Study Design and Interpretation? Ambulatory Care Pharmacy Specialty Examination Content Outline.
   1. Translation of evidence into practice (14%) (Domain 3)
   2. Task statements
      a. Retrieve biomedical literature applicable to ambulatory care pharmacy practice.
      b. Interpret biomedical literature with regard to study design methodology, statistical analysis, and significance and applicability of reported data and conclusions.
      c. Respond to requests for information from patients and health care professionals by using evidence-based literature.
      d. Use the principles and strategies of project and research design to generate and disseminate information in ambulatory care.
      e. Enlist evidence-based strategies to effectively teach students, residents, pharmacists, and other health care professionals.

B. Examples of Online Statistical and Study Design Tools
   1. www.graphpad.com/quickcalcs/

II. VARIOUS ISSUES IN STUDY DESIGN

A. Research Design Classification
   1. Study purpose: Descriptive versus analytic
   2. Time orientation: Prospective versus retrospective design
      a. Prospective: Begin in the present and progress forward, collecting data from subjects whose outcomes lie in the future.
      b. Retrospective: Begin and end in the present; however, this design involves a major backward look to collect information about events that occurred in the past.
   3. Investigator orientation: Interventional versus quasi-experimental
   4. Experimental setting
      a. Randomized controlled trials
      b. Observational trials

B. Relative Strength of Evidence: Hierarchy of Study Designs
C. Validity in Study Design
   1. Internal validity
      a. Validity within the confines of the study methods
      b. Does the study design adequately and appropriately test/measure what it purports?
      c. Does the study adequately and appropriately address bias, confounding, and measurement of endpoints?
   2. External validity
      a. Validity related to generalizing the study results outside the study setting
      b. Can the results be applied to other groups, patients, or systems?
      c. Addresses issues of generalizability and representativeness

D. Bias in Study Design
   1. Definition: Systematic, nonrandom variation in study methodology and conductance, ultimately introducing error in outcome interpretation. Bias can occur in all aspects of the study design.
   2. Examples of bias
      a. Selection bias: An error in the selection of or sampling of individuals for a clinical study. Classic example: Subjects chosen for the case and control groups differ in one or more characteristics that alter the outcome of a study.
      b. Observational or information bias: An error in the recording of individual factors of a study, such as inaccurate recording of a patient’s risk factor, inaccurate recording of the timing of a blood sample
      c. Recall bias: Classic example: Studies of birth defects secondary to medications
      d. Interviewer bias: Classic example: Interviews are not conducted in a uniform manner (or by the same person) for all study participants.
      e. Misclassification bias
         i. Differential
         ii. Non-differential
   3. Controlling for bias
      a. Design: For example, selection of study population
      b. Means of collecting data
c. Sources of information (regarding disease and exposure)
d. Analysis: May be difficult to interpret

E. Confounding in Study Design
   1. A variable that affects the independent or dependent variable, altering the ability to determine the true effect on the measured outcome. These factors may hide or exaggerate a true association.
   2. To minimize the potential for missing a confounding variable, all relevant information should be collected and evaluated.
   3. Controlling for confounding
      a. During the design of a study
         i. Randomization
         ii. Restriction
         iii. Matching
      b. Analysis
         i. Stratification
         ii. Multivariate analysis

F. Causality
   1. Temporality: Cause before effect
   2. Strength: Plausibility increases with strength of relationship
   3. Biological gradient: Dose-response?
   4. Consistency: Observations for several settings
   5. Specificity: Single cause for effect
   6. Plausibility: Biologically plausible
   7. Coherence: Consistency with existing knowledge
   8. Analogy: Preclinical expectation applied to clinical testing
   9. Experiment: Randomized controlled trials

III. CASE REPORTS/CASE SERIES

A. Document and Describe Experiences, Novel Treatments, and Unusual Events. Allows hypothesis generation that can be tested with other study designs. Note that the title does not state “study.”
   1. Possible adverse drug reactions in one or more patients: QT-interval prolongation associated with fluoroquinolone antibiotics
   2. Case report: One patient
   3. Case series: More than one patient with a similar experience or many case reports combined into a descriptive review
   4. Reports should provide sufficient detail to allow readers to recognize same/similar cases at their center/practice.

B. Advantages and Disadvantages
   1. Advantages: Hypotheses are formed, which may be the first step in describing an important clinical problem. Easy to perform and inexpensive.
   2. Disadvantages: Does not provide explanation other than conjecture and does not establish causality or association
IV. OBSERVATIONAL STUDY DESIGNS

A. Design Does Not Involve Investigator Intervention, Only Observation. It is essential to remember that observational study designs investigate associations—not, in most cases, causes.

