INTRODUCTION TO STATISTICAL CONCEPTS NEEDED FOR CLINICAL PHARMACOLOGY

Jose de Leon, MD

10/29/15

Educational Objectives

At the conclusion of this presentation, the participant should be able to understand some basic statistical concepts needed to interpret efficacy and safety articles. They include:

1. Number Needed to Treat/Harm
2. Statistical Significance
3. Confidence Interval
4. Relationship Between Significance and Confidence Interval
5. Meta-Analysis
6. Randomized Clinical Trial (RCT)

Warning

- If you hate “Statistics”, this presentation may be hazardous to your health since it has a substantial number of statistical concepts. Moreover, they are annoyingly repeated throughout the presentation to help with learning. When Dr. de Leon lectures residents face-to-face, he introduces these concepts as “tolerated” by the audience. In this case, you will need to set your pace for tolerance.
- Other presentations in this series have as little statistical material as possible.

Argument Against Avoiding Statistics

- Be aware that in the 21st century, it is not a good idea for you to use the words "in my experience" to answer basic pharmacological questions such as, *Which antipsychotic drug is better?"* which can be answered by reading published articles as long as the reader has some understanding of basic statistical concepts.
- To be an expert in psychiatry, a psychiatrist should have a basic understanding of statistics for interpreting published psychopharmacological studies.

Abbreviations (used in this presentation)

- ADR: adverse drug reaction
- AP: antipsychotic
- BPRS: Brief Psychiatric Rating Scale
- CATIE: Clinical Antipsychotic Trials of Intervention Effectiveness
- CGI: Clinical Global Impression
- CI: confidence interval
- FGAP: first-generation antipsychotic
- IM: intramuscular
- MDD: major depressive disorder
- NNH: number needed to harm
- NNT: number needed to treat
- OR: odds ratio
- PANSS: Positive and Negative Syndrome Scale
- RCT: Randomized Clinical Trial
- SGAP: second-generation antipsychotic
- SMD: standardized mean difference

AP Abbreviations (used in this presentation)

- AMI: amisulpiride, SGAP available in Europe
- ARI: aripiprazole
- ASE: asenapine
- CLO: clozapine
- CPZ: chlorpromazine
- HAL: haloperidol
- ILO: iloperidone
- LUR: lurasidone
- OLA: olanzapine
- PAL: paliperidone
- QUE: quetiapine
- RIS: risperidone
- ZIP: ziprasidone

Definitions of Efficacy and Safety

- Efficacy is how well the desired effect is obtained in the patient.
- Safety's goal is to avoid adverse drug reactions (ADRs). Psychiatric textbooks use the old terminology "side effects" instead of ADRs.

These two concepts have been described in the prior presentation "Introduction to Clinical Pharmacology".

Lecture Content

1.Defining Statistical Concepts on Efficacy/Safety

- 1.1. Number Needed to Treat/Harm
- 1.2. Statistical Significance
- 1.3. Confidence Interval (CI)
- 1.4. Relationship Between Significance and CI

2.Practicing Statistical Concepts on AP Efficacy

- 2.1. Meta-Analysis
- 2.2. Randomized Clinical Trial (RCT)
- 2.3. NNTs To Compare AP RCTs
- 2.4. Standardized Mean Difference (SMD)

3.Practicing Statistical Concepts on AP Safety

- 3.1. Other Statistical Terms
- 3.2. SMDs in Weight Gain
- 3.3. Odds Ratios (ORs) in Extrapyramidal Symptoms
- 3.4. On the Concept of ORs

1. Defining Statistical Concepts on Efficacy/Safety

1. Defining Statistical Concepts on Efficacy/Safety

1.1. Number Needed to Treat/Harm

1.2. Statistical Significance

1.3. Confidence Interval (CI)

1.4. Relationship between Significance and CI

1.1. Number Needed to Treat/Harm

1.1. Number Needed to Treat/Harm

- The outcome of patients in a randomized clinical trial (RCT) are summarized by:
  - The number needed to treat (NNT), which reflects whether they were *improved* or *not*. This is a measure of efficacy.

    And
  - The number needed to harm (NNH), which reflects whether they were *harmed* or *not*. This is a measure of safety.

1.1. Number Needed to Treat/Harm

- More precisely,
  NNT/NNH = the number of patients needed to treat to achieve one additional favorable/unfavorable outcome when compared with placebo (or other drug).

1.1. Number Needed to Treat/Harm

http://www.medscape.org/viewarticle/748336

- Imagine a response rate on AP: 45% (0.45)
  placebo: 15% (0.15)
- To calculate the NNT:
  - difference: 0.30 (0.45-0.15)
  - inverse: 1/0.30= 3.3
  - round up: NNT= 4

You need at least 4 patients on the drug to see an additional response beyond the placebo response.

1.1. Number Needed to Treat/Harm

http://www.medscape.org/viewarticle/748336

- In psychopharmacology-approved drugs: NNT usually <10.
- Antipsychotic NNTs:
  - IM SGAP vs. placebo: NNT as low as 2
  - SGAP for adjuvant treatment for MDD vs. placebo: NNT around 10
  - NNT for SGAP vs. FGAP is > NNT for SGAP vs. placebo. FGAPs provide better response than placebo and they are intermediate: some SGAP>FGAP>placebo.

1.1. Efficacy and Safety: Statistics

- To properly frame the issue in statistical language, remember that you are comparing a difference between:
  - a drug versus
  - placebo (or other drug).
- This difference can be:
  - significant or not, and
  - large or small, this is called by statisticians: effect size.

Both concepts are needed to understand a third concept, confidence interval (CI).

1.2. Statistical Significance

1.2. Statistical Significance

Can you explain
in your own words
what we mean
when we say
that a NNT is
statistically significant?

1.2. Statistical Significance

Can you explain
in your own words
what we mean
when we say
that a NNT is
statistically significant?

We have trouble if you answered, “No”.

1.2. Statistical Significance

- We have trouble because:
  - You are like most physicians.
  - You studied statistics to pass your exams.
  - You have not used statistical concepts enough to incorporate them into your language.
- Dr. de Leon is not sure he can teach you these concepts in this presentation to the point that you can use them.

1.2. Statistical Significance

- Let's recapitulate. You probably know:
  - To publish a study, you need to get a "statistically significant" value with a $p<0.05$.
- You probably do not know:
  - This is an arbitrary convention started by statisticians long ago, and
  - A NNT with a p value = 0.02 simply means that there is a probability of 2/100 (or 2%) that this NNT is explained by chance. In simple words, if it is explained by chance, the difference may not be "real".

1.2. Statistical Significance

- You compare drugs vs. placebo and have:
  - Drug A: NNT of 8 with a p value=0.02, and
  - Drug B: NNT of 80 with a p value=0.02.

1.2. Statistical Significance

What can you say about

a NNT=8 (p=0.02)

and

a NNT=80 (p=0.02)?

1.2. Statistical Significance

- Both NNTs have a p value = 0.02:
  - These differences are not likely to be explained by chance.
  - As a matter of fact, there is only 2% probability that they are explained by chance.
  - In summary, both NNTs are “significant” according to the standard convention $p<0.05$.
- Both significant NNTs are measures of the effect size. The effect size is what should worry clinicians, once they know that a statistical difference is significant.

