MEASURES OF EFFECT SIZE FOR RESEARCH AND PRACTICE

Chittaranjan Andrade, MD
Professor in Psychopharmacology
National Institute of Mental Health and Neurosciences
Bangalore, India

WORKSHOP OBJECTIVES

- To provide the audience with an understanding of:
  - Standardized mean difference (SMD)
  - Relative risk (RR)
  - Odds ratio (OR)
  - Numbers needed to treat/harm (NNT/H)
  - Confidence intervals
- To teach the audience how to calculate
  - SMD, RR, OR, NNT/H
- To teach the audience how to teach

AUDIENCE REQUIREMENTS

- Assumptions:
  - That the audience understands basic concepts in research
    [E.g. What RCTs and observational studies are]
  - That the audience understands basic concepts in statistics
    [E.g. What M(SD), normal distribution, statistical significance testing etc. are]

WHY THESE ARE NECESSARY

- Knowledge about SMD, RR, OR, NNT/H, CI is important when
  - Reading research
  - Writing papers
  - Applying the findings of research to clinical practice
- It enriches our understanding of the findings
  - Above and beyond usual information such as M(SD), difference between means, difference between proportions, absolute risk, and statistical significance

FOR EXAMPLE: 1

- The M(SD) age of onset of schizophrenia is 25(4) years
  - How sure are we that this number is right?
- Ziprasidone resulted in greater improvement than placebo by 8 points on the PANSS.
  - Is this a big or a small advantage?
- The response rate with venlafaxine vs placebo is 60% vs 40%.
  - Is this a big or a small advantage?

FOR EXAMPLE: 2

- Using measures of effect size, readers can
  - Get a better perspective about the findings
  - Better distinguish statistical significance from clinical significance (even a small relationship can be statistically significant if N is very large).

EFFECT SIZE

- Standardized mean difference

EFFECT SIZE (ES)

- Correct term: standardized mean difference (SMD)
  - Because RRs, ORs etc are also measures of effect size.
- Difference between groups expressed in units of standard deviation (SD).
- Cohen's d: Uses pooled SD
- Glass' Delta: Uses SD of the control group (because the intervention may change not only the mean but also the SD)

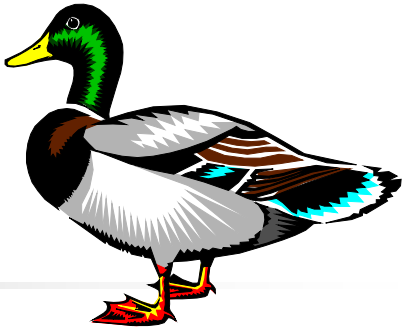
SMD

- Difference between the means of the experimental and control groups divided by the pooled standard deviation (SD)
  - Or the SD of the control group, if the pooled SD is unavailable
- Tells us how different the means are in units of SD.
- Traditional interpretation of effect sizes:
  - 0.2-0.5 = small
  - 0.5-0.8 = moderate
  - >0.8 = large

SMD: EXAMPLE

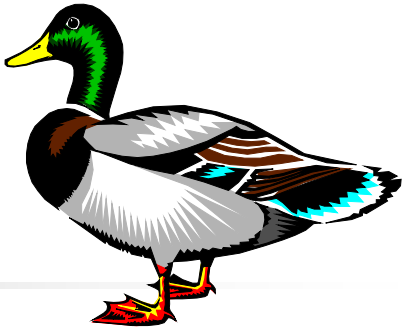
- M (SD) final HRSD score with venlafaxine was 12 (3).
- M (SD) final HRSD score with placebo was 16 (4).
- Effect size = (16-12)/4 = 1.
- i.e., treatment with venlafaxine lowered the endpoint HRSD score by an average of 1 SD.
- What does SMD=0.5 mean?

INTERPRETING EFFECT SIZE



- ES=0.2; 85% overlap in the distributions of the two groups
- ES=0.5, 67% overlap
- ES=0.8, 53% overlap

INTERPRETING EFFECT SIZE



- An ES of 1.0 means that about 84% of the venlafaxine group has final HRSD scores <16 (the mean final HRSD score of the control group).
- For an ES of 2.0, this figure is 98%.
- Important: Effect size is a RELATIVE ESTIMATE. Examine its clinical importance in the context of the absolute difference between groups.
  - Because ES depends on the width of the SD.
  - Examples