B. Case-Control Study: Study Exposure in Those With and Without the Outcome of Interest

![Figure 2. Case-control study design.](image)

1. Determine the association between exposures/risk factors and disease/condition. Classic example: Aspirin use and Reye syndrome
2. Retrospective studies
3. Useful method (and perhaps the only practical way) to study exposures in rare diseases or diseases that take long periods to develop
4. Critical assumptions to minimize bias
   a. Cases are selected to be representative of those who have the disease.
   b. Controls are representative of the general population that does not have the disease and are as identical as possible to the cases, minus the presence of the disease.
   c. Information is collected from cases and controls in the same way.
5. Examples
   Purpose of study:
   i. To estimate in women the association between hemorrhagic stroke and the use of appetite suppressants containing PPA
   ii. To estimate the association between any use of PPA (in appetite suppressant or cough or cold remedy) and hemorrhagic stroke
iii. To estimate in men and women the association between hemorrhagic stroke and the type of exposure to PPA
iv. Disease: Hemorrhagic stroke (several types). Exposure: PPA
v. Cases: Symptomatic subarachnoid or intracerebral hemorrhage (n=702). Controls: Matched by sex, race, and age (n=1376)
vi. Exposure assessed by structured questionnaire, product photographs, and ingredient confirmation

6. Advantages
   a. Inexpensive and can be conducted quickly
   b. Allows investigation of several possible exposures or associations

7. Disadvantages
   a. Confounding must be controlled for.
   b. Observational and recall bias: Looking back to recall exposures and their possible levels of exposure
   c. Selection bias: Case selection and control matching are difficult.

8. Measure of association: OR (odds ratio): In some cases, this can be an estimate of the relative risk/risk ratio (RR). The OR is interpreted as the odds of exposure to a factor in those with a condition or disease compared with those who do not have the condition or disease. Interpretation of these concepts will be presented in the following text.

C. Cohort Study
1. Determines the association between exposures/factors and disease/condition development. Allows an estimation of the risk of outcome (and the RR between the exposure groups). Study outcome of interest in those with and without the exposure of interest. Classic examples follow:
   a. Framingham Study. A “cohort” of subjects from Framingham, Massachusetts, were (and are) studied over time to evaluate the relationship between a variety of conditions (exposures) on the development of cardiovascular disease.
   b. Nurses’ Health Study: Investigated the potential long-term consequences of the use of oral contraceptives
2. Describes the incidence or natural history of a disease/condition and measures it in time sequence
3. “Retrospective” (historical): Begins and ends in the present but involves a major backward look to collect information about events that occurred in the past

![Figure 3. “Retrospective” (historical) cohort study design.](image-url)
a. Advantages: Less expensive and time consuming; no loss to follow-up, ability to investigate issues not amenable to a clinical trial or ethical or safety issues
b. Disadvantages: Only as good as the data available, little control of confounding variables through nonstatistical approaches, recall bias

4. Prospective or longitudinal: Begin in the present and progress forward, collecting data from subjects whose outcomes lie in the future

Figure 4. Prospective cohort study design.

a. Example: Prospective, observational study: Postmenopausal hormone use and secondary prevention of coronary events in the Nurses’ Health Study (Ann Intern Med 2001;135:1-8)
b. Advantages: Can control for confounding factors to a greater extent, easier to plan for data collection
c. Disadvantages: More expensive and time intensive, loss of subject follow-up, difficult to study rare diseases/conditions at a reasonable cost

5. Measure of association: RR: The risk of an event or development of a condition relative to exposure; the risk of someone developing a condition when exposed compared with someone who has not been exposed

D. Cross-sectional (a.k.a. prevalence study)
1. Identify the prevalence or characteristics of a condition in a group of individuals.
2. Examples
   b. Cross-sectional analysis of data from a large cohort study: Maternal characteristics and migraine pharmacotherapy during pregnancy (Cephalgia 2009;29:1267-76)
3. Advantages: Easy design, “snapshot in time,” all data collected at one time, studies are accomplished by questionnaire, interview, or other available biomedical information (e.g., laboratory values).
4. Disadvantages: Does not allow the study of a factor (or factors) in individual subjects over time, just at the time of assessment; difficult-to-study, rare conditions
V. INCIDENCE, PREVALENCE, RELATIVE RISK/RISK RATIOS, AND ODDS RATIOS

A. Incidence
1. Measure of the probability of developing a disease
2. Incidence rate: Number of new cases of disease per population in a specified time
3. Calculated by dividing the number of individuals who develop a disease during a given period by the number of individuals who were at risk of developing a disease during the same period

B. Prevalence
1. Measure of the number of individuals who have a condition/disease at any given time
2. Point prevalence: Prevalence on a given date
3. Period prevalence: Prevalence in a period (e.g., year, month)

C. Interpreting RR/ORS
1. Estimate the magnitude of association between exposure and disease. Key point: For observational studies, this is not cause and effect; it is an association.
2. The incidence of disease in the exposed group divided by the incidence of disease in the unexposed group
3. The RR (a.k.a. risk ratio) cannot be directly calculated for most case-control studies; instead, the OR is usually an estimate of the RR.
4. The RR and OR are interpreted on the basis of their difference from unity (1.0). If the 95% CI includes unity, no statistical difference is indicated. The CI also gives us an idea of the spread within which the true effect lies.
5. Interpretation of the index of risk
   a. Direction of risk

Table 1. Direction of Risk Associated with OR and RR

<table>
<thead>
<tr>
<th>RR</th>
<th>OR</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>Negative association</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RR: Risk of disease is lower in the exposed group</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR: Odds of exposure is lower in the diseased group</td>
</tr>
<tr>
<td>= 1</td>
<td>= 1</td>
<td>No association</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RR: Risk of disease in the two groups is the same</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR: Odds of exposure in the two groups is the same</td>
</tr>
<tr>
<td>&gt; 1</td>
<td>&gt; 1</td>
<td>Positive association</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RR: Risk of disease is greater in the exposed group</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR: Odds of exposure is greater in the diseased group</td>
</tr>
</tbody>
</table>

OR = odds ratio; RR = relative risk/risk ratio.

   b. Magnitude of risk
Table 2. Magnitude of Risk Associated with OR and RR

<table>
<thead>
<tr>
<th>RR</th>
<th>OR</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>0.75</td>
<td>25% reduction in the risk/odds</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>No difference in risk/odds</td>
</tr>
<tr>
<td>1.5</td>
<td>1.5</td>
<td>50% increase in the risk/odds</td>
</tr>
<tr>
<td>3.0</td>
<td>3.0</td>
<td>3-fold (or 200%) increase in the risk/odds</td>
</tr>
</tbody>
</table>

OR = odds ratio; RR = relative risk/risk ratio.