1.2. Statistical Significance

Can you compare
a NNT=8 (p=0.02) in drug A
and
a NNT=80 (p=0.02) in drug B?

1.2. Statistical Significance

- Both differences are significant but their effect sizes are different.
- The effect sizes are:
  - relatively large for drug A (NNT=8), and
  - relatively small for drug B (NNT=80).
- For drug A, NNT=8 means that
  - you only need to study 8 patients with the drug to see a difference compared with placebo.
- For drug B, NNT=80 means that
  - you need to study 80 patients with the drug to see a difference from the placebo.

1.2. Statistical Significance

- Very large studies >1000 patients can easily provide significant differences.
  Be careful, these differences may be small (small effect sizes) and may be irrelevant in the clinical environment.
- It is difficult to find statistical differences in small studies < 25 patients. These significant differences need to have large effect sizes.
  Be careful; other studies may not be able to replicate the significant results first estimated in small studies.

1.2. Statistical Significance

- To plan a study you need to find the right size for providing a significant difference. Statisticians call this estimating the statistical power of the sample.
- Another problem with a statistical difference is that a p value reflects the probability that a specific comparison is significant. If you do multiple comparisons you may need to adjust/correct by the number of comparisons. Statisticians use different techniques to control for multiple comparisons. Multivariate analysis is described in Section 3.

1.2. Statistical Significance

- It is important to reach statistical significance to interpret an effect size. If you have an effect size that is large but it is not significant you cannot comment on it since it may be explained by chance.

- One way to increase the possibility of significance of an effect size that is consistent in multiple small studies but not reaching significance in the individual studies is to combine the small studies in a meta-analysis (discussed in Section 2).

1.3. Confidence Interval (CI)

1.3. Confidence Interval (CI)

- Framing the issue in statistical language: NNT/NNH is a measure of effect size and needs to be accompanied by a CI.
- CI is a measure of accuracy and variability:
  - very large studies provide small CIs, and
  - small studies provide large CIs.
- There is a convention to report 95 % CIs, which are somewhat related to significance.

1.3. Confidence Interval (CI)

Can you explain
in your own words
the meaning of a 95% CI
of an NNT=8
in an RCT?

1.3. Confidence Interval (CI)

- If you do, Dr. de Leon is wrong; there is hope that physicians know statistics.
- If you do not know, let's imagine you have a real difference not explained by chance:
  - You repeat another RCT with the same drug and placebo.
  - Even assuming the NNT=8 is correct, in this second RCT it is not likely you are going to get exactly an NNT=8. It may be around 8, it may be 6, or it may be 10.
- A 95% CI tells you how close it will be to 8.

1.3. Confidence Interval (CI)

- You repeat same RCT with
  - the same drug versus placebo,
  - in the same type of patient, and
  - using the same design.
- If you repeat it 100 times,
  - statistical estimation tells you that in 95 of 100 RCTs, the NNT is between 6 and 10. This range "6 to 10" is the 95% CI. This range also can be presented as:
    - NNT=8 (95% CI 6-10) or
    - NNT=8 (95% CI 6, 10). This presentation style with "commas" is better than the '-', because there is no confusion with negative values. In this presentation we will use the "to".

1.3. Confidence Interval (CI)

Can you explain
in your own words
what is the meaning of an
NNT=8 (95% CI 6 to 10)
in an RCT?

1.3. Confidence Interval (CI)

- The NNT was 8 in this RCT.
  - In 95 of 100 repeated RCTs with similar characteristics, the NNT would be between 6 and 10.

    OR
  - There is a 95% chance that the NNT is between 6 and 10.

1.4. Relationship between Significance and CI

1.4. Relationship between Significance and CI

- In the prior example:
  - NNT=8 (95% CI 6 to 10)
  - Notice that the CI range is between 6 and 10.

    Between 6 and 10 you can find 7, 8 and 9, but not 1. This tells you the CI is significant with a p value <0.05.

1.4. Relationship between Significance and CI

- The approximate rule for the relationship between a significant NNT with a p value <0.05 and 95% CI is:
  - When the 95% CIs do not include 1, NNT/NNHs are significant.
  - When the 95% CIs include 1, NNT/NNHs are not significant.

1.4. Relationship between Significance and CI

- Two real examples from a summary of the NNTs from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE)

http://www.ncbi.nlm.nih.gov/pubmed/20428308

- CATIE phase IIE compared discontinuation rates from all Phase 1s, due to inadequate efficacy or other reasons.
  The patients were re-randomized to:
  - open-label CLO, or
  - double-blinded RIS,
    OLA, or
    QUE.

NNTs compare the discontinuation rates.

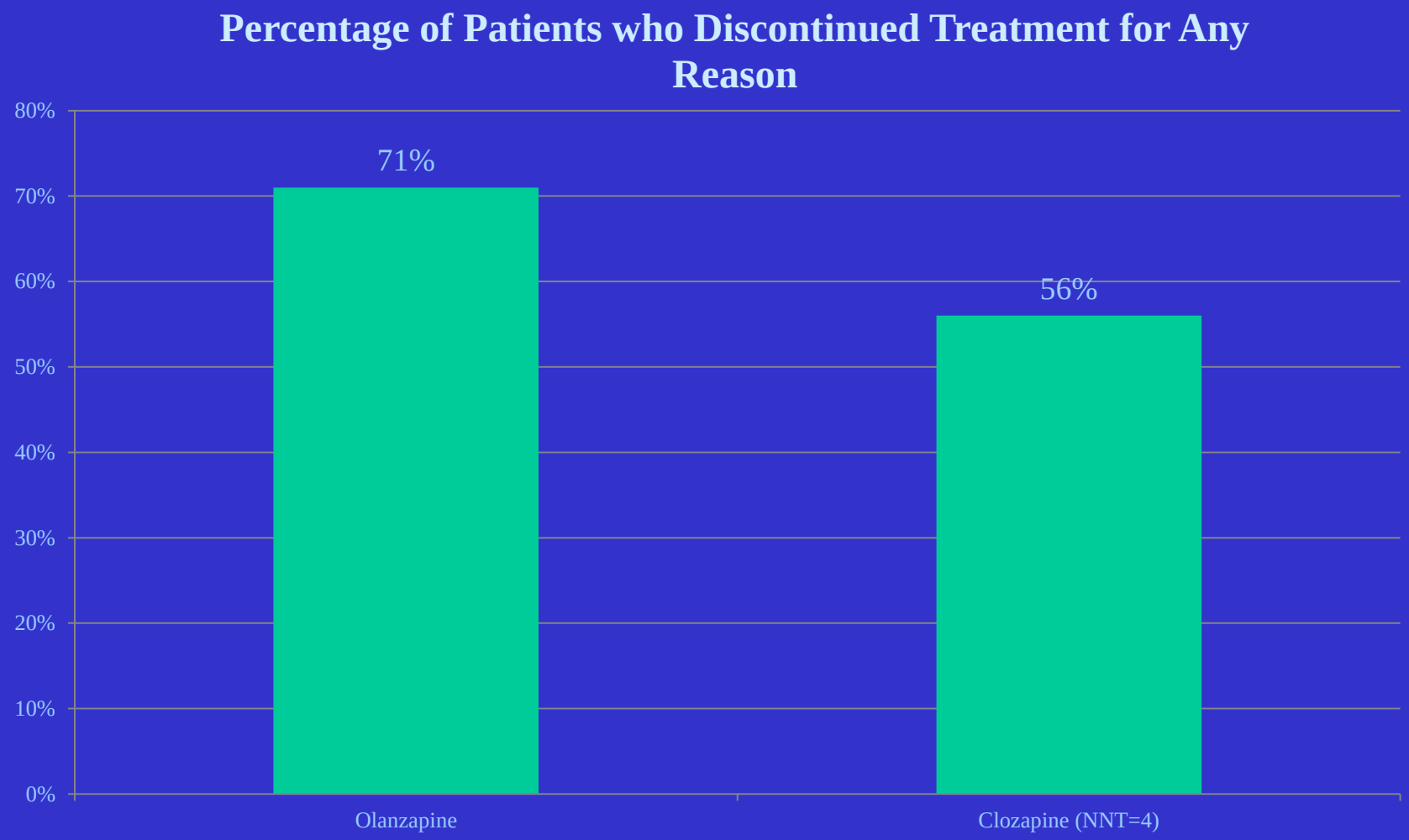
1.4. Relationship between Significance and CI

