BIOSTATISTICS AND TRIAL DESIGN:
AN INTERACTIVE REVIEW FOR PHARMACISTS
FRIDAY/8:45-10:45AM

ACPE UAN: 0107-9999-18-261-L01-P 0.2 CEU/2.0 hr

Activity Type: Application-Based

Learning Objectives for Pharmacists:
Upon completion of this CPE activity participants should be able to:
1. Describe basic terms in biostatistics related to types of data studied, measures of central tendency, and descriptors of population distribution variability.
2. Compare and contrast different statistical tests to determine significance and apply them to a given clinical trial.
3. Discuss study designs and the level of evidence each design can achieve.
4. Appraise the results of a biomedical study and its applicability to practice including bias and confounding, internal and external validity, and numbers needed to treat/harm.

Speakers:

Darla Eastman, PharmD, BCPS
Darla Eastman is an associate professor at Drake University College of Pharmacy and Health Sciences and a clinical pharmacist at Iowa Methodist Medical Center where she is part of the Trauma Surgical Team. She is an active member of the Iowa Pharmacy Association, American College of Clinical Pharmacy, and the American Association of Colleges of Pharmacy. Dr. Eastman’s areas of interest include infectious disease and critical care.

Geoff Wall, PharmD, FCCP, BCPS, CGP
Geoffrey C. Wall is a Professor at Drake University College of Pharmacy and Director of the Drake Drug Information Center. His clinical practices include the Internal Medicine and Medical Intensive Care Teaching Services at Iowa Methodist Medical Center in Des Moines, IA. He received his Bachelor of Science in Pharmacy from the University of Utah and Doctor of Pharmacy from Idaho State University. He completed an ASHP-accredited Internal Medicine Specialty Residency at Scott and White Memorial Hospitals and Clinics. Dr. Wall has written several peer-reviewed papers and textbook chapters on a variety of topics and has designed or participated in several clinical trials.

Speaker Disclosure: Darla Eastman reports no actual or potential conflicts of interest in relation to this CPE activity. Geoffrey Wall reports that he is on the speaker’s bureau with Boehringer Ingelheim, Janssen and La Jolla Pharmaceuticals. Off-label use of medications will not be discussed during this presentation.
“Lies, Dang* Lies, and Statistics”: Biostatistics and Trial Design: An Interactive Primer for Pharmacists

* = apologies to Twain

Presented by: Geoffrey C. Wall, Pharm.D., FCCP, BCPS
Professor of Pharmacy Practice, Drake University
Internal Medicine Clinical Pharmacist
Iowa Methodist Medical Center

Darla Eastman, Pharm.D., BCPS
Associate Professor of Pharmacy Practice, Drake University
Trauma Clinical Pharmacist
Iowa Methodist Medical Center

Disclosure

• Geoff Wall reports no actual or potential conflicts of interest associated with this presentation
• Darla Eastman reports no actual or potential conflicts of interest associated with this presentation
Learning Objectives

• Upon successful completion of this activity, participants should be able to:
  • Describe basic terms in biostatistics related to types of data studied, measures of central tendency, and descriptors of population distribution variability
  • Compare and contrast different statistical tests to determine significance and apply them to a given clinical trial
  • Discuss study designs and the level of evidence each design can achieve
  • Appraise the results of a biomedical study and its applicability to practice including bias and confounding, internal and external validity, and numbers needed to treat/harm

Case #1

• As a hospital pharmacist you are asked to review a drug for your hospital's P&T Committee
• This drug, just approved by the FDA, has two major Phase 3 studies published to date
• In the “statistics” section of the paper the authors used numerous tests you have never heard of
• Do you assume that “since the journal approved the paper for publication, the stats must be correct?”
Lies…er, Misconceptions about Biostats

- Only statisticians can understand them
  - NOT true, understanding basic biomedical stats and trial design is not difficult
- If a paper is published in a peer-reviewed journal the stats are ok
  - In many cases statistics are only given a cursory glance in peer-review
- Every paper needs a statistician among the authors or the stats are suspect
  - It certainly helps to, but again, BASIC stats are not difficult to do

- Like everything else in Pharmacy, getting good at reviewing studies takes practice more than anything else

The purpose of this session

- Review basic concepts of biostats:
  - Types of data, measures of central tendency, populations, hypothesis testing
- Review different types of evidence/studies
  - RCT vs Cohort vs Case control
- Review types of statistical tests
- Pearls for reviewing studies

- This is just a BASIC review. Far more detailed study activities do exist for those interested
Case #2

- As an independent pharmacy owner, you have decided to add in-home delivery of medications to the services provided. To get an idea of the scope and success of this service, you ask patients who fill prescriptions to take a quick 4-question survey. Each question uses a Likert (1-5) scale to assess the feelings your patients have about this service.
- What type of data is this?