6. Calculating RR/OR/contingency tables

Table 3. Contingency Table for Estimating RR and OR

<table>
<thead>
<tr>
<th>Exposure?</th>
<th>Disease?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>A</td>
</tr>
<tr>
<td>No</td>
<td>C</td>
</tr>
</tbody>
</table>

a. RR = \([A/(A + B)]/[C/(C + D)]\)
b. OR = \((A/C)/(B/D)\) or \(= (A \times D)/(B \times C)\)


Table 4. Contingency Table

<table>
<thead>
<tr>
<th>Exposure?</th>
<th>Hemorrhagic Stroke in Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appetite suppression use</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>6</td>
</tr>
<tr>
<td>No</td>
<td>377</td>
</tr>
</tbody>
</table>


a. OR = \((6/377)/(1/749) = 12\)
b. Data from the PPA study as stated previously related to appetite suppressant and development of hemorrhagic stroke

Table 5. Use of PPA and Appetite Suppressants and the Risk of Developing Hemorrhagic Stroke

<table>
<thead>
<tr>
<th></th>
<th>Cases (+ hemorrhagic stroke) n=383</th>
<th>Controls (- hemorrhagic stroke) n=750</th>
<th>Adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appetite suppressant: Women</td>
<td>6</td>
<td>1</td>
<td>16.6 (1.51–182)</td>
</tr>
<tr>
<td>Appetite suppressant: Men</td>
<td>0</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Appetite suppressant: Either</td>
<td>6</td>
<td>1</td>
<td>15.9 (1.38–184)</td>
</tr>
<tr>
<td>PPA: Women</td>
<td>21</td>
<td>20</td>
<td>1.98 (1.00–3.90)</td>
</tr>
<tr>
<td>PPA: Men</td>
<td>6</td>
<td>13</td>
<td>0.62 (0.20–1.92)</td>
</tr>
<tr>
<td>PPA: Either</td>
<td>27</td>
<td>33</td>
<td>1.49 (0.84–2.64)</td>
</tr>
</tbody>
</table>

CI = confidence interval; OR = odds ratio; PPA = phenylpropanolamine.
c. What do these numbers mean?
d. Can you interpret the point estimate and 95% CI in all cases?
   i. What does the point estimate mean?
   ii. What does the CI mean?
   iii. Which ones are statistically significant?

D. Causation
1. REMEMBER: In general, we do not prove or show causality with observational studies, but there is
   some general “guidance” to consider when evaluating them. It is important to recognize that, in many
   situations, the conduct of studies to establish causality is not possible, practical, or ethical.
2. Types of causality
   a. Sufficient cause
   b. Necessary cause
   c. Risk factor
3. Questions used to evaluate causality
   a. Was statistical significance observed?
   b. What was the strength of the association, as measured by the OR or the RR?
   c. Were dose-response relationships evaluated?
   d. Was there a temporal relationship between exposure and disease/outcome?
   e. Have the results been consistently shown?
   f. Is there biologic plausibility to the association?
   g. Is there any experimental (e.g., animal, in vitro) evidence?

VI. RANDOMIZED CONTROLLED TRIAL DESIGN

A. Characteristics
1. Experimental or interventional, investigator makes intervention and evaluates cause and effect. Examine
   etiology, cause, efficacy, using comparative groups.
2. Some previous background information or studies should exist to suggest that the intervention used will
   likely be beneficial.
3. Design allows assessment of causality.
   a. Sufficient cause
   b. Necessary cause
   c. Risk factor
4. Minimizes bias through randomization and/or stratification
   a. Randomization
   b. Block randomization
   c. Stratification
   d. Cluster randomization
5. Treatment controls
   a. Placebo controlled
   b. Active controlled
   c. Historical control
6. Blinding methods
   a. Single-blind: Either subjects or investigators are unaware of subject assignment to active/control.
   b. Double-blind: Both subjects and investigators are unaware of subject assignment to active/control.
c. Triple-blind: Both subjects and investigators are unaware of subject assignment to active/control; in addition, an analysis group is unaware.
d. Double-dummy: Two placebos necessary to match active and control therapies
e. Open-label: Everyone is aware of subject assignment to active/control.

7. May use parallel or crossover design (see additional information in the following text)
a. Crossover provides practical and statistical efficiency.
b. Crossover is not appropriate for certain types of treatment questions (e.g., effect of treatment on a disease that worsens quickly over time or worsens during the study period).