http://www.ncbi.nlm.nih.gov/pubmed/18479317

- Comparison of CLO vs. OLA:
  (discontinuation rate: 56% vs. 71%)
  The NNT=7 (95% CI -3 to 10).
  Notice that the 95% CI includes 1.
  1 is between -3 and 10; thus, the NNT is not significant. Notice that the p value is not reflected in the figure in the next slide.

1.4. Relationship between Significance and CI

CATIE Phase IIE http://www.ncbi.nlm.nih.gov/pubmed/18479317

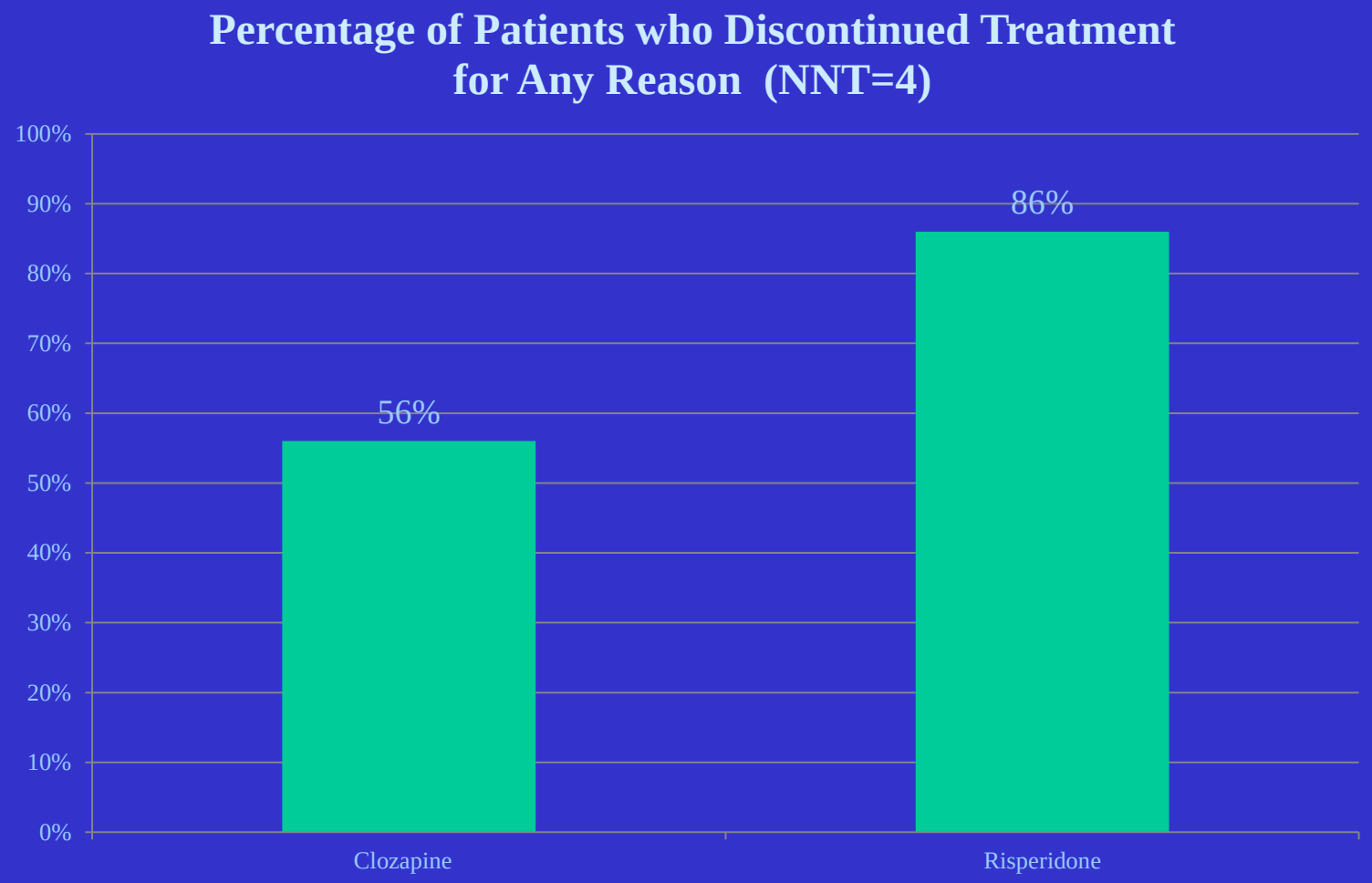


1.4. Relationship between Significance and CI

http://www.ncbi.nlm.nih.gov/pubmed/18479317

- Comparison of CLO vs. RIS:
  (discontinuation rate: 56% vs. 86%)
  NNT=4 (95% CI 2 to 15).
  Notice that the 95% CI does not include 1.
  1 is NOT between 2 and 15; thus, the NNT is significant and is reflected in the figure as “NNT 4”.

1.4. Relationship between Significance and CI CATIE Phase IIE http://www.ncbi.nlm.nih.gov/pubmed/18479317



2. Practicing Statistical Concepts on Efficacy

2.3.1.1. Efficacy: Statistical Concepts

- The next slides use NNTs to summarize AP efficacy. They are an example of how statistics can be an excellent instrument for summarizing drug efficacy.
- Meta-analyses frequently use NNTs.
- This second set of statistical slides discusses the concepts of:
  - meta-analysis,
  - randomized clinical trials (RCTs),
  - using NNTs to compare AP RCTs, and
  - standardized mean difference (SMD).

2. Practicing Statistical Concepts on Efficacy

- 2.1. Meta-Analysis
- 2.2. Randomized Clinical Trial (RCT)
- 2.3. NNTs to compare AP RCTs
- 2.4. Standardized Mean Difference (SMD)
- 2.5. Interpreting Multiple SMDs

2.1. Meta-Analysis

2.1. Meta-Analysis

What is a meta-analysis?

2.1. Meta-Analysis

- Meta-analysis http://www.ncbi.nlm.nih.gov/mesh/68015201
  - is a quantitative method of:
    - combining the results of independent studies (usually drawn from the published literature) and
    - synthesizing summaries and conclusions
  - which may be used to:
    - evaluate therapeutic effectiveness,
    - plan new studies, etc.,
  - with application chiefly in the areas of
    - research and
    - medicine.