Types of data

- Dictates
  - Measures of central tendency
  - Appropriate statistical test to use
Types of Data

- **Nominal Data**: Dichotomous data,
  - Discrete data
  - e.g.: yes or no, alive or dead, pet owner or not
- **Ordinal Data**: Ranked data with no level of concrete measurement between levels
  - Discrete data
  - e.g. NYHA, Most Likert scales, APACHE II, Ramsay scale of sedation, etc.
- **Interval Data**: Data in specific order with defined levels of magnitude, but no set point for zero
  - Continuous data
  - e.g. Degrees Fahrenheit—this is rarely used in the biomedical literature
- **Ratio Data**: Interval data with a fixed zero
  - Continuous data
  - e.g. blood pressure, time, pulse, etc.

Case #2

- As a independent pharmacy owner you have decided to add in-home delivery of medications to the services provided. To get an idea of the scope and success of this service you ask the patients who fill prescription to take a quick 4 question survey. Each question uses a Likert (1-5) scale to assess the feelings your patients have about this service
- What type of data is this?
Case #3

• As an independent pharmacist you and your tech are reviewing time to receiving payments from PBMS (stop laughing…). Of the 15 PBMs you review 13/15 will reimburse sometime within day 1 and day 28 of the claim (I said STOP LAUGHING) and 2/15 take between 28 and 60 days to reimburse.

• What is the best measure of central tendency to describe this data?

Mean (μ)

• The arithmetic average (add all of the scores together, then divide by the number of scores)

\[ \mu = \frac{\sum x}{n} \]
Median

- The middle number (just like the median strip that divides a highway down the middle; 50/50)
- Used when data is not normally distributed
- Often hear about the median price of housing

Mode

- The most frequently occurring number (score, measurement, value, cost)
- On a frequency distribution, it’s the highest point (like the à la mode on pie)
Case #3

- As an independent pharmacist you and your tech are reviewing time to receiving payments from PBMS (stop laughing…). Of the 15 PBMs you review 13/15 will reimburse sometime within day 1 and day 28 of the claim (I said STOP LAUGHING) and 2/15 take between 28 and 60 days to reimburse

- What is the best measure of central tendency to describe this data?
Confidence Intervals (CI)

**Confidence Interval**: An interval of values computed from the sample, that is almost sure to cover the true population value.

We make confidence intervals using values computed from the sample, not the known values from the population.

**Interpretation**: In 95% of the samples we take, the true population proportion (or mean) will be in the interval.

This is also the same as saying we are 95% confident that the true population proportion (or mean) will be in the interval.

SO if 1 (or “parity”) is in the CI of a result we are saying that “no difference” is as statistically likely to the real answer as the one found. This means NO statistical difference in the result (no p-value needed!).

**So what?**

- In most clinical trials it is these measures of central tendency that are compared against each other in clinical trials
  - **Mean** blood pressure decrease between two drugs
  - **Median** time to remission of disease in a chemotherapy regimen
  - **Modes** actually rarely used in clinical trials
Case #4

- A industry representative speaks to the director of a hospital pharmacy about a new drug for acute lowering of blood pressure after a head injury. The paper she brings to review compares 40 patients with a head injury who received either standard of care (SOC) or her company’s new drug. Mean SBP decreased 19 mmHg with the SOC and 13 mmHg with the new drug (p > 0.05). She claims that “both drugs are equal in lowering BP”
- What could be a possible flaw in this study?
We make mistakes!

**Alpha level**
- Set BEFORE we collect data, run statistics
- Defines how much of an error we are willing to make to say we made a difference
- If we’re wrong, it’s an alpha error or Type 1 error

**p value**
- Calculated AFTER we gather the data
- The calculated probability of a mistake by saying it works
- AKA: level of significance
- Describes the percent of the population/area under the curve (in the tail) that is beyond our statistic

---

**β Error and Power**

- **Type II Error (β) = False Negative**
  - In other words not finding a difference between groups when one, in fact, exists,

- **Power** is the probability of making a correct decision when the Null Hypothesis (Ho) is false
  - In other words, can you detect a difference if one exists

- Power is increased by: increasing n, increasing alpha, increasing differences between populations
Case #4

• A industry representative speaks to the director of a hospital pharmacy about a new drug for acute lowering of blood pressure after a head injury. The paper she brings to review compares 40 patients with a head injury who received either standard of care (SOC) or her company’s new drug. Mean SBP decreased 19 mmHg with the SOC and 13 mmHg with the new drug (p > 0.05). She claims that “both drugs are equal in lowering BP”

• What could be a possible flaw in this study?

2-Tailed Test

• The critical value is the number that separates the “blue zone” from the middle (± 1.96 this example)
• In a t-test, in order to be statistically significant the t score needs to be in the “blue zone”
• If α = .05, then 2.5% of the area is in each tail
1-Tailed Test

• The critical value is either + or -, but not both.
• In this case, you would have statistical significance ($p < .05$) if $t \geq 1.645$.