8. Factorial design: Designed to answer two separate research questions in a single group of subjects

9. Examples
a. Clinical trial: Comparison of two drugs, comparison of two behavioral modifications, etc.
b. Educational intervention: Online course versus lecture class format
c. Health care intervention: Pharmacist-based health care team versus non–pharmacist-based health care team

B. Randomized Controlled Trial: Parallel Design

C. Randomized Controlled Trial: Crossover Design

**Figure 5.** Randomized controlled trial: parallel design.

C. Randomized Controlled Trial: Crossover Design

**Figure 5.** Randomized controlled trial: parallel design.
D. Examples of Considerations for Controlled Trials
1. Are the results of the study valid (methods)?
   a. Did the subjects undergo randomization, and what was the randomization technique? Did the randomization process result in equal baseline characteristics?
   b. Were all subjects who entered the trial accounted for? Was follow-up complete? If not, how many were lost to follow-up, from which groups did they leave, and why?
   c. Were subjects analyzed in the groups to which they were randomized? Was intention-to-treat, per-protocol, or actual treatment analysis used?
   d. How was blinding conducted (e.g., subject, investigator), if applicable?
   e. Were the inclusion and exclusion criteria appropriate, or were they too restrictive or inclusive? Were the groups similar at the start of the trial?
   f. Was the sample size sufficient, and was a power calculation included?
   g. Were the groups handled the same way, aside from the intervention(s)?
   h. Were the statistical tests appropriate and understandable?
   i. What was assessed: Surrogate markers or true outcomes? Were \textit{a priori} subgroup analyses performed?
2. What were the results?
   a. How large was the treatment effect?
   b. How precise was the effect (based on CIs significant)?
   c. Did the authors properly interpret the results?
3. Can I apply the results of this study to my patient population? Will they help me care for my patients?
   a. Can the results of this study be applied to general practice?
   b. Was a representative population studied? Can I apply this to my setting?
   c. Do the patients I care for fulfill the enrollment criteria for this study?
   d. Do the patients I care for fulfill the subgroup criteria evaluated?
   e. Do the expected benefits outweigh the expected and/or unanticipated risks?
VII. OTHER ISSUES TO CONSIDER IN CONTROLLED TRIALS

A. Subgroup Analysis
1. Important part of controlled clinical trials (if set a priori)
2. Many times, they are overused and overinterpreted, leading to unnecessary research, misinterpretation of results, and/or suboptimal patient care.
3. Many potential pitfalls in identifying and interpreting
   a. Failure to consider several comparisons or to adjust p-values
   b. Problems with sample size (power), classification, and lack of assessment of interaction

B. Composite End Points: Often, the impression is that this practice is not a good practice.
1. The primary end point is one of the most important decisions to make in the design of a clinical study.
2. A composite end point combines several end points.
   a. For example, cardiovascular death, nonfatal myocardial infarction (MI), and cardiac arrest with resuscitation
   b. Usually combines measures of morbidity and mortality
   c. What does the following statement mean? Our findings show that ramipril reduces the rates of death, MI, stroke, revascularization, cardiac arrest, heart failure, complications related to diabetes, and new cases of diabetes in a broad spectrum of high-risk patients. Treating 1000 patients with ramipril for 4 years prevents about 150 events in around 70 patients.
      i. Was there a reduction in all the end points or just in some?
      ii. Are all the outcomes just as likely to occur?
      iii. Why would the investigators of this trial have been interested in all of these outcomes?
3. What are the positives for using composite end points?
   a. No single primary outcome
   b. To alleviate problems of multiple testing
   c. To increase number of events, which decreases sample size and cost to the investigator
4. What are the problems?
   a. Difficulties in interpreting composite end points; consider our earlier example
   b. Misattribution of statistically beneficial effects of composite measure to each of its component end points
   c. Dilution of effects, negative results for relatively common component of composite end point “hide” real differences in other end points. Undue influence exerted on composite end point by “softer” component end points.
   d. “Averaging” of overall effect: Problems when component end points move in opposite directions; a sign the composite end point should be abandoned without valid conclusions being drawn
   e. Should all end points be weighed the same, or should death “weigh” more?
5. The results for each individual end point should be reported together with the results for the composite.

C. Surrogate End Points
1. Variables thought to be associated with clinical outcomes
   a. Blood pressure and stroke prevention
   b. Low-densitylipoprotein cholesterol (LDL-C) reduction and cardiovascular death reduction
      i. Statins: Yes
      ii. Hormone replacement therapy: No
   c. Premature ventricular contraction suppression and reduced mortality
2. Surrogate outcomes do not always predict clinical outcomes.
3. Short-duration studies that evaluate surrogate end points may not be large enough to detect uncommon adverse events.
D. Superiority Versus Equivalence Versus Noninferiority
1. A superiority trial is designed to detect a difference between experimental treatments. This is the typical design in a clinical trial.
2. An equivalence trial is designed to confirm the absence of a meaningful difference(s) between treatments, neither better nor worse (both directions). The key is the definition of the specified margin. What difference is important? One example is a bioequivalence trial.
3. A noninferiority trial is designed to investigate whether a treatment is not clinically worse (not less effective than stated margin, or inferior) than an existing treatment.
   a. It may be the most effective, or it may have a similar effect.
   b. Useful when placebo administration is not possible for ethical reasons
   c. ONTARGET (The Ongoing Telmisartan Alone and in Combination with Ramipril Global Endpoint Trial)
      i. Designed to evaluate telmisartan, ramipril, or their combination in patients with a high risk of vascular disease
      ii. Objective was to determine whether telmisartan was noninferior to ramipril in the incidence of cardiovascular deaths.
      iii. Noninferior difference was defined as 13% or less.
   d. Essentials of noninferiority design
      i. Control group must be effective.
      ii. Current study similar to previous study with control and with equal doses, clinical conditions, and design used
      iii. Adequate power is essential, and, usually, larger sample sizes are required.