2.1. Meta-Analysis

- A basic concept that even a physician can remember:
  - RCTs are summarized and averaged by meta-analysis.

2.2. Randomized Clinical Trial (RCT)

2.2. Randomized Clinical Trial (RCT)

Write how you define a randomized clinical trial.

2.2. Randomized Clinical Trial (RCT)

- Two concepts are used in the literature and are abbreviated as RCT:
  - randomized clinical trial (RCT), and
  - randomized controlled trial (RCT).

  Both are synonymous. Dr. de Leon will be using randomized clinical trial for RCT.
- Important words to remember and that your RCT definition should include:
  - "comparing" treatments,
  - under "controlled" conditions,
  - by using "random" treatment selection.

2.2. Randomized Clinical Trial (RCT)

- PubMed uses the words "randomized controlled trial" and defines it as:
  Clinical trials that involve
  - at least
    - one test treatment and
    - one control treatment,
  - concurrent enrollment and
  - follow-up of
    - the test- and
    - control-treated groups,

    and
  - in which the treatments to be administered are selected by a random process, such as the use of a random-numbers table.

http://www.ncbi.nlm.nih.gov/mesh/68016032

2.2. Randomized Clinical Trial (RCT)

- The problem with the statistical measures described in this presentation to summarize RCTs is that they measure average effects. This mean effect may not apply to a specific patient. Some patients may improve more or less than the average.

- Another problem with NNT or other statistical measures obtained in RCTs is that they only reflect the duration of the trial. Most drug trials only last a few weeks. CATIE was a pragmatic trial; therefore, it was longer than pharmaceutical AP RCTs.

2.2. Randomized Clinical Trial (RCT)

- The relevance of RCTs and meta-analysis is reviewed in the PowerPoint presentation on evidence-based medicine.
- That presentation also discusses the problems of applying these concepts to the treatment of individual patients (personalized medicine).
- You will have to decide if you can take more "statistical torture" or not in the form of the presentation or article in which is based on. The presentation is easier than the article, that has no abstract: http://www.ncbi.nlm.nih.gov/pubmed/22367661. The article pdf is available

http://uknowledge.uky.edu/psychiatry_facpub/41/

2.3. NNTs to Compare AP RCTs

2.3. NNTs to Compare AP RCTs

- Leucht et al. published multiple AP meta-analyses: In 2009 they published one comparing SGAPs vs. FGAPs in *Lancet* http://www.ncbi.nlm.nih.gov/pubmed/19058842
  - CLO, OLA & RIS were better than FGAPs
    They are presented in order of NNTs:
    (*response = a ↓of 50% on BPRS/PANS scores or a much-improved CGI rating)*
    The order is CLO >OLA >RIS
    NNT: 7 > 11 > 15
    Remember, the lower the NNT, the larger the effect size of the drug.
  - ARI, QUE & ZIP were not better than FGAPs.

2.3. NNTs to Compare AP RCTs

- Leucht et al. published another 2009 meta-analysis in the *Am J Psychiatry* http://www.ncbi.nlm.nih.gov/pubmed/19015230

They compared SGAPs among themselves:

- OLA was better than: ARI, QUE, RIS & ZIP.
- RIS was better than: QUE & ZIP.
- CLO in doses >400 mg/d was better than: RIS.

2.3. NNTs to Compare AP RCTs

- More recently, Citrome reviewed new SGAPS:
  - Iloperidone (ILO) vs. placebo:
    - NNT=5 (for a response ≥ **20**% of the total scale)
  - Asenapine (ASE) vs. placebo:
    - NNT=6 (for a response ≥ **20**% of the total scale)
    - NNT=8 (*for a response ≥ **30**% of the total scale*)
  - Lurasidone (LUR) vs. placebo:
    - NNT=3-6 (for a response ≥ **20**% of the total scale)
    - NNT=7-13 (*for a response ≥ **30**% of the total scale*).

http://www.medscape.org/viewarticle/748336

2.4. Standardized Mean Difference (SMD)

2.4. Standardized Mean Difference (SMD)

- If you have reviewed any of the Leucht et al. AP meta-analyses from 2009, you have seen they also use another way of summarizing results called: standardized mean difference (SMD).
- Remember, NNT compares response percentages. That is a dichotomous variable: □ response, or
  □ no response.
- If you are using the total from a psychiatric scale:
  □ BPRS, or
  □ PANSS,
  it is a continuous numerical variable. For comparing means or mean decrease from baseline, you use standardized mean differences (SMDs).

2.4. Standardized Mean Difference (SMD)

- The concept of standardized mean difference (SMD) appears to be too complex for Dr. de Leon's residents, or he is too dumb to explain it correctly.
- Please remember two things about SMDs:
  - It is used in meta-analyses
  - to compare means in a standardized way.

2.4. Standardized Mean Difference (SMD)

- Recapitulating NNT/NNH:
  - When the 95% CIs do not include 1, NNT/NNHs are significant.
  - When the 95% CIs include 1, NNT/NNHs are not significant.
- SMDs are similar but not the same:
  - When the 95% CIs do not include **0**, SMDs are significant.
  - When the 95% CIs include **0**, SMDs are not significant.

2.5. Interpreting Multiple SMDs

2.5. Interpreting Multiple SMDs

Can you interpret the following 13 SMDs?

2.5. Interpreting Multiple SMDs

- Leucht et al. 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  SMDs in order (95% CI) (drug versus placebo)
  - CLO: -0.88 (-1.03 to -0.73)
  - AMI: -0.66 (-0.78 to -0.53)
  - OLA: -0.59 (-0.65 to -0.53)
  - RIS: -0.56 (-0.63 to -0.50)
  - PAL: -0.50 (-0.60 to -0.39)
  - HAL: -0.45 (-0.51 to -0.31)
  - QUE: -0.44 (-0.52 to -0.35)
  - ARI: -0.43 (-0.52 to -0.34)
  - ZIP: -0.39 (-0.49 to -0.30)
  - CPZ: -0.38 (-0.54 to -0.23)
  - ASE: -0.38 (-0.51 to -0.25)
  - LUR: -0.33 (-0.45 to -0.21)
  - ILO: -0.33 (-0.43 to -0.22)

2.5. Interpreting Multiple SMDs

- Please do not panic; relax, breathe.
- Start remembering three issues regarding SMDs:
  - They are a measure of the effect size of the difference (in this case compared with placebo).
  - They can be significant or non-significant.
  - The SMD's 95% CI tell you about the significance of that SMD.

    Moreover, 95% CIs can be used to compare SMDs and whether the comparison is significant or not.
- The following slides review these steps one at a time.

2.5. Interpreting Multiple SMDs

2.5.1. Range of SMDs

2.5.2. Significance vs. Placebo

2.5.3. Comparison Among SMDs

2.5.1. Range of SMDs

2.5.1. Range of SMDs

What is the range of SMDs?

(CIs are deleted to simplify.)