Case #5

• You are writing a P&T monograph for your hospital's formulary committee concerning one of the new PCSK9 drugs for hypercholesterolemia. One of the papers you read is a “small crossover study where participants are given one dose of the drug or placebo, had their lipid profiles measured and then 1 week later given the opposite treatment (the active drug group now gets placebo and vice-versa)

• What issues would you have concerning the accuracy of this study?
General Study Designs

• Many clinical trial study designs fall into the categories of parallel group, dose-ranging, cross-over and factorial designs

• There are many other possible designs and variations on these designs
General Study Designs

• Dose-Ranging Studies

- high dose
- medium dose
- low dose
- control

General Study Designs

• Cross-Over Designs

- wash-out
- A → B
- B → A
General Study Designs

- Factorial Designs

- Cross-Over Designs
  - Subjects are randomized to sequences of treatments (A then B or B then A)
  - Uses the patient as his/her own control
  - Often a “wash-out” period (time between treatment periods) is used to avoid a “carry over” effect (the effect of treatment in the first period affecting outcomes in the second period)
  - Can have a cross-over design with more than 2 periods
Cross-Over Designs

- Advantage: treatment comparison is only subject to within-subject variability not between-subject variability
  ➞ reduced sample sizes, so easier and cheaper to do
- Disadvantages:
  - strict assumption about carry-over effects
  - inappropriate for certain acute diseases (where a condition may be cured during the first period)
  - drop outs before second period

Cross-Over Designs

- Appropriate for conditions that are expected to return to baseline levels at the beginning of the second period

Examples:
  - Treatment of chronic pain
  - Comparison of hearing aids for hearing loss
  - Mouth wash treatment for gingivitis
Case #5

- You are writing a P&T monograph for your hospital’s formulary committee concerning one of the new PCSK9 drugs for hypercholesterolemia. One of the papers you read is a “small crossover study where participants are given one dose of the drug or placebo, had their lipid profiles measured and then 1 week later given the opposite treatment (the active drug group now gets placebo and vice-versa)

- What issues would you have concerning the accuracy of this study?

Factorial Designs

- Attempts to evaluate two interventions compared to a control in a single experiment (simplest case)
- An important concept for these designs is interaction (sometimes called effect modification)

Interaction: The effect of treatment A differs depending upon the presence or absence of intervention B and vice-versa.
**Factorial Designs**

- **Advantages:**
  - If no interaction, can perform two experiments with less patients than performing two separate experiments
  - Can examine interactions if this is of interest

- **Disadvantages:**
  - Added complexity
  - Potential for adverse effects due to "poly-pharmacy"

---

**Example:** Physician’s Health Study

Physicians randomized to:
- aspirin (to prevent cardiovascular disease)
- beta-carotene (to prevent cancer)
- aspirin and beta-carotene
- neither (placebo)

---

Cohort Design

• A cohort consists of a group of individuals from a well-defined population (exposure and characteristics known) followed over time to observe what happens to which groups and when

Framingham Cohort Study

• Followed up 5,573 people from 1968 onwards
  • White
  • Initially free of cardiovascular disease
  • Included people with hypertension and diabetes
• Developed predictive algorithms for CHD we use to this day
Retrospective Cohort Design

- A group of subjects or cohort is identified in the past
- Follow-up is then to the present
- Most MUEs are this design
- Advantages
  - Data already collected and events have occurred
  - Quick and less expensive design
  - Cannot remove bias entirely.
  - Regression model or propensity matching may help account for confounders

Case-Control Design

- Group of subjects with disease or condition are identified (Cases)
- Suitable Control group identified without the condition
- Frequency of exposure or risk factor compared in cases and controls
- Decreases confounding factors that might explain differences between groups
Case-Control Design

CASES

Past Present
Unexposed Exposed
Exposed Unexposed

CONTROLS

Case #6

• You are working for a company that conducts clinical trials locally for large drug companies. You are reviewing a potential study your business may participate in where new DM2 patients are randomized in a double-blind fashion to metformin 1000mg bid or a new DPP-IV inhibitor. Outcomes will include Hgb-A1c, quality of life and adverse effects.

• What issues do you think this study may have? Should you agree to help conduct the trial?
Why is blinding important?