SMD: APPLICATIONS

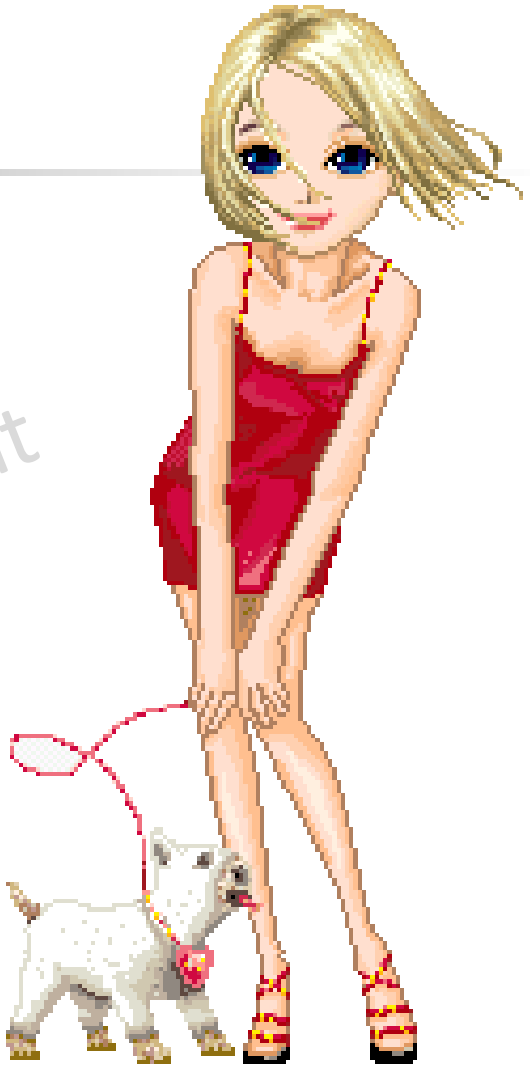
- Tells us how big an effect is.
  - E.g. Risperidone reduces scores on a QoL scale by 3 points.
  - Is this a small improvement or a big improvement?
  - When we are familiar with a scale (e.g. PANSS), we can interpret absolute changes easily).
  - When we are unfamiliar with a scale (e.g. this QoL scale), the SMD can help.
- Can be used to convert data from different rating instruments into a common unit for meta-analysis.

META-ANALYSIS AND EFFECT SIZES

- Average effect size can be calculated across studies.
- Useful when different studies use different rating scales.
- Effect sizes can be weighted for quality of studies.
- Regression analysis can be done to find out what design variables predict effect size.

RISK

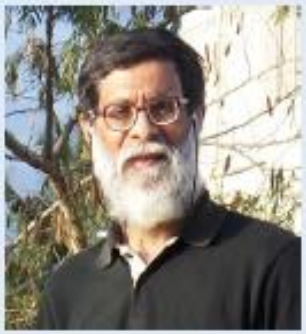
- Absolute risk
- Relative risk
- Attributable risk



Clinical and Practical **Psychopharmacology**

Understanding Relative Risk, Odds Ratio, and Related Terms:
As Simple as It Can Get

Chittaranjan Andrade, MD



Each month in his online column, Dr Andrade considers theoretical and practical ideas in clinical psychopharmacology with a view to update the knowledge and skills of medical practitioners who treat patients with psychiatric conditions.

Department of Psychopharmacology, National Institute of Mental Health and Neurosciences, Bangalore, India (candrade@psychiatrist.com).

ABSTRACT

Risk, and related measures of effect size (for categorical outcomes) such as relative risks and odds ratios, are frequently presented in research articles. Not all readers know how these statistics are derived and interpreted, nor are all readers aware of their strengths and limitations. This article examines several measures, including

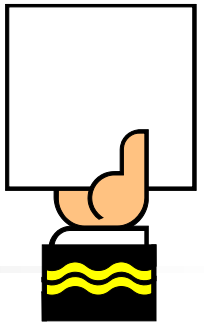
Introduction

Many research papers present findings as odds ratios (ORs) and relative risks (RRs) as measures of effect size for categorical outcomes. Whereas these and related terms have been well explained in many articles,[1–5] this article presents a version, with examples, that is meant to be both simple and practical. Readers may note that the explanations and examples provided apply mostly to randomized controlled trials (RCTs), cohort studies, and case-control studies. Nevertheless, similar principles operate when these concepts are applied in epidemiologic research. Whereas the terms may be applied slightly differently in different explanatory texts, the general principles are the same.