2.5.1. Range of SMDs

- Leucht et al. 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  SMDs in order (95% CI) (drug versus placebo)
  - CLO: -0.88
  - AMI: -0.66
  - OLA: -0.59
  - RIS: -0.56
  - PAL: -0.50
  - HAL: -0.45
  - QUE: -0.44
  - ARI: -0.43
  - ZIP: -0.39
  - CPZ: -0.38
  - ASE: -0.38
  - LUR: -0.33
  - ILO: -0.33

2.5.1. Range of SMDs

- Go back and look at the values. All are negative, reflecting that AP ↓ symptoms more than placebo.
- Values range between -1 and 0, but did not reach 0. A 0 means that the drug is not different from placebo.
- Remember, SMDs a measure of the effect size:
  - If they are close to -1: large effect size. The largest is CLO: -0.88.
  - If they are close to 0: smaller effect size. The smallest is ILO: -0.33.

2.5.2. SMD: Significance of Comparison Drug vs. Placebo

2.5.2. SMDs' Significance

Are these SMDs statistically significantly different from placebo?

2.5.2. SMDs' Significance

- Leucht et al. 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  SMDs in order (95% CI) (drug versus placebo)
  - CLO: -0.88 (-1.03 to -0.73)
  - AMI: -0.66 (-0.78 to -0.53)
  - OLA: -0.59 (-0.65 to -0.53)
  - RIS: -0.56 (-0.63 to -0.50)
  - PAL: -0.50 (-0.60 to -0.39)
  - HAL: -0.45 (-0.51 to -0.31)
  - QUE: -0.44 (-0.52 to -0.35)
  - ARI: -0.43 (-0.52 to -0.34)
  - ZIP: -0.39 (-0.49 to -0.30)
  - CPZ: -0.38 (-0.54 to -0.23)
  - ASE: -0.38 (-0.51 to -0.25)
  - LUR: -0.33 (-0.45 to -0.21)
  - ILO: -0.33 (-0.43 to -0.22)

2.5.2. SMDs' Significance

- Example: ILO: -0.33 (-0.43 to -0.22)
  0 is not found between -0.43 and -0.22.
- Go back and look at the 95% CI values:
  None of the 95% CIs include 0.
  All of these SMDs are significant.

2.5.3. Comparing SMDs

2.5.3. Comparing SMDs

How can you compare SMDs?

2.5.3. Comparing SMDs

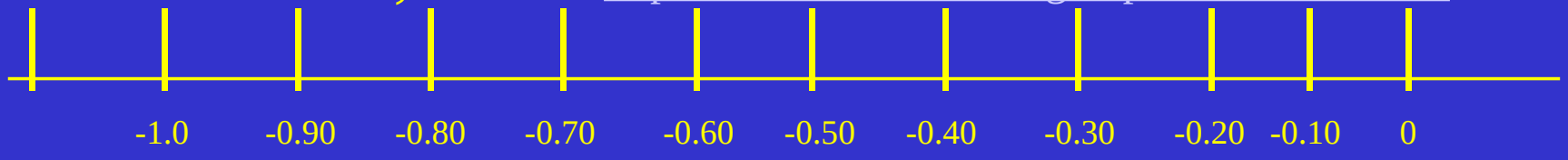
- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  SMDs in order (95% CI) (drug versus placebo)
  - CLO: -0.88 (-1.03 to -0.73)
  - AMI: -0.66 (-0.78 to -0.53)
  - OLA: -0.59 (-0.65 to -0.53)
  - RIS: -0.56 (-0.63 to -0.50)
  - PAL: -0.50 (-0.60 to -0.39)
  - HAL: -0.45 (-0.51 to -0.31)
  - QUE: -0.44 (-0.52 to -0.35)
  - ARI: -0.43 (-0.52 to -0.34)
  - ZIP: -0.39 (-0.49 to -0.30)
  - CPZ: -0.38 (-0.54 to -0.23)
  - ASE: -0.38 (-0.51 to -0.25)
  - LUR: -0.33 (-0.45 to -0.21)
  - ILO: -0.33 (-0.43 to -0.22)

2.5.3. Comparing SMDs

- We are going to simplify by only using the CI of 4 SMDs.
- Moreover, we are going to represent them in the space between -1 to 0.

2.5.3. Comparing SMDs

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019



-1.03 to -0.73

CLO

-0.63,-0.53

RIS

-0.43,-0.34

ARI

-0.45 to -0.21

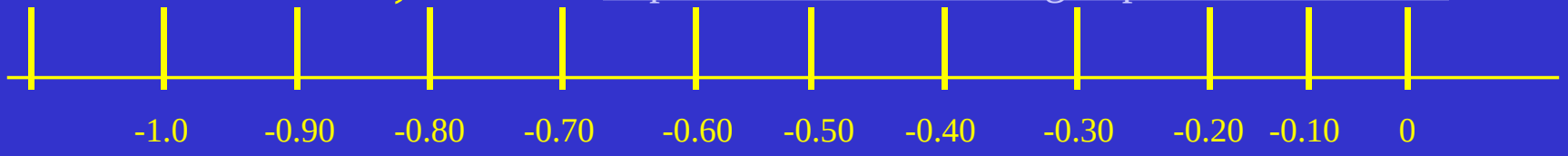
LUR

2.5.3. Comparing SMDs

- To compare 95% CIs, notice whether they overlap or not:
  - If two 95% CIs do **NOT** overlap, it can be approximated that they are significantly different with a $p<0.05$.
  - If two 95% CIs overlap, it can be approximated that they are **NOT** significantly different.
- Let's start with the largest effect size: CLO. Important values are in red:
  - lowest range of CI for CLO and
  - highest range of other CIs.

2.5.3. Comparing SMDs

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019



-1.03 to -0.73

CLO

-0.63,-0.53

RIS -0.63 does not overlap with -0.73 of CLO.

CLO and RIS are significantly different.

-0.43,-0.34

ARI -0.43 does not overlap with -0.73.

CLO and ARI are significantly different.

-0.45 to -0.21

LUR -0.45 does not overlap with -0.73.

CLO and LUR are significantly different.

Moreover, CLO is significantly better than RIS, ARI and LUR.

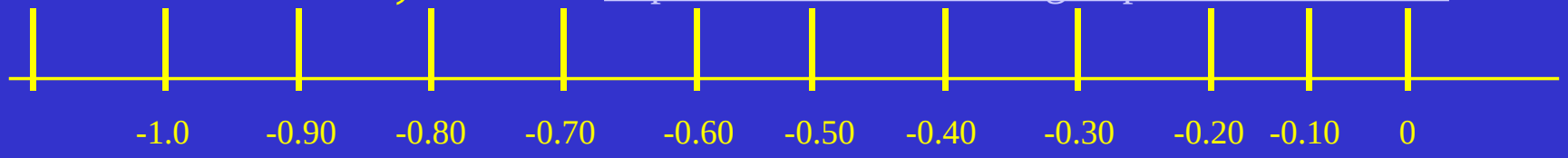
2.5.3. Comparing SMDs

- The three next slides will compare:
  - RIS CI with others,
  - ARI CI with others, and
  - LUR CI with others.

  Red is used for the AP's high CI range and those that will be compared.
  Light blue is used for the AP's low CI range and those that will be compared.

2.5.3. Comparing SMDs

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019



-1.03 to -0.73

CLO -0.73 does not overlap with -0.63 of RIS

RIS is significantly worse than CLO.

-0.63,-0.53

RIS

-0.43,-0.34

ARI -0.43 does not overlap with -0.53

RIS is significantly better than ARI.