• Why blind patients?
  • Patients expectations can influence response
  • Might report more adverse events if known to be on treatment
  • Might assume no efficacy if on placebo
• Why blind investigators?
  • May subconsciously influence outcome measures
  • Some endpoints controlled by investigators and could be influenced by knowledge of treatment

General Notes on Trial Design

• Trial design helps determine causality (RCT) or associations (Everything else)
• Look at internal vs external validity (esp. inclusions/exclusions)—in other words are the patients in the study similar to the patients you see?
• Is it biologically plausible or supported by previous work?
• Level of evidence?
Types of Statistical Tests

- Ok, there are a ton of these out there BUT
  - For clinical studies a few tried and true tests work nearly all the time
  - If a paper does not use them, think about why??
  - Being able to appropriately analyze a paper really means knowing if the data and circumstances match the test used, not if they did the math right!!
### Dropouts and Non-Compliance

- **Intention to Treat Principle**
  - Results analyzed as randomized—even if they only were in the study briefly—commonly done in studies with drugs
  - Evaluates the effect of a treatment “policy”
  - More “real world” but power decreases

- **Per protocol**
  - Exclude dropouts
  - Adjust for compliance or dose received
  - Evaluates the effect of the “active ingredient”
  - But makes it harder to assess “real world” effects of taking medication

---

<table>
<thead>
<tr>
<th>TYPE OF VARIABLE</th>
<th>2 SAMPLES (INDEPENDENT) (PARALLEL DESIGN)</th>
<th>2 SAMPLES (RELATED) (CROSS-OVER OR PRE-POST DESIGN)</th>
<th>≥3 SAMPLES (INDEPENDENT) (PARALLEL DESIGN)</th>
<th>≥3 SAMPLE (RELATED) (CROSS-OVER DESIGN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>X² or Fisher’s exact</td>
<td>McNemar’s test</td>
<td>X² (Bonferroni)</td>
<td>COCHRAN’S Q (Bonferroni)</td>
</tr>
<tr>
<td></td>
<td>1 confounder</td>
<td>Mantel-Haenszel</td>
<td>rare</td>
<td>rare</td>
</tr>
<tr>
<td></td>
<td>≥2 confounders</td>
<td>Logistic regression</td>
<td>Logistic regression</td>
<td>rare</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Wilcoxon rank sum or Mann Whitney U</td>
<td>Wilcoxon signed rank test</td>
<td>Kruskal-Wallis ANOVA (MCP or Bonferroni)</td>
<td>Friedman ANOVA</td>
</tr>
<tr>
<td></td>
<td>1 confounder</td>
<td>2-way repeated ANOVA ranks</td>
<td>2-way ANOVA ranks</td>
<td>2-way repeated ANOVA ranks</td>
</tr>
<tr>
<td></td>
<td>≥2 confounders</td>
<td>Repeated measures regression</td>
<td>ANCOVA ranks</td>
<td>Repeated measures regression</td>
</tr>
<tr>
<td>Continuous</td>
<td>Student’s t-test</td>
<td>Paired Student’s t-test</td>
<td>1-way ANOVA (MCP)</td>
<td>Repeated measures ANOVA (MCP)</td>
</tr>
<tr>
<td></td>
<td>1 confounder</td>
<td>2-way ANOVA</td>
<td>2-way ANOVA</td>
<td>2-way repeated ANOVA</td>
</tr>
<tr>
<td></td>
<td>≥2 confounders</td>
<td>ANCOVA</td>
<td>ANCOVA</td>
<td>Repeated measures regression</td>
</tr>
</tbody>
</table>
Superiority vs. Equivalence
Trial Design

• **Superiority:**
  • Hypothesis: Treatment A is better than Treatment B
  • Statistical testing: Prove that Tx A is not equal to Tx B (disprove Null hypothesis)

Non-Inferiority (Clinical Equivalence):
• Hypothesis: Tx A is at least as good as Tx B
• Statistical testing: Prove that Tx A is not worse than Tx B

• **Equivalence:**
  • Statistical testing: Prove that Tx B is not worse (and not better) than Tx A

Why do a non-inferiority study?

• Placebo control trial unethical but still want to demonstrate that the new treatment is at least as effective as the current standard of therapy
• New therapy may offer important advantages over currently available effective therapies
  • Improved safety
  • Better tolerability/fewer side effects
  • Ease of use
  • Less expensive (ha!)
  • Increased market competition
• Most common reason: easier and cheaper to do
Design of Noninferiority Trials

- Region of non-inferiority must be defined in advance

- If the upper bound 95% Confidence Interval of the difference between two treatments lies entirely below the pre-specified boundary then these treatments may be considered clinically equivalent
Outcome Assessment

- Absolute vs relative risk
- Is the outcome clinically as well as statistically significant?
- Do other factors outweigh this outcome?
  - Safety?
  - Cost?
  - Ease of use?
Example

- New drug for acute myocardial infarction to reduce mortality
- First studied in a high risk population:
  - 40% mortality at 30 days among untreated
  - e.g., elderly, heart failure, anterior wall infarction

Ref: [http://www.cche.net/usersguides/ebm_tips.asp](http://www.cche.net/usersguides/ebm_tips.asp)

Example

- New drug for acute myocardial infarction to reduce mortality
- First studied in a high risk population:
  - 40% mortality at 30 days among untreated
  - e.g., elderly, heart failure, anterior wall infarction
  - 30% mortality among treated
- How would you describe the effect of the new drug?