VIII. CONTROLLED CLINICAL TRIALS: ANALYSIS
A. Controlled Clinical Trial: Application (JAMA 1998;280:605-13)
1. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease (CHD) in postmenopausal women
2. Objective: To determine whether estrogen plus progestin therapy alters the risk of CHD in postmenopausal women with established CHD
3. Randomized, blinded, placebo controlled
   a. Two treatment arms: ERT-P (conjugated equine estrogen 0.625 mg/day plus medroxyprogesterone acetate 2.5 mg/day) and placebo—n=2763 with coronary artery disease, younger than 80 years; mean age 66.7 years
   b. Follow-up averaged 4.1 years; 82% of patients undergoing hormone replacement therapy still taking hormone at the end of 1 year; 75% at the end of 3 years
4. End points
   a. Primary: Nonfatal MI or CHD death
   b. Secondary: Many, including all-cause mortality. Are these composite outcomes appropriate?
5. Statistical analysis
   a. Baseline characteristics: t-test and c²: Is comparing baseline characteristics necessary in this type of trial?
   b. Power analysis and sample size calculation
   c. Kaplan-Meier with Cox proportional hazards model
6. Surrogate end point: LDL-C lowered
7. Results:

Table 6. Death and Secondary End Points by Treatment Group (Overall)

<table>
<thead>
<tr>
<th></th>
<th>ERT-P</th>
<th>Placebo</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary CHD events</td>
<td>12.5%</td>
<td>12.7%</td>
<td>0.99 (0.80–1.22)</td>
</tr>
<tr>
<td>CHD death</td>
<td>5.1%</td>
<td>4.2%</td>
<td>1.24 (0.81–1.75)</td>
</tr>
<tr>
<td>Any thromboembolic event</td>
<td>2.5%</td>
<td>0.9%</td>
<td>2.89 (1.50–5.58)</td>
</tr>
<tr>
<td>Gallbladder disease</td>
<td>6.1%</td>
<td>4.5%</td>
<td>1.38 (1.00–1.92)</td>
</tr>
</tbody>
</table>

CHD = coronary heart disease; CI = confidence interval; ERT-P = conjugated equine estrogen. 0.625 mg/day plus medroxyprogesterone acetate 2.5 mg/day; HR = hazard ratio.

8. Significant time trend: More CHD events in the treatment group than in placebo in year 1 and fewer in years 4 and 5

9. Author's conclusions
   a. During follow-up, ERT-P did not reduce overall rate of CHD events.
   b. Treatment increased rate of thromboembolic events and gallbladder disease.

B. Questions to Consider in Evaluating and Interpreting a Clinical Trial

1. Study design
   a. Was the studied sample representative of the population or the individual to whom the results were being applied?
   b. Were the inclusion/exclusion criteria appropriate, or were they overly restrictive or inclusive?
   c. Sufficient sample size, power, and so forth? Was a power analysis included?
   d. Was a study objective and/or hypothesis provided?
   e. Was the study blinded and to whom? (Subject, investigator, study personnel, or all?)
   f. Was a run-in phase used? If so, why? Did it affect the interpretation of the trial?
   g. What type of randomization method was performed? Did the randomization process produce equal baseline characteristics between all groups?

2. Outcomes/assessments
   a. Were the primary and/or secondary outcomes identified, were they reasonable, and did they apply to clinical practice?
   b. Was a composite outcome used, and were all the individual components identified and clearly stated in the methods and results?
   c. Were surrogate markers used instead of (or in addition to) clinically relevant outcomes?

3. Analysis
   a. What analysis technique was used: Intention to treat, actual treatment, or per protocol?
   b. Were the statistical tests appropriate?

4. Interpretation: Was the author’s interpretation appropriate and within the confines of the study design?

5. Extrapolation
   a. Are you applying the results to similar patients in a similar setting?
   b. Are there possible additional adverse effects that were not measured in this study?
IX. COMMON APPROACHES TO ANALYZING CLINICAL TRIALS

A. Intention-to-Treat Analysis
   1. Compares outcomes on the basis of initial group assignment or “as randomized.” The allocation to
groups was how they were “intended to be treated,” even though they may not have taken the medication
for the duration of the study, dropped out, and did not comply with the protocol.
   2. Determines effect of treatment under usual conditions of use. Analogous to routine clinical practice in
which a patient receives a prescription but may not adhere to the prescribed drug regimen.
   3. Gives a conservative estimate of differences in treatments; may underestimate treatment benefits
   4. Most common approach to assessing clinical trial results
   5. This is the preferred type of analysis in a superiority trial.

B. Per-Protocol Analysis
   1. Subjects who do not adhere to allocated treatment are not included in the final analysis; only those who com-
   pleted the trial and adhered to the protocol (based on some predetermined definition [e.g., 80% adherence]).
   2. Provides additional information about treatment efficacy and provides more generous estimates of
differences between treatments
   3. Subject to several issues because of factors such as lower sample size and definitions of adherence.
   Results are more difficult to interpret and would be validly applied only to adherent patients like those
   in the trial; not necessarily generalizable to all patients.