Clinical Situation

Consider a hypothetical RCT in which 76 depressed patients were randomly assigned to receive either venlafaxine (n = 40) or placebo (n = 36) for 8 weeks. During the trial, new-onset sexual dysfunction was identified in 8 patients treated with venlafaxine and in 3 patients treated with placebo. These results are presented in Table 1. Using these data, we can calculate the values for a variety of terms, as illustrated in the sections that follow.

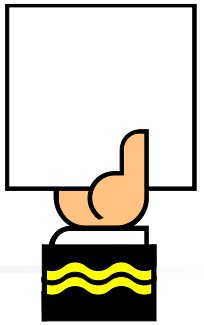
RISK: DEFINITIONS



- Risk
  - Used in the context of benefit as well as AEs
  - Think of it as ‘chance’
  - E.g. When tossing a coin, the risk of heads is 0.5

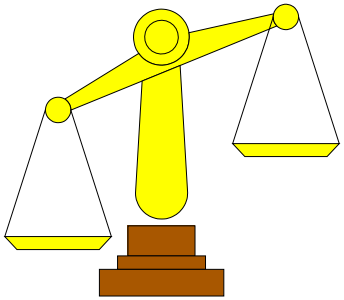
- Absolute risk: the probability of occurrence of an event.
  - E.g., the lifetime risk of bipolar disorder is 0.5-1.0%.

RISK: DEFINITIONS



- Relative risk (RR): the probability of occurrence of an event in one group relative to the probability of occurrence of the event in another group.
  - E.g. the RR of unipolar depression in women is 1.5 relative to men.
  - E.g. Boys are 5 times as likely to develop ADHD as girls (RR=5).
  - Calculated from RCTs, cohort studies

REFERENCE GROUPS FOR RRs



- A placebo-treated group
  - For anorgasmia with fluoxetine
  - For response to fluoxetine
- A control group
  - For schizophrenia in persons with and without a positive family history
- The lowest quartile or quintile
  - For risk of IHD in persons stratified by cholesterol levels

WORKED EXAMPLE

- In an RCT, 11 of 50 patients developed nausea with venlafaxine.
  - So, the absolute risk of nausea with venlafaxine is 22%.
- Only 4 of 40 patients developed nausea with placebo.
  - So, the absolute risk of nausea with placebo is 10%.
- The RR of nausea with venlafaxine relative to placebo is 22/10 = 2.2.

RR=2.2

- Explained in English, this means
  - Nausea is twice as common with venlafaxine as with placebo
  - Nausea is 2.2 times as common with venlafaxine as with placebo
  - Nausea is slightly more than twice as common with venlafaxine as with placebo
  - The risk of nausea with venlafaxine is 2.2 times the risk of nausea with placebo.

INTERPRETING THE RR

- An RR = 1 means that the risk is identical.
- An RR < 1 means that the risk is reduced.
- An RR > 1 means that the risk is increased.
- Note: The RR cannot be negative.

- During 12 months of maintenance therapy, venlafaxine was associated with a lower risk of relapse into depression than placebo (RR=0.3).
  - Explain this in a single sentence, in English.

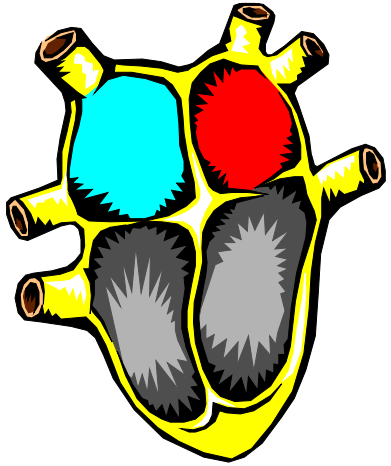
RR=0.3

- The 1-year relapse rate with venlafaxine was about one-third that with placebo.
- Venlfaxine reduced the relapse rate by 70%.

HOW 'BIG' IS THE RR?

- In general, RRs <0.5 and RRs > 2.0 are considered clinically significant
  - Parallel with interpretation of Cohen's d
  - However, look at the absolute/base risk (the risk in the control group or in the general population)
  - Also look at the importance of the outcome (e.g. a minor AE vs a life-threatening AE.

INTERPRETING RRs: Importance of the base rate



- Doubling the risk may be unimportant if the base rate is low.
  - E.g. Lithium and Ebstein's anomaly
- A small increase in risk may be important if the base rate is high.
  - E.g. Sexual impairment with venlafaxine in depression

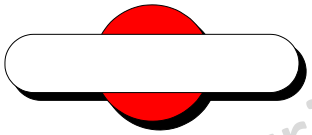
INTERPRETING RRs: Emotional issues: 1

- For the occurrence of an adverse effect with drug relative to placebo, which of the following is the worst?
- RR=2
- Doubled risk
- Risk increased by 100%
- A 200% risk

INTERPRETING RRs: Emotional issues: 2

- A drug lowers the 5-year risk of myocardial infarction by 50%. Impressive?
  - But the absolute risk is only 1%. Still impressive?
  - The NNT is 1 in 200. That is 200 patients must receive the drug for 5 years for myocardial infarction to be prevented in 1 person. Still impressive?