Ref: [http://www.cche.net/usersguides/ebm_tips.asp](http://www.cche.net/usersguides/ebm_tips.asp)
Example

Absolute risk reduction (also called the risk difference) is the simple difference in the event rates (40%−30%=10%).

Relative risk reduction is the difference between the event rates in relative terms. Here, the event rate in the treatment group is 25% less than the event rate in the control group (i.e., the 10% absolute difference expressed as a proportion of the control rate reduction is 10/40 or 25% less.)

Ref: http://www.cche.net/usersguides/ebm_tips.asp
This is where NNT and NNH come in

- **Number Needed to Treat (NNT):**
  - Number of persons who would have to receive an intervention for 1 to benefit.

- **Number Needed to Harm (NNH):**
  - Number of persons who would have to receive an intervention for 1 to experience an adverse event.

**NNT and NNH Calculation**

\[
\text{NNT} = \frac{100}{\text{ARR}} \quad \text{or} \quad \frac{1}{\text{ARR}}
\]

\[
\text{NNH} = \frac{100}{\text{ARI}} \quad \text{or} \quad \frac{1}{\text{ARI}}
\]

Note:
- Absolute Risk Reduction (ARR)
- Absolute Risk Increase (ARI)
A new pain medication with both serotonergic and opioid properties is approved by the FDA. In reading the Phase 3 study that merited this approval you find that in a study of patients with chronic low back pain, 128 of 560 patients achieved the primary endpoint of a 50% improvement in pain compared to only 85 of 545 patients in a placebo arm. The drug costs an AWP of $500/month. How can the above be discussed with the average patient in language they can understand?

Case #7

- **NNT = 1/ARR**
  - **ARR = 128/560 (active) – 85/545 (placebo)**
  - **ARR = 22.8% - 15.6% = 7.2% or 0.072**
  - **NNT = 1/0.072 = 13.8 → 14 patients**

- At $500/month/patient for 14 patients = $84,000/year for one patient to achieve a 50% improvement in pain
What is a meta analysis?

Quantitative approach for systematically combining results of previous research to arrive at conclusions about the body of research.

What does it mean?

• Quantitative: numbers
• Systematic: methodical
• Combining: putting together
• Previous research: what's already done
• Conclusions: new knowledge
Why do a meta-analysis?

- Small studies don’t give a real idea of effect of treatment
- Conflicting results of smaller studies (tie-breaker)
- Larger population to look for dropouts or harm
- Are meta-analyses equivalent to RCTs (i.e prove causality)??
  - even now, a contentious subject
- GIGO!! (garbage in, garbage out)

What to look for in reading a meta-analysis

- Three Outcome Measures
  - Effect Size (Odds Ratio)—how large was the clinical effect
  - Variance with 95% Confidence Interval—does it favor one treatment over another?
  - Test of heterogeneity
- Two Graphs
  - Forest Plot
  - Funnel Plot
Some indication that exercise is about as effective as medication.
Too few studies to be conclusive.

Some indication that exercise is may add to effects beyond medication.
Too few studies to be conclusive.
Forrest Plot Example

**Review:** Mefloquine for preventing malaria in non-immune adult travellers
**Comparison:** Mefloquine versus alternative chemoprophylaxis
**Outcome:** Diarrhoea

<table>
<thead>
<tr>
<th>Study or sub-category</th>
<th>Treatment</th>
<th>Control</th>
<th>Peto OR 95% CI</th>
<th>Weight %</th>
<th>Peto OR 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthur 1991</td>
<td>64/119</td>
<td>58/119</td>
<td>24.65</td>
<td>0.96</td>
<td>10.55, 1.571</td>
</tr>
<tr>
<td>Fournier 1983</td>
<td>16/200</td>
<td>19/156</td>
<td>12.14</td>
<td>0.61</td>
<td>10.30, 1.241</td>
</tr>
<tr>
<td>Croft 1997a</td>
<td>29/203</td>
<td>103/176</td>
<td>33.94</td>
<td>0.16</td>
<td>10.20, 0.20</td>
</tr>
<tr>
<td>Croft 1997b</td>
<td>25/247</td>
<td>28/245</td>
<td>27.53</td>
<td>0.99</td>
<td>10.55, 1.771</td>
</tr>
<tr>
<td>Kolakowski 1987</td>
<td>22/69</td>
<td>9/69</td>
<td>9.04</td>
<td>0.07</td>
<td>1.96, 6.931</td>
</tr>
<tr>
<td>Chil 1997</td>
<td>7/68</td>
<td>4/67</td>
<td>3.87</td>
<td>1.77</td>
<td>0.52, 6.061</td>
</tr>
</tbody>
</table>

Total (95% CI)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>895</td>
<td>822</td>
</tr>
</tbody>
</table>

Test for heterogeneity: Q = 61.09, df = 5 (P = 0.00001), I² = 91.8%
Test for overall effect: Z = 4.36 (P = 0.0001)