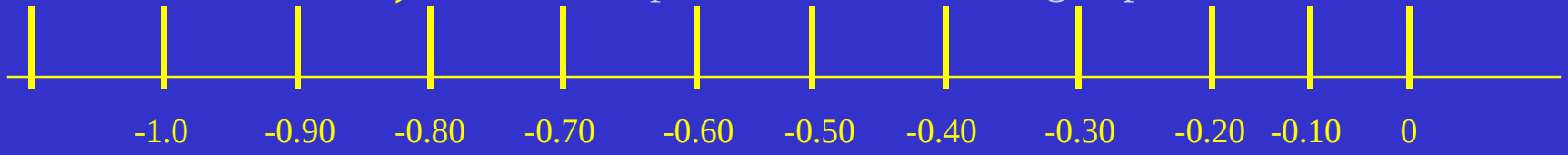
-0.45 to -0.21

LUR -0.45 does not overlap with -0.53

RIS is significantly better than LUR.

2.5.3. Comparing SMDs

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019



-1.03 to -0.73

CLO -0.73 does not overlap with -0.43 of ARI

ARI is significantly worse than CLO.

-0.63,-0.53

RIS -0.53 does not overlap with -0.43 of ARI

ARI is significantly worse than RIS.

-0.43,-0.34

ARI

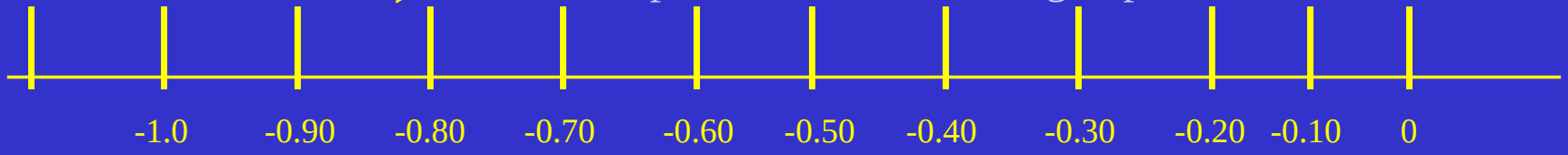
-0.45 to -0.21

LUR -0.45 overlaps with -0.34

ARI and LUR are not significantly different.

2.5.3. Comparing SMDs

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019



-1.03 to -0.73

CLO -0.73 does not overlap with -0.45 of LUR.
LUR is significantly worse than CLO.

-0.63, -0.53

RIS -0.53 does not overlap with -0.45 of LUR.
LUR is significantly worse than RIS.

-0.43, -0.34

ARI -0.34 overlaps with -0.45 of LUR.

ARI and LUR are not significantly different.

-0.45 to -0.21

LUR

3. Practicing Statistical Concepts on Safety

3. Practicing Statistical Concepts on Safety

- 3.1. Other Statistical Terms
- 3.2. SMDs in Weight Gain
- 3.3. Odds Ratios (Ors) in Extrapyramidal Symptoms
- 3.4. On the concept of Odds Ratios

3.1. Other Statistical Terms

3.1. Other Statistical Terms

- If you look for AP meta-analyses in PubMed, you will see:
  - Leucht et al. has published many of them.
  - They love statistics and describe other statistical terms beyond NNTs/NNHs or SMDs. They appear to be trying to strain psychiatrists' brains.

3.1. Other Statistical Terms

- Dr. de Leon is not confident in his skills when trying to explain to psychiatrists about all the possible statistical terms that could be found in meta-analysis, such that psychiatrists would not become confused.
- However, the rules for approximating 95% CIs and significance are consistent.

3.1. Other Statistical Terms

- NNT/NNH are ratios:
  - When the 95% CIs do not include 1, NNT/NNH are significant.
  - When the 95% CIs include 1, NNT/NNH are not significant.
- This approximate rule applies to other ratios.
  Other ratios are:
  - risk ratios (RRs),
  - relative risks (RRs),
  - odds ratios (ORs)…

3.1. Other Statistical Terms

- SMDs are differences between means:
  - When the 95% CIs do not include 0, SMDs are significant.
  - When the 95% CIs include 0, SMDs are not significant.
- This approximate rule applies to other mean differences. Other mean differences are:
  - weighted mean differences (WMDs)…

3.2. SMDs in Weight Gain

3.2. SMDs in Weight Gain

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  SMDs in order (95% CI) (drug versus placebo)
  - OLA: 0.74 (0.67 to 0.81)
  - CLO: 0.65 (0.31 to 0.99)
  - ILO: 0.62 (0.49 to 0.74)
  - CPZ: 0.55 (0.34 to 0.76)
  - QUE: 0.43 (0.34 to 0.53)
  - RIS: 0.42 (0.33 to 0.50)
  - PAL: 0.38 (0.27 to 0.48)
  - ASE: 0.23 (0.07 to 0.31)
  - AMI: 0.20 (0.05 to 0.35)
  - ARI: 0.17 (0.05 to 0.28)
  - LUR: 0.10 (-0.02 to 0.21)
  - ZIP: 0.10 (-0.02 to 0.22)
  - HAL: 0.09 (-0.00 to 0.17)

2.5.2. SMDs' Significance

Which of these SMDs are statistically different from placebo?

2.5.2. SMDs' Significance

- When the 95% CI does not include 0, SMDs are significant.
  Mark significant SMDs in red.

3.2. SMDs in Weight Gain

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  SMDs in order (95% CI) (drug versus placebo)
  - OLA: 0.74 (0.67 to 0.81)
  - CLO: 0.65 (0.31 to 0.99)
  - ILO: 0.62 (0.49 to 0.74)
  - CPZ: 0.55 (0.34 to 0.76)
  - QUE: 0.43 (0.34 to 0.53)
  - RIS: 0.42 (0.33 to 0.50)
  - PAL: 0.38 (0.27 to 0.48)
  - ASE: 0.23 (0.07 to 0.31)
  - AMI: 0.20 (0.05 to 0.35)
  - ARI: 0.17 (0.05 to 0.28)
  - LUR: 0.10 (-0.02 to 0.21)
  - ZIP: 0.10 (-0.02 to 0.22)
  - HAL: 0.09 (-0.00 to 0.17)

2.5.2. SMDs' Significance

- See HAL: 0.09 (-0.00 to 0.17)
  The lower CI of 0.0 means p is approximately 0.05.
  In this situation different methods of calculating significance can provide small variations.
  - With one method HAL may be statistically different from placebo because $p<0.05$.
  - With another method HAL may not be statistically different from placebo because p was NOT <0.05.

2.5.2. SMDs' Significance

Which of these SMDs are statistically significantly different from PAL SMD?

3.2. SMDs in Weight Gain

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  SMDs in order (95% CI) (drug versus placebo)
  - OLA: 0.74 (0.67 to 0.81)
  - CLO: 0.65 (0.31 to 0.99)
  - ILO: 0.62 (0.49 to 0.74)
  - CPZ: 0.55 (0.34 to 0.76)
  - QUE: 0.43 (0.34 to 0.53)
  - RIS: 0.42 (0.33 to 0.50)
  - PAL: 0.38 (0.27 to 0.48)
  - ASE: 0.23 (0.07 to 0.31)
  - AMI: 0.20 (0.05 to 0.35)
  - ARI: 0.17 (0.05 to 0.28)
  - LUR: 0.10 (-0.02 to 0.21)
  - ZIP: 0.10 (-0.02 to 0.22)
  - HAL: 0.09 (-0.00 to 0.17)

2.5.2. SMDs' Significance

- When the AP's SMD (95% CI, lower range) does not overlap with PAL SMD (95% CI, upper range, 0.48), then
  AP SMDs are significantly higher than PAL SMD.
  Mark them in red.
- When the AP's SMD (95% CI, upper range) does not overlap with PAL SMD (95% CI, lower range, 0.27), then
  AP SMDs are significantly lower than PAL SMD.
  Mark them in light blue.