ATTRIBUTABLE RISK (Risk Difference)



- Difference between two absolute risks, as in the attributable risk of nausea with venlafaxine = (risk with venlafaxine – risk with placebo).
- NNT is the reciprocal of the attributable risk.

ODDS RATIO

- When tossing a coin, the odds of the coin falling heads are 50:50 (1:1)
- How this is different from risk or probability (1:2)

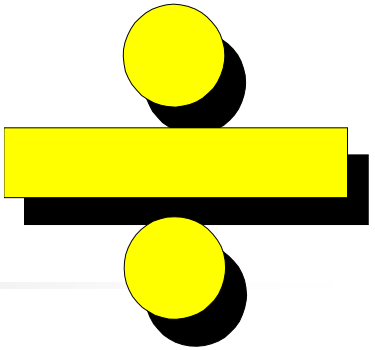
WHY ODDS RATIO? WHY NOT RR?

- In an RCT, we know
  - How many patients received venlafaxine and how many developed nausea.
  - Ditto for placebo
- In a case-control study
  - We know how many cases of GI bleeding received SSRIs and how many did not
  - Ditto for age- and sex-matched controls
  - We DON'T know how many SSRI-treated patients did and did not suffer bleeds; ditto for untreated patients.

ODDS RATIO: WORKED EXAMPLE

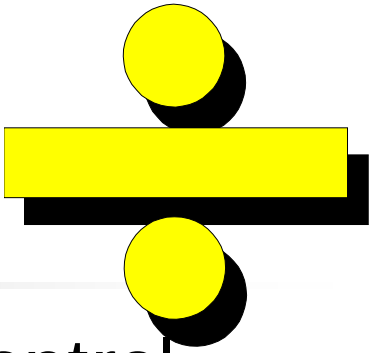
- In a case control study, 20 of 200 cases of GI bleeding were receiving SSRIs. Therefore, the odds of SSRI treatment in cases = 20:180.
- Only 5 of 200 controls without GI bleeding were receiving SSRIs. Therefore, the odds of SSRI treatment in controls = 5:195.
- Therefore, the odds ratio for GI bleeding with SSRIs is 20:180/5:195; i.e., 4.33.
- RR is inapplicable here because we don't know how many SSRI patients didn't suffer bleeds.

ODDS RATIOS



- RRs are calculated from RCTs.
- Odds ratios are calculated from case-control studies as we don't know the population at risk.
- RRs and ORs are presented and interpreted similarly.
- RRs are more easy to conceptualize than ORs.
- ORs, like RRs, cannot be negative.
- ORs similar to RRs when the event is rare but underestimate or overestimate RRs, otherwise.

ODDS RATIOS



- Odds ratios are calculated from case-control studies as we don't know the population at risk.
- ORs can be derived from logistic regressions which control for confounds.
  - The value of the OR is not fixed, as is the value of the RR in an RCT.
  - The value of the OR increases with an increasing number of independent variables entered in the logistic regression equation.

PETO ODDS RATIO

- Used only in the context of meta-analysis
  - Employs the inverse variance method
- Useful only when:
  - The event of interest is very rare (e.g. suicide in antidepressant RCTs)
  - Experimental and control groups are reasonably balanced in size

NNT/NNH

Clinical and Practical **Psychopharmacology**

The Numbers Needed to Treat and Harm (NNT, NNH) Statistics: What They Tell Us and What They Do Not

Chittaranjan Andrade, MD

Each month in his online column, Dr Andrade considers theoretical and practical ideas in clinical psychopharmacology with a view to update the knowledge and skills of medical practitioners who treat patients with psychiatric conditions.

Department of Psychopharmacology, National Institute of Mental Health and Neurosciences, Bangalore, India (candrade@psychiatrist.com).

Clinical Question

A meta-analysis of randomized controlled trials (RCTs) found that antidepressants were effective in pediatric depression; the number needed to treat (NNT) was 9.[1] What information does NNT = 9 provide?