**Funnel Plots**

Shows effect size by precision. Expect a funnel shape if fixed-effects sampling distribution. Asymmetry suggests pub bias; excess variance suggests heterogeneity (moderators). This one shows pretty good symmetry but lots of variability (mortality rates by hospital).
How to deal with heterogeneity

Some researchers ignore heterogeneity: using *fixed effect model*

Some allow for heterogeneity: using *random effects model*

Either might be reasonable depending on variables and is beyond the scope of this talk

Statistical measures of heterogeneity

- The Chi$^2$ test measures the amount of variation in a set of trials, and tells us if it is more than would be expected by chance
- Small p values suggest that heterogeneity is present
- This test is not very good at detecting heterogeneity. Often a cut-off of $p<0.10$ is used, but lack of statistical significance does not mean there is no heterogeneity
Statistical measures of heterogeneity (2)

- The $I^2$ is available and is probably the measure of choice to determine heterogeneity.
- $I^2$ is the proportion of variation that is due to heterogeneity rather than chance.
- Large values of $I^2$ suggest heterogeneity.
- Roughly, $I^2$ values of 25%, 50%, and 75% could be interpreted as indicating low, moderate, and high heterogeneity.

Accuracy of Diagnostic Studies
Case #8

- As a hospital pharmacist you are tasked with initiating your facility’s antimicrobial stewardship program. One facet of this program you feel would be helpful is targeted PCN allergy skin testing. However your physicians are concerned the test does not identify PCN allergic patients reliably. The guidelines on PCN skin testing note it has a 97% negative predictive value (NPV).
- What should you report to your physicians?

Measures of Test Accuracy

<table>
<thead>
<tr>
<th>Disease</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td></td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

- Sensitivity and Specificity
- Predictive values
- Likelihood ratios
- Diagnostic odds ratio
Accuracy of Diagnostic Studies Summary

• **Sensitivity:** Who really has the disease and is identified as such
  • So sensitive tests RULE OUT a condition

• **Specificity:** Who really does not have the disease and is correctly identified as such
  • So specific tests RULE IN a condition

• **Positive Predictive Value** (PPV) is the proportion of the people who test positive who truly have the disease

• **Negative Predictive Value** is the proportion of the people who test Negative who truly do not have the disease

Case #8

• As a hospital pharmacist you are tasked with initiating your facility’s antimicrobial stewardship program. One facet of this program you feel would be helpful is targeted PCN allergy skin testing. However your physicians are concerned the test does not identify PCN allergic patients reliably. The guidelines on PCN skin testing note it has a 97% negative predictive value (NPV).

• What should you report to your physicians?
Take Home Points

• Stats and study assessment can seem overwhelming
• But like nearly every skill in pharmacy—practice makes perfect—well, better anyway
• Peer-reviewed publication does not guarantee a perfect paper
• Practice at your site—journal club is a great tool

Further References for Stats and Lit Review

• http://www.dartmouth.edu/~library/biomed/guides/research/ebm-az-list.html
• http://guides.library.harvard.edu/hms/ebm
• https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3121570/
Application Cases

Trial Design Case

- The hospital pharmacy has asked you to evaluate use of Kcentra (4-factor PCC) in reversing warfarin.
- Why type of trial design would you use?
Trial Design Case

• Most common would be to use a retrospective cohort design.
• Could evaluate all patient’s who received Kcentra in a certain historical time period and examine pre- and post-INR values or other outcome.
• Why might a different trial design be chosen?
  • If hospital was evaluating different dosing.
  • Evaluating Kcentra compared with vitamin K, FFP, or both

Romosozumab Trial


Romosozumab or Alendronate for Fracture Prevention in Women with Osteoporosis.


Author information

Abstract

BACKGROUND: Romosozumab is a monoclonal antibody that binds to and inhibits sclerostin, increases bone formation, and decreases bone resorption.

METHODS: We enrolled 4093 postmenopausal women with osteoporosis and a fragility fracture and randomly assigned them in a 1:1 ratio to receive monthly subcutaneous romosozumab (210 mg) or weekly oral alendronate (70 mg) in a blinded fashion for 12 months, followed by open-label alendronate in both groups. The primary end points were the cumulative incidence of new vertebral fracture at 24 months and the cumulative incidence of clinical fracture (nonvertebral and symptomatic vertebral fracture) at the time of the primary analysis (after clinical fractures had been confirmed in ≥330 patients). Secondary end points included the incidences of nonvertebral and hip fracture at the time of the primary analysis. Serious cardiovascular adverse events, osteonecrosis of the jaw, and atypical femoral fractures were adjudicated.
Results Evaluation

<table>
<thead>
<tr>
<th>Romosozumab → Alendronate</th>
<th>Alendronate → Alendronate</th>
<th>Statistical Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Vertebral Fractures</td>
<td>6.2%</td>
<td>11.9%</td>
</tr>
<tr>
<td>Clinical Fractures</td>
<td>9.7%</td>
<td>13.0%</td>
</tr>
<tr>
<td>Serious Cardiac Event</td>
<td>2.5%</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

• What is the Relative Risk Reduction (RRR) and ARR of new vertebral fractures for the romosozumab group?
  • ARR = 11.9%-6.2% = 5.7%
  • RRR = 5.7%/11.9% = 47.9%
• What is the NNT to prevent one new vertebral fracture?
  • NNT = 100/5.7% = 17.5 = 18 people
• How can we evaluate the cardiac event risk?


