Disclosure

- We do not have a vested interest in or affiliation with any corporate organization offering financial support or grant monies for this continuing education activity, or any affiliation with an organization whose philosophy could potentially bias my presentation.

Learning Objectives

1. Review the advantages and disadvantages of superiority and noninferiority study designs in the medical literature.
2. Highlight the most commonly used statistical tests in the medical literature.
3. Calculate the odds ratio (OR), hazard ratio (HR), number needed to treat (NNT), and number needed to harm (NNH) using clinical trials as examples.
4. Explain the difference between clinical versus statistical significance.

Quality of Trials Supporting FDA Approval

- More than half of trials used a placebo for comparison
- Majority of trials evaluating treatment for chronic diseases used study durations of < 1 year
- Most pivotal trials use surrogate markers for the primary outcome
- Median # of pivotal trials per indication is was 2
  - ~37% of indications gained approval with 1 trial
**CONSORT: Guidance For Critiquing and Conducting Clinical Trials**

**Null Hypothesis**
- There is NO difference between the study group and control group.

**Superiority**
- Most commonly encountered in literature.
- Designed to demonstrate that one treatment is better than another.

**Noninferiority**
- Designed to demonstrate that one treatment is no worse than an active (standard of care).

**Common Clinical Study Designs**

**Superiority**
- To determine if study drug is better than control.
- Intent to treat (ITT).

**Noninferiority**
- To determine if study drug is NO WORSE than control by a predefined margin.
- ITT and Per-protocol.
Statistics are one piece of the puzzle, not the entire picture.

“Statistics are like a bikini. What they reveal is suggestive, but what they conceal is vital.”
-Aaron Levenstein

Biostatistics7,8

- In providing EVIDENCE-BASED medicine, we highly rely on evidence from clinical trials to make recommendations in practice.

- Statistical Tests
- P values
- Not due to chance

Confidence Intervals7-9

- Confidence Intervals (CIs)
  - Provide an estimate of the range of values that is likely to represent the true mean of the universal population.
  - Two Types
    - Mean CIs: significant if it does not touch or cross “0”
      - Primary outcome: weight loss (kg)
      - 95% CI -2.4 kg (-3.1 - 2.4 kg)
    - Measures of Association CIs: significant if it does not touch or cross “1”
      - Primary outcome incidence of myocardial infarction
      - 95% CI HR 0.77 (0.71 - 0.85)

P values7-9

- P value
  - Probability that an observed effect is due to chance alone
  - When the correct statistical tests are employed, p \leq 0.05 typically indicates statistical significance

Alpha and Beta Errors1,7-9

<table>
<thead>
<tr>
<th>Decision from statistical test</th>
<th>Null False (There is a difference)</th>
<th>Null True (There is no difference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject Null (There is a difference)</td>
<td>Correct, acceptable level of error: The statistical test shows there is a difference when one exists (e.g., p \leq 0.05)</td>
<td>Incorrect. Type I (Alpha) error: The statistical test shows there is a difference when one does not exist (e.g., p &gt; 0.05), (false positive). Controlled for by use of the correct statistical test.</td>
</tr>
<tr>
<td>Accept Null (There is no difference)</td>
<td>Incorrect. Type II (Beta) error: Statistical test concludes there is no difference when one exists (e.g., p &gt; 0.05), (false negative). Possibly due to a lack of adequate power.</td>
<td>Correct, acceptable level of error: Statistical test concludes no difference when one does not exist. (e.g., p &gt; 0.05).</td>
</tr>
</tbody>
</table>

Stepwise approach to DDDIIG into statistics10-12

Assess the study Design: Is it independent (parallel) or dependent (crossover or repeated measures)?

Determine the Data types present: [IRON] - found in the baseline characteristics, outcome results, and adverse drug reactions
- interval/ratio (laboratory data)
- ordinal (rating/ Likert scales)
- nominal (yes/no; %)

Investigate if the Distribution is normal:
- Choosing between a parametric and nonparametric test depends on the assumptions made about the population from which the data were selected
  - Individuals in EACH group (n) < 30 or \geq 307 (at least 30 in EACH group)
  - 2 SD away from mean still within possible range? Mean = \pm 2 X SD
- For parametric tests, variances from the population should also be equal or near equal (i.e., baseline patient characteristics should be similar)

Investigate the number of Groups evaluated: 2 groups or \geq 2 groups?
EB is a 67 YO white male who presents to the ER complaining of gradual worsening of shortness of breath from his COPD. Glucocorticoids are one part of the pillar of treatment for the treatment of acute exacerbation. The physician asks if you reviewed the recent JAMA article that discusses 5 day vs. 14 day treatment. “You think we should try 5 days?”

Leuppi JD, et al. JAMA. 2013;309:2223-31

EB is a 67 YO white male who presents to the ER complaining of gradual worsening of shortness of breath from his COPD. Glucocorticoids are one part of the pillar of treatment for the treatment of acute exacerbation. The physician asks if you reviewed the recent JAMA article that discusses 5 day vs. 14 day treatment. “You think we should try 5 days?”

Leuppi JD, et al. JAMA. 2013;309:2223-31

EB is a 67 YO white male who presents to the ER complaining of gradual worsening of shortness of breath from his COPD. Glucocorticoids are one part of the pillar of treatment for the treatment of acute exacerbation. The physician asks if you reviewed the recent JAMA article that discusses 5 day vs. 14 day treatment. “You think we should try 5 days?”

Leuppi JD, et al. JAMA. 2013;309:2223-31
WW is a 55 YO Hispanic female who presents to a primary care clinic where she was recently diagnosed with type 2 diabetes (A1C 8%). WW has a past medical history significant for a myocardial infarction one year ago. In determining the optimal medication to initiate, you review an article published in Diabetes Care investigating cardiovascular benefits of using metformin versus a sulfonylurea. You question “Should we initiate metformin or a sulfonylurea?”


**SPREAD-DIMCAD Trial**

- **Patients**: 304 Chinese patients with type 2 diabetes and coronary artery disease
- **Intervention Comparison**
  - Metformin IR 0.75-1.5 g daily for 3 years
  - Glipizide IR 15-30 mg daily for 3 years
- **Outcomes**
  - 1°: Composite of recurrent CVD events
  - 2°: New or worsening: angina, heart failure, arrhythmia, peripheral vascular events; ADRs


**Calculating NNT and NNH**

- When it is appropriate to calculate
  - Superiority study design
  - Primary outcome nominal data type
  - Primary outcome statistically significant
  - Clinically relevant adverse drug reactions
- Information needed to calculate
  - Experimental event rate (EER) and control event rate (CER)
  - Study duration and dosing regimens
  - NNT (to prevent) and NNH (to cause)

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>NO</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EER</strong> Metformin</td>
<td>39</td>
<td>117</td>
<td>156</td>
</tr>
<tr>
<td>(25%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CER</strong> Glipizide</td>
<td>52</td>
<td>96</td>
<td>148</td>
</tr>
<tr>
<td>(35.1%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RR: 25% / 35.1% = 0.71
ARR: 35.1% - 25% = 10.1%
NNT: 1/0.101 = 9.9 = 10


**Do not lose sight of the forest for the trees!**

References


References