C. As-Treated Analysis
   1. Subjects are analyzed by the actual intervention received. If subjects were in the active treatment group
but did not take active treatment, the data would be analyzed as if they were in the placebo group.
   2. This analysis essentially ignores/destroys the randomization process for those who did not adhere to the
study design. Results should be interpreted with caution.

X. SYSTEMATIC REVIEW/META-ANALYSIS

A. Introduction
   1. Dramatic increase in the number of these types of papers
   2. First meta-analysis probably published in 1904: Assessment of typhoid vaccine effectiveness

B. Systematic Review
   1. Summary that uses explicit methods to perform a comprehensive literature search, critically appraise it,
   and synthesize the world literature on a specific topic
   2. Differs from a standard literature review: The study results are more comprehensively synthesized and
   reviewed.
   3. As with a controlled clinical trial (or other studies), the key is a well-documented and well-described
   systematic review.
   4. Some systematic reviews will attempt to statistically combine results from many studies.
   5. Differs from other reviews, which combine evaluation with opinions

C. Meta-analysis
   1. Systematic review that uses mathematical/statistical techniques to summarize the results of the evaluated
   studies
2. These techniques may improve on the following:
   a. Calculation of effect size
   b. Increase in statistical power
   c. Interpretation of disparate results
   d. Reduction in bias
   e. Answers to questions that may not be addressable with individual studies

3. Reliant on criteria for inclusion of previous studies and statistical methods to ensure validity. Details of included studies are essential.

4. Elements of trial methodology
   a. Research question
   b. Identification of available studies
   c. Criteria for trial inclusion/exclusion
   d. Data collection and presentation of findings
   e. Calculation of summary estimate: Ideally with Forest plot (Figure 7)
   f. Assessment of heterogeneity
      i. Statistical heterogeneity
      ii. $\chi^2$ and Cochran Q are common tests for heterogeneity.
      iii. Assessment of publication bias: Funnel plot
   g. Sensitivity analysis

### Table 1

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Lactobacillus</th>
<th>Placebo</th>
<th>Risk Ratio M-H Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.3.1 Age &gt; 18yo</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ahuja, M. (2002)</td>
<td>0 545 28 196 3.95</td>
<td>0.01 [0.00, 0.11]</td>
<td></td>
</tr>
<tr>
<td>Armazi, A. (2001)</td>
<td>1 30 8 30 6.1%</td>
<td>0.13 [0.02, 0.94]</td>
<td></td>
</tr>
<tr>
<td>Beausoleil, M. (2007)</td>
<td>7 44 15 45 12.7%</td>
<td>0.45 [0.20, 0.95]</td>
<td></td>
</tr>
<tr>
<td>Cremonini, F. (2002)</td>
<td>1 21 5 21 6.0%</td>
<td>0.17 [0.02, 1.27]</td>
<td></td>
</tr>
<tr>
<td>Golz, V. (1979)</td>
<td>3 43 9 36 9.9%</td>
<td>0.25 [0.08, 0.95]</td>
<td></td>
</tr>
<tr>
<td>Thomas, M. (2001)</td>
<td>33 133 40 134 14.9%</td>
<td>0.95 [0.68, 1.42]</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>816 451 53.5%</td>
<td>0.24 [0.08, 0.75]</td>
<td></td>
</tr>
<tr>
<td><strong>Total events</strong></td>
<td>51 105</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Heterogeneity: Tau² = 1.41; $\chi^2$ = 30.19; df = 5 (P < 0.0001) P = 83%
| Test for overall effect: Z = 2.46 (P = 0.01) |
| **1.3.2 Age < 18yo** |               |         |                             |
| Arvela, T. (1999) | 3 61 9 58 9.7% | 0.32 [0.09, 1.11] |
| Ruszczynski, M. (2008) | 3 120 9 120 9.6% | 0.33 [0.09, 1.20] |
| Tankanow, R. (1990) | 10 15 15 23 14.5% | 0.95 [0.61, 1.50] |
| Vanderhoof, J. (1999) | 7 93 25 95 12.7% | 0.29 [0.13, 0.63] |
| Subtotal (95% CI) | 289 216 46.5% | 0.44 [0.18, 1.08] |
| **Total events** | 23 59 |         |                             |
| Heterogeneity: Tau² = 0.59; $\chi^2$ = 12.63; df = 3 (P = 0.006) P = 76%
| Test for overall effect: Z = 1.79 (P = 0.01) |
| **Total (95% CI)** | 1105 757 100.0% | 0.35 [0.19, 0.67] |
| **Total events** | 74 164 |         |                             |
| Heterogeneity: Tau² = 0.66; $\chi^2$ = 42.93; df = 9 (P < 0.0001) P = 79%
| Test for overall effect: Z = 3.22 (P = 0.001) |
| Test for subgroup differences: Not applicable |

**Figure 7.** Forest plots.

CI = confidence interval; yo = years old.


D. Meta-analysis: Additional example (Arch Intern Med 2008;168:687-94)
XI. SUMMARY MEASURES OF EFFECT

A. Absolute and Relative Differences
   1. Absolute differences or absolute changes
   2. Relative differences or relative changes
   3. Absolute differences are more important than relative differences, although the authors of many clinical studies highlight the differences observed in trials with relative differences because they are numerically larger. Why? Larger numbers are more convincing to practitioners and patients. Most drug advertisements (both directly to patients and to health care professionals) quote relative differences.