3.2. SMDs in Weight Gain

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  SMDs in order (95% CI) (drug versus placebo)
  - OLA: 0.74 (0.67 to 0.81)
  - CLO: 0.65 (0.31 to 0.99)
  - ILO: 0.62 (0.49 to 0.74)
  - CPZ: 0.55 (0.34 to 0.76)
  - QUE: 0.43 (0.34 to 0.53)
  - RIS: 0.42 (0.33 to 0.50)
  - PAL: 0.38 (0.27 to 0.48)
  - ASE: 0.23 (0.07 to 0.31)
  - AMI: 0.20 (0.05 to 0.35)
  - ARI: 0.17 (0.05 to 0.28)
  - LUR: 0.10 (-0.02 to 0.21)
  - ZIP: 0.10 (-0.02 to 0.22)
  - HAL: 0.09 (-0.00 to 0.17)

3.3. Odds Ratios (ORs) in Extrapyramidal Symptoms

3.3. ORs in Extrapyramidal Symptoms

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  ORs in order (95% CI) (drug versus placebo)
  - HAL: 4.76 (3.70 to 6.04)
  - CPZ: 2.65 (1.33 to 4.76)
  - LUR: 2.46 (1.55 to 3.72)
  - RIS: 2.09 (1.54 to 2.78)
  - PAL: 1.81 (1.17 to 2.69)
  - ASE: 1.66 (0.85 to 2.93)
  - ZIP: 1.61 (1.05 to 2.37)
  - AMI: 1.60 (0.88 to 2.65)
  - ILO: 1.58 (0.55 to 3.65)
  - ARI: 1.20 (0.73to 1.85)
  - QUE: 1.01 (0.68 to 1.44)
  - OLA: 1.00 (0.73 to 1.33)
  - CLO: 0.3 (0.12 to 0.62)

3.3. ORs in Extrapyramidal Symptoms

Which of these ORs are statistically significantly different from placebo?

3.3. ORs in Extrapyramidal Symptoms

- When the 95% CIs do not include 1,
  ORs are significant.
  Remember ORs are ratios.
  - Mark those significantly higher than placebo in red.
  - Mark those significantly lower than placebo in light blue.

3.3. Odds Ratios in Extrapyramidal Symptoms

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  ORs in order (95% CI) (drug versus placebo)
  - HAL: 4.76 (3.70 to 6.04)
  - CPZ: 2.65 (1.33 to 4.76)
  - LUR: 2.46 (1.55 to 3.72)
  - RIS: 2.09 (1.54 to 2.78)
  - PAL: 1.81 (1.17 to 2.69)
  - ASE: 1.66 (0.85 to 2.93)
  - ZIP: 1.61 (1.05 to 2.37)
  - AMI: 1.60 (0.88 to 2.65)
  - ILO: 1.58 (0.55 to 3.65)
  - ARI: 1.20 (0.73to 1.85)
  - QUE: 1.01 (0.68 to 1.44)
  - OLA: 1.00 (0.73 to 1.33)
  - CLO: 0.3 (0.12 to 0.62)

3.3. Odds Ratios in Extrapyramidal Symptoms

- Please see ASE: 1.66 (0.85 to 2.93)
  OR is not significant.
  If more studies are done, it will likely become significant.
- Please see CLO: 0.3 (0.12 to 0.62)
  According to this meta-analysis OR, CLO is significantly better than placebo for extrapyramidal symptoms.
  Some data on Parkinson disease suggest that it may be true; CLO may help with parkinsonian tremors.

http://www.ncbi.nlm.nih.gov/pubmed/22095576

3.3. ORs in Extrapyramidal Symptoms

Which of these ORs are statistically significantly different from ZIP OR?

3.3. ORs in Extrapyramidal Symptoms

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  ORs in order (95% CI) (drug versus placebo)
  - HAL: 4.76 (3.70 to 6.04)
  - CPZ: 2.65 (1.33 to 4.76)
  - LUR: 2.46 (1.55 to 3.72)
  - RIS: 2.09 (1.54 to 2.78)
  - PAL: 1.81 (1.17 to 2.69)
  - ASE: 1.66 (0.85 to 2.93)
  - ZIP: 1.61 (1.05 to 2.37)
  - AMI: 1.60 (0.88 to 2.65)
  - ILO: 1.58 (0.55 to 3.65)
  - ARI: 1.20 (0.73to 1.85)
  - QUE: 1.01 (0.68 to 1.44)
  - OLA: 1.00 (0.73 to 1.33)
  - CLO: 0.3 (0.12 to 0.62)

3.3. ORs in Extrapyramidal Symptoms

- When the AP's OR (95% CI, lower range) does not overlap with ZIP OR (95% CI, upper range, 2.37), then AP ORs are significantly higher than ZIP OR. Mark them in red.

- When the AP OR (95% CI, upper range) does not overlap with ZIP OR (95% CI, lower range, 1.05), then AP ORs are significantly lower than ZIP OR. Mark them in light blue.

3.3. OR in Extrapyramidal Symptoms

- Leucht et al., 2013: http://www.ncbi.nlm.nih.gov/pubmed/23810019
  ORs in order (95% CI) (drug versus placebo)
  - HAL: 4.76 (3.70 to 6.04)
  - CPZ: 2.65 (1.33 to 4.76)
  - LUR: 2.46 (1.55 to 3.72)
  - RIS: 2.09 (1.54 to 2.78)
  - PAL: 1.81 (1.17 to 2.69)
  - ASE: 1.66 (0.85 to 2.93)
  - ZIP: 1.61 (1.05 to 2.37)
  - AMI: 1.60 (0.88 to 2.65)
  - ILO: 1.58 (0.55 to 3.65)
  - ARI: 1.20 (0.73to 1.85)
  - QUE: 1.01 (0.68 to 1.44)
  - OLA: 1.00 (0.73 to 1.33)
  - CLO: 0.3 (0.12 to 0.62)

3.4. The Concept of OR

(Skip this section if your brain is saturated with statistics)

3.4. The Concept of OR

- OR is a widely used statistical concept.
- All kinds of psychiatric articles can use it. You should try to better understand it.

3.4. The Concept of OR

How do you define an OR?

3.4. The Concept of OR

- You go to PubMed: http://www.ncbi.nlm.nih.gov/mesh/?term=odd+ratio
  - The ratio of two odds.
  - The exposure-odds ratio for case control data is the ratio of the odds in favor of exposure among cases to the odds in favor of exposure among noncases.
  - The disease-odds ratio for a cohort or cross-section is the ratio of the odds in favor of disease among the exposed to the odds in favor of disease among the unexposed.
  - The prevalence-odds ratio refers to an odds ratio derived cross-sectionally from studies of prevalent cases.