**Impact of pharmacological treatment of diabetes mellitus on dementia risk: systematic review and meta-analysis.**

McMillan JM1,2, Mele BS2, Hogan DB1, Leung AA1,2.

**BACKGROUND:** The association between diabetes mellitus (DM) treatment and dementia is not well understood.

**OBJECTIVE:** To investigate the association between treatment of diabetes, hypoglycemia, and dementia risk.

**RESEARCH DESIGN AND METHODS:** We performed a systematic review and meta-analysis of pharmacological treatment of diabetes and incident or progressive cognitive impairment. We searched Ovid MEDLINE, Embase, Cochrane Central Registry of Controlled Trials, and PsychINFO from inception to 18 October 2017. We included cross-sectional, case-control, cohort, and randomized controlled studies. The study was registered with PROSPERO (ID CRD42017077953).

**RESULTS:** We included 37 studies into our systematic review and 13 into our meta-analysis. Ten studies investigated any antidiabetic treatment compared with no treatment or as add-on therapy to prior care. Treatment with an antidiabetic agent, in general, was not associated with incident dementia (risk ratio (RR) 1.01; 95% CI 0.93 to 1.10). However, we found differential effects across drug classes, with a signal of harm associated with insulin therapy (RR 1.21; 95% CI 1.06 to 1.39), but potentially protective effects with thiazolidinedione exposure (RR 0.71; 95% CI 0.55 to 0.93). Severe hypoglycemic episodes were associated with a nearly twofold increased likelihood of incident dementia (RR 1.77; 95% CI 1.35 to 2.33). Most studies did not account for DM duration or severity.

**CONCLUSIONS AND LIMITATIONS:** The association between treatment for diabetes and dementia is differential according to drug class, which is potentially mediated by hypoglycemic risk. Not accounting for DM duration and/or severity is a major limitation in the available evidence base.
Questions to Ask as you review this Meta-Analysis

• What kind of studies did they review?
• Were they the same type of study?
• Where they the same type of patients?
• How did they account for bias?
• How did they account for different outcomes and how they were measured?
• What was the heterogeneity of the studies reviewed? How can you assess?
• So what can you take from this study??
Relative risk of developing dementia based on the occurrence of one or more hypoglycemic events.

<table>
<thead>
<tr>
<th>Hypoglycemia Exposure (Author, Year)</th>
<th>RR (95% CI)</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypoglycemia - two or more events (Chin et al., 2016)</td>
<td>4.07 (1.04, 15.00)</td>
<td>3.99</td>
</tr>
<tr>
<td>Hypoglycemia - one event (Chin et al., 2016)</td>
<td>2.09 (1.00, 5.09)</td>
<td>7.14</td>
</tr>
<tr>
<td>Hypoglycemia - one event (Whitmer et al., 2003)</td>
<td>1.26 (1.01, 1.27)</td>
<td>30.46</td>
</tr>
<tr>
<td>Hypoglycemia - one event (Yaffe et al., 2013)</td>
<td>2.10 (1.00, 4.40)</td>
<td>9.75</td>
</tr>
<tr>
<td>Hypoglycemia - three or more events (Whitmer et al., 2003)</td>
<td>1.04 (1.02, 2.06)</td>
<td>23.51</td>
</tr>
<tr>
<td>Hypoglycemia - two events (Whitmer et al., 2003)</td>
<td>1.00 (1.07, 2.00)</td>
<td>25.26</td>
</tr>
<tr>
<td>Overall (I²-squared = 64.4%, p = 0.015)</td>
<td>1.77 (1.36, 2.30)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis.

Jacqueline M McMillan et al. BMJ Open Diab Res Care 2018;6:e000563
©2018 by American Diabetes Association

Rivaroxaban versus enoxaparin for the prevention of venous thromboembolism after total knee arthroplasty: A meta-analysis.

OBJECTIVE: This article analyzed the clinical efficacy and tolerability of rivaroxaban and enoxaparin in patients undergoing total knee arthroplasty (TKA) surgery.

METHODS: Five randomized, controlled clinical trials on rivaroxaban versus enoxaparin in patients who underwent TKA were identified and included in this meta-analysis.