B. Number Needed to Treat (NNT)
   1. Characteristics
      a. Another means to characterize changes or differences in absolute risk
      b. Definition: The reciprocal of the absolute risk reduction (ARR)
         i. NNT = 1/(ARR).
         ii. Rounded to the next highest whole number is the most conservative approach
      c. Applied to clinical outcomes with dichotomous data (e.g., yes/no, alive/dead, MI/no MI)
      d. Caution: Assumes the baseline risk is the same for all patients (or that it is unrelated to RR)
      e. Extrapolation beyond studied time points
      f. NNTs should be provided only for statistically significant effects.
      g. Number needed to harm
   2. NNT application
      b. Study evaluated the effect of ramipril on cardiovascular events in high-risk patients.
      c. Prospective randomized double-blind study
         i. 9297 high-risk patients received ramipril or matching placebo once daily for an average follow-up of 5 years.
         ii. Primary outcome: Composite of MI, stroke, or death from cardiovascular causes
      d. Results (data taken from previously referenced article). NNTs = 1/(0.178 − 0.140) = 1/0.038 = 26.3, rounded up to 27.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Ramipril, %</th>
<th>Placebo, %</th>
<th>Relative Risk</th>
<th>RRR</th>
<th>ARR</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>14.0</td>
<td>17.8</td>
<td>0.79</td>
<td>0.21</td>
<td>0.038</td>
<td>27</td>
</tr>
<tr>
<td>Death from CV causes</td>
<td>6.1</td>
<td>8.1</td>
<td>0.74</td>
<td>0.25</td>
<td>0.02</td>
<td>50</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>9.9</td>
<td>12.3</td>
<td>0.80</td>
<td>0.20</td>
<td>0.024</td>
<td>42</td>
</tr>
<tr>
<td>Stroke</td>
<td>3.4</td>
<td>4.9</td>
<td>0.68</td>
<td>0.31</td>
<td>0.015</td>
<td>67</td>
</tr>
</tbody>
</table>

ARR = absolute risk reduction; CV = cardiovascular; NNT = number needed to treat; RRR = relative risk reduction.

   e. Online calculator: http://araw.mede.uic.edu/cgi-bin/nntcalc.pl
   C. OR to NNT calculator: http://kteclearinghouse.ca/cebm/practise/ca/calculators/
XII. REPORTING GUIDELINES FOR CLINICAL STUDIES

A. The Consolidated Standards of Reporting Trials (CONSORT)
   1. Initially published in 1996 and updated several times since—most recently, in 2010
   2. Created in an effort to improve, standardize, and increase the transparency of the reporting of clinical trials and to facilitate the improvement of literature evaluation
   3. Available at www.consort-statement.org/
   4. The CONSORT statement has been endorsed by several publications and published in these journals.
   5. The CONSORT statement
      a. The checklist: 25-item checklist pertaining to the content of the following:
         i. Title
         ii. Abstract
         iii. Introduction
         iv. Methods
         v. Results
         vi. Discussion
         vii. Other information
      b. The flow diagram: Intended to depict the passage of study participants through the randomized controlled trial
   6. Extensions of the CONSORT statement
      a. Design extensions
         i. Cluster trials
         ii. Noninferiority and equivalence trials
         iii. Pragmatic trials
      b. Intervention extension
         i. Herbal medicinal interventions
         ii. Nonpharmacologic interventions
         iii. Acupuncture interventions
      c. Data extensions
         i. Patient-reported outcomes
         ii. Harms
         iii. Abstracts

B. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement
   1. Initially published in 2007
   2. “An international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies”
   3. Available at www.strobe-statement.org
   4. Endorsed by several publications and published in these journals
   5. The STROBE checklist: 22-item checklist, same basic concepts as the CONSORT checklist, with alterations germane to observational trials

C. Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA)
   1. Established in 1996 (as QUOROM), renamed in 2009, updated in 2015
   2. Evidence-based minimum set of items for reporting systematic reviews and meta-analyses
   3. Available at www.prisma-statement.org/
      a. The PRISMA checklist: 27-item checklist with alterations germane to systematic reviews and meta-analyses
b. The PRISMA flow diagram: Four-stage diagram, depicting the flow of information through the systematic review

c. The PRISMA explanation and elaboration document: Intended to enhance the use and understanding of the PRISMA statement

D. Enhancing the Quality and Transparency of Health Research (EQUATOR) Network
   1. International initiative to improve the reliability and value of medical research literature by promoting transparent and accurate reporting of research studies
   2. Does not have its own statements but promotes the use of key reporting guidelines
   3. Many other statements regarding study types not addressed in the discussion related to CONSORT, STROBE, and PRISMA are listed on the EQUATOR network website (www.equator-network.org).

XIII. PHARMACOECONOMIC STUDIES

A. Cost-Minimization Analysis
   1. Differences in cost among comparable therapies are evaluated
   2. Only useful to compare therapies that have similar outcomes

B. Cost-Effectiveness Analysis
   1. Outcome: Clinical units or cost per unit health outcome (outcome examples: years of life saved, number of symptom-free days, blood glucose, blood pressure, etc.)
   2. Useful to measure the cost impact when health outcomes are improved

C. Cost-Utility Analysis
   1. Assigns utility weights to outcomes so the impact can be measured in relation to cost (outcome example: quality-adjusted life-years)
   2. Compares outcomes related to mortality when mortality may not be the most important outcome

D. Cost-Benefit Analysis
   1. Monetary value is placed on both therapy costs and beneficial health outcomes.
   2. Allows analysis of both the cost of treatment and the costs saved with beneficial outcomes

XIV. SENSITIVITY/SPECIFICITY/PREDICTIVE VALUES

A. Sensitivity: Proportion of True Positives That Are Correctly Identified by a Test; a test with a high sensitivity means that a negative test can rule OUT the disorder.