3.4. The Concept of OR

- You conclude PubMed is a not good place to understand statistical concepts.
- Dr. de Leon will start all over.

3.4. The Concept of OR

- We have a dichotomous/binary variable we want to study.
  It is a variable that can be:
  - present, or
  - absent.
  The study variable is called the dependent variable by statisticians.
  Let's imagine it is bipolar disorder.

3.4. The Concept of OR

- Then we have another dichotomous/binary variable; we want to find out if it is statistically associated with bipolar disorder. This other variable is called the independent variable by statisticians. Let's imagine it is smoking.

3.4. The Concept of OR

- We have a bipolar sample with 100 individuals:
  - 45 (45%) are current smokers, &
  - 55 (55%) are not current smokers.

  These are called cases.
- We have 100 controls from the general population:
  - 25 (25%) are current smokers, &
  - 75 (75%) are not current smokers.

  These are called controls.

This is a case-control study.

3.4. The Concept of OR

You want to find out if bipolar disorder is significantly associated with current smoking.

3.4. The Concept of OR

- You do a cross-tabulation:

		Bipolar		Total
		Yes	No	
Smoking	Yes	45	25	70
	No	55	75	130
		100	100	200

- If you remember from statistics class, a Pearson chi-square test gives you the significance.

3.4. The Concept of OR

- Let's imagine that you can calculate it using a statistical program.
  The results are:
  - Pearson chi-square=8.971
  - degrees of freedom=1
  - significance (two-sided): p=0.003

In summary, bipolar disorder is significantly associated with current smoking in this case-control study.

3.4. The Concept of OR

How do you find the effect size of this significant association between bipolar disorder and smoking?

3.4. The Concept of OR

- The OR gives you the effect size of this association.

- OR >1 indicates a positive association (↑ risk)
  <1 indicates a negative association (↓ risk)

- To orient you:
  - OR= 2 or (0.50) indicates a reasonable association.
  - OR= 5 or (0.20) indicates a strong association.

3.4. The Concept of OR

- Let's imagine that you can calculate the OR using a statistical program. The results are:
  - odds ratio=2.455
  - lower limit of 95% CI=1.347
  - upper limit of 95% CI=4.473

  The 95% CI does not include 1, but we already know by the Pearson chi-square that the difference was statistically significant.

3.4. The Concept of OR

Do you have any problems with these results?

3.4. The Concept of OR

- This significant association between bipolar disorder with current smoking:
  - p value=0.003
  - OR=2.5 (95% CI 1.3 to 3.5)

  can be explained by another factor that we are not measuring. As a matter of fact, in many countries smoking is associated with being a male.

3.4. The Concept of OR

- We are going to
  - explore the effect of male status, and
  - correct for the effect of male status.

3. 4. OR Bipolar and Current Smoking

3.4.1. Exploring the Effect of Male Status

3.4.2. Correcting for the Effect of Male Status

3.4.1. Exploring the Effect of Male Status

3.4.1. Exploring for the Effect of Male Status

- You do a cross-tabulation:

	Male		Total
	Yes	No	
Smoking Yes	45	25	70
No	25	110	130
	70	135	200

- The statistical program gives you:
  - a p value<0.001
  - OR=9.9 (95% CI 5.0 to 19.5)

Current smoking is significantly associated with male status.

3.4.1. Exploring for the Effect of Male Status

- Current smoking is significantly associated with male status by ↑ the risk:
  - OR=9.9 (95% CI 5.0 to 19.5)
- Current smoking is also significantly associated with female status by ↓ the risk:

  You can estimate female OR and 95% CI by calculating the inverse
  - female OR=1/9.9=0.10

    female upper lower 95% CI=1/5.0=0.20

    female lower 95% CI=1/19.5=0.05

    OR=0.10 (95% CI 0.05 to 0.20)

3.4.1. Exploring for the Effect of Male Status

How can you further explore the effect of being male?

3.4.1. Exploring for the Effect of Male Status

- You can explore the association between bipolar disorder and current smoking:
  - in males, and
  - females.

  Statisticians call this stratification.

3.4.1. Exploring for the Effect of Male Status

- You do a cross-tabulation in males:

Males		Bipolar		Total
		Yes	No	
Smoking	Yes	25	20	45
	No	10	10	20
		35	30	65

- The statistical program gives you:
  - a p value=0.68
  - OR=1.3 (95% CI 0.44 to 3.6)

Current smoking is not significantly associated with bipolar disorder.

3.4.1. Exploring for the Effect of Male Status

- You do a cross-tabulation in females:

Females		Bipolar		Total
		Yes	No	
Smoking	Yes	20	65	85
	No	45	5	50
		65	70	135

- The statistical program gives you:
  - a p value<0.001
  - OR=5.8 (95% CI 2.0 to 16.5)

Current smoking is significantly associated with bipolar disorder.

3.4.1. Exploring for the Effect of Male Status

- The association between current smoking and bipolar disorder is:
  - significant in females, but
  - not significant in males.

In summary, this exploration was confusing.

3.4.2. Correcting for the Effect of Male Status

3.4.2. Correcting for the Effect of Male Status

- Logistic regression can be used to correct/adjust the ORs of other variables. It is a multivariate analysis (multiple independent variables).
- In the case-control study:
  - OR bipolar disorder & current smoking =2.5 (95% CI 1.3 to 3.5),
  - OR being male & current smoking =9.9 (95% CI 5.0 to 19.5).

These are usually called univariate ORs (only 1 independent variable).

3.4.2. Correcting for the Effect of Male Status

- Logistic regression:
  - OR bipolar disorder & current smoking corrected by the effect of male status: =2.9 (95% CI 1.4 to 5.8), and
  - OR male status & current smoking corrected by the effect of bipolar disorder: =10.7 (95% CI 5.2 to 21.2).

References for Those "Thirsty" for More Statistics

1) This free article in PubMed appears a good refernce for number needed to treat http://www.ncbi.nlm.nih.gov/pubmed/10356018
2) Andrade C at J Clin Psychiatry provides summaries of statistical concepts for clinicians: a) confidence intervals http://www.ncbi.nlm.nih.gov/pubmed/25742216 b) number needed to treat http://www.ncbi.nlm.nih.gov/pubmed/25830454 and c) odds ratios and other related terms http://www.ncbi.nlm.nih.gov/pubmed/26231012
3) Citrome L is a clinician who has many articles on statistical concepts. The presentation provides articles on APs a) Medscape http://www.medscape.org/viewarticle/748336 b) free on PubMed http://www.ncbi.nlm.nih.gov/pubmed/20428308 and) other on PubMed http://www.ncbi.nlm.nih.gov/pubmed/18479317 Other article introduced how NNT and NNH can be combined http://www.ncbi.nlm.nih.gov/pubmed/23574101

Questions

-Please review the 10 questions on the pdf titled "Questions on the Presentation: Introduction to Statistical Concepts Needed for Clinical Pharmacology".

-You will find the answers on the slide after the "Thank you" slide. No peeking until you have answered all the questions.

-If you do not answer all the questions correctly, please review the Power Point presentation one more time to reinforce the statistical concepts.

Thank you

Dr. de Leon is grateful that you survived
so that you can read
his other presentations focused on pharmacology.

Answers

1. A
2. A
3. C
4. B
5. A
6. A
7. A
8. D
9. A
10. B