RESULTS: The meta-analysis indicated that rivaroxaban prophylaxis was associated with lower rates of symptomatic venous thromboembolism (VTE) (relative risk [RR]: 0.55; 95% confidence interval [CI]: 0.35-0.86; P=.009), symptomatic deep vein thrombosis (DVT) (RR: 0.44; 95% CI: 0.25-0.80; P=.007), asymptomatic DVT (RR: 0.57; 95% CI: 0.37-0.89; P=.01), distal DVT (RR: 0.62; 95% CI: 0.45-0.85; P=.003) and proximal DVT (RR: 0.42; 95% CI: 0.24-0.75; P=.004). Compared with the enoxaparin group, the incidence of symptomatic pulmonary embolism (PE) (RR: 0.48; 95% CI: 0.19-1.24; P=.13) in the rivaroxaban group was not significantly different. A nonsignificant trend towards all-cause death (RR: 0.38; 95% CI: 0.03-4.92; P=.46) or major bleeding (RR: 1.59; 95% CI: 0.77-3.27; P=.21) risk between rivaroxaban and enoxaparin prophylaxis was found.

CONCLUSION: Compared with the enoxaparin group, the group using rivaroxaban after TKA had a significantly lower rate of symptomatic VTE, symptomatic DVT, asymptomatic DVT, distal DVT, and proximal DVT. Our study shows that rivaroxaban after TKA is more effective than enoxaparin and did not increase major bleeding or all-cause mortality.
Questions to Ask as you review this Meta-Analysis

- What kind of studies did they review?
- Were they the same type of study?
- Where they the same type of patients?
- How did they account for bias?
- How did they account for different outcomes and how they were measured?
- What was the heterogeneity of the studies reviewed? How can you assess?
- So what can you take from this study??

Characteristics of included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Study style</th>
<th>Group</th>
<th>Dose and average use time</th>
<th>No. of patients</th>
<th>Gender, M/F</th>
<th>Age, years</th>
<th>BMI, kg/m²</th>
<th>Average operation time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turg(19)</td>
<td>2005</td>
<td>RCT</td>
<td>Recombination</td>
<td>Oral 2.5, 5, 10, 20 or 30 mg bid daily, 5–9 days after surgery</td>
<td>104</td>
<td>47/57</td>
<td>66</td>
<td>31.8</td>
<td>99.5</td>
</tr>
<tr>
<td>Lassen(16)</td>
<td>2008</td>
<td>RCT</td>
<td>Recombination</td>
<td>Oral 30 mg once daily, at least day 10 and up to day 14 after surgery</td>
<td>1220</td>
<td>30/757</td>
<td>67.6</td>
<td>29.5</td>
<td>96.0</td>
</tr>
<tr>
<td>Turg(19)</td>
<td>2000</td>
<td>RCT</td>
<td>Recombination</td>
<td>Oral 10 mg once daily oral, day 11 to day 15 after surgery</td>
<td>1520</td>
<td>519/1007</td>
<td>64.4</td>
<td>30.0</td>
<td>100.4</td>
</tr>
<tr>
<td>Zou(11)</td>
<td>2014</td>
<td>RCT</td>
<td>Recombination</td>
<td>Oral recombinant at a dose of 0.4 mg/day, treated for 14 days after surgery</td>
<td>102</td>
<td>30/70</td>
<td>63.5</td>
<td>27.5</td>
<td>84.8</td>
</tr>
<tr>
<td>Ke(17)</td>
<td>2017</td>
<td>RCT</td>
<td>Recombination</td>
<td>A full dose of recombinant 0.4 mg of 4000 IU was subcutaneously administered once daily, 15 days after surgery</td>
<td>95</td>
<td>22/74</td>
<td>65.2</td>
<td>25.4</td>
<td>69.0</td>
</tr>
</tbody>
</table>

Note: BMI = body mass index, RCT = randomized controlled clinical trial.
 Outcome: Symptomatic VTE

Rivaroxaban versus enoxaparin: symptomatic venous thromboembolism after total knee arthroplasty.

Outcome: All-cause mortality

Rivaroxaban versus enoxaparin: all-cause mortality after total knee arthroplasty.
Biostatistical Pearls

• Data type (Nominal, Ordinal, Interval, Ratio)
  • Assess central tendency
  • Appropriate statistical tests
• Study design limitations and strengths
• Superiority v noninferiority
• Significance (NNT, NNH)
• Meta-analysis evaluation

“Lies, Dang* Lies, and Statistics”: Biostatistics and Trial Design: An Interactive Primer for Pharmacists

* = apologies to Twain

Presented by:
Geoffrey C. Wall, Pharm.D., FCCP, BCPS
Professor of Pharmacy Practice, Drake University
Internal Medicine Clinical Pharmacist
Iowa Methodist Medical Center

Darla Eastman, Pharm.D., BCPS
Associate Professor of Pharmacy Practice, Drake University
Trauma Clinical Pharmacist
Iowa Methodist Medical Center