B. Specificity: Proportion of True Negatives That Are Correctly Identified by a Test; a test with high specificity means that a positive test can rule IN the disorder.

C. Positive Predictive Value: Proportion of Patients with a Positive Test Result Who Actually HAVE the Disease
D. Negative Predictive Value: Proportion of Patients with a Negative Test Result Who Actually DO NOT HAVE the Disease

E. Example: Tables 8 and 9

Table 8. Relationship Between Test and Correct Diagnosis Identified by Disease

<table>
<thead>
<tr>
<th>Test</th>
<th>Disease Present</th>
<th>Disease Absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test positive</td>
<td>True positive</td>
<td>False positive</td>
<td>TP + FP</td>
</tr>
<tr>
<td>Test negative</td>
<td>False negative</td>
<td>True negative</td>
<td>TN + FN</td>
</tr>
<tr>
<td>Total</td>
<td>TP + FN</td>
<td>FP + TN</td>
<td>Total</td>
</tr>
</tbody>
</table>

1. Sensitivity = TP/(TP + FN)
2. Specificity = TN/(TN + FP)
3. Positive predictive value = TP/(TP + FP)
4. Negative predictive value = TN/(TN + FN)
5. Positive likelihood ratio = sensitivity/(1 − specificity)
6. Negative likelihood ratio = (1 − sensitivity)/specificity

Table 9. Relationship Between Test and Correct Diagnosis Identified by Disease in a Published Study

<table>
<thead>
<tr>
<th>Test</th>
<th>Positive disease</th>
<th>Negative disease</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>231 (true positive)</td>
<td>32 (false positive)</td>
<td>263</td>
</tr>
<tr>
<td>Negative</td>
<td>27 (false negative)</td>
<td>54 (true negative)</td>
<td>81</td>
</tr>
<tr>
<td>Total</td>
<td>258</td>
<td>86</td>
<td>344</td>
</tr>
</tbody>
</table>

REFERENCES

ANSWERS AND EXPLANATIONS TO SELF-ASSESSMENT QUESTIONS

1. Answer: B
Recall bias is always a potential concern for case-control studies because of the amount of time that passes between the study and the drug “ingestion.” Because risk factors were not included in the study design, this is of concern (Answer B is correct). Although a study may be susceptible to many types of bias, the other choices would not pose as much risk (if any) compared with recall bias (Answers A, C, and D are incorrect).

2. Answer: B
Internal validity is greatly jeopardized because the study is not designed to protect against this possible bias. In a sense, this design flaw jeopardizes external validity (how well does a study apply to other patients with this condition/disease?), but a lack of internal validity is most affected (Answer B is correct). The other answers can be adequately controlled for in the design and conduct of the study (Answers A, C, and D are incorrect).

3. Answer: B
Intention-to-treat analysis generally considers the approach, which gives the best estimate of use effectiveness (use under typical clinical trial conditions), whereas per-protocol analysis gives a better estimate of method effectiveness (use under ideal conditions) (Answer B is correct; Answer C is incorrect). Intention-to-treat analysis is the most common approach to data analysis for randomized controlled trials and may underestimate the treatment effect (Answer D is incorrect). Recall bias is not a concern with randomized controlled trials (Answer A is incorrect).

4. Answer: B
The CI of the difference in recurrence rate between the two groups includes zero; thus, there is no statistically significant difference between the two groups (Answer B is correct). Answer A is incorrect because the 95% CI contains zero and is therefore not statistically significant. Answer C is incorrect because not enough information is provided. Answer D is incorrect because all the previously stated information can be determined without the benefit of reported p-values.

5. Answer: D
Answers A and C are incorrect calculations. Calculating the NNT to prevent one recurrence using twice-daily therapy is as follows: 0.044 – 0.029 = 0.015 and 1/0.015 = 66.7…67; however, the NNT should not be calculated when the end point of interest is nonsignificant (Answer B is incorrect; Answer D is correct).

6. Answer: C
Answers A and B are incorrect because the margarine with ALA did not significantly reduce the risk of cardiovascular events (the 95% CI includes 1 [no difference in risk]). Answer D is incorrect because the p-value is not required for interpreting statistical significance when the 95% CI is provided. Answer C is correct because the p-value corresponds to the 95% CI.

7. Answer: A
Clinical trials are usually adequately powered to compare primary end points (Answer A is correct). Because Answer B is part of the composite outcomes, the study was likely not powered to detect this outcome independently. Similarly, even though the subgroup analysis was determined a priori, the study is not typically designed to have sufficient power to make this comparison (Answers C and D are incorrect).

8. Answer: D
Answers A and B are incorrect because each implies that one drug is more effective than the other. In this type of study design, neither drug is more/less effective. Answer C is incorrect because the CI of the OR does not include 1; thus, the finding is statistically significant at the 5% level, making Answer D correct.