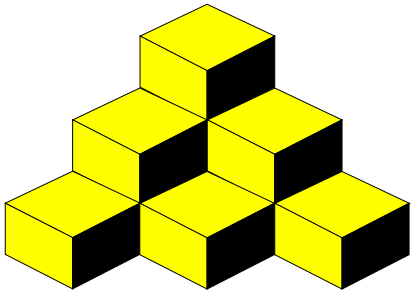
Introduction

Research results are presented in the form of summary statistics such as the mean improvement or the mean response rate in different treatment groups. These summary statistics can be directly compared, such as to determine whether the mean improvement or the mean response rate is significantly greater in one group versus the other.

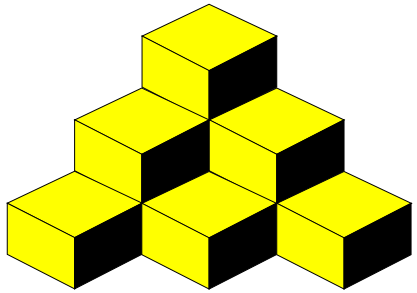
Readers may wish to know whether an identified advantage is small or large. For example, in an RCT, an antidepressant may outperform

NUMBERS NEEDED TO TREAT (a measure of ES for proportions)

- Response to drug = 60%.
- Response to placebo = 40%.
- Hence, out of 100 treated persons, 20 extra patients respond to drug.
- Hence, out of 5 treated persons, 1 extra patient responds to drug.
- NNT=5.
- Ditto for NNH (for AEs)

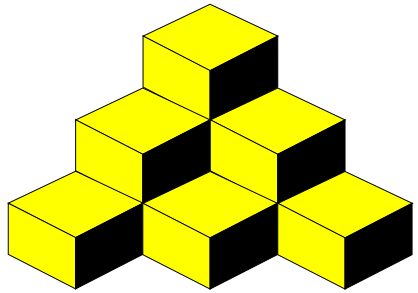


NNT & NNH



- For any given drug (relative to control):
  - A smaller NNT implies greater efficacy
  - A larger NNH implies greater safety

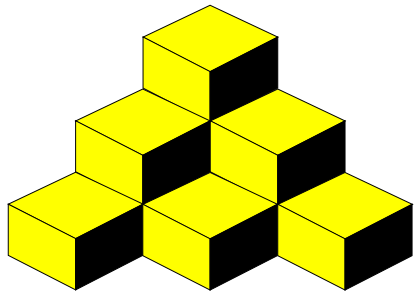
NNT: IMPORTANCE: 1



- Response to drug = 60%.
- Response to placebo = 40%.
- Please calculate the RR and the NNT

- Response to drug = 6%
- Response to placebo =4%
- Please calculate the RR and the NNT

NNT: IMPORTANCE: 2



- Response to drug = 60%.
- Response to placebo = 40%.
- RR = 1.5
- NNT = 5
- Response to drug = 6%
- Response to placebo =4%
- RR = 1.5
- NNT = 50

NNT: Questions, 1

- Define NNT.
- In a depression study, provide examples for the exposure (treatment) variable.
- What is the reference group when calculating NNT?
- In a depression study, what might be the outcome variable for calculating NNT?
- In what kind research design is NNT calculated?

NNT: Answers, 1

- The number of persons who need to be exposed to [whatever] for one extra person to benefit.
  - 'Exposure': experimental treatment, e.g. drug, CBT, etc.
  - 'Benefit' is dichotomized (e.g. response vs nonresponse; alive vs dead)
  - Reference group: Persons exposed to the control treatment, e.g. placebo, treatment as usual, etc.
  - NNT is usually calculated from RCT data (but can be obtained from cohort studies, too).

NNT: Questions, 2

- What is the ideal value for NNT?
- What is the maximum value for NNT?
- Can the NNT be 0?
- Can the NNT be negative?

NNT: Answers, 2

- The ideal value for NNT is 1
  - Everybody responds to drug, nobody responds to placebo (NNT=100/100).
  - [For every patient treated with drug, one extra patient improves.]
- NNT can run into thousands for a weak effect.
  - 95% upper bound CI can be infinity (not significant).
  - NNT is infinity if response rate is the same in experimental and control groups [for every patient treated with drug, zero extra patients improve].

NNT: Answers, 2 (contd.)

- NNT cannot lie between 0 and 1
  - E.g. You cannot treat zero (or half) patients for one extra patient to respond. The minimum value must be 1.
- NNT can be negative if response rate is greater in the placebo group.

NNT: OTHER NOTES

- NNT is time-sensitive
  - 5 patients need to receive venlafaxine for 8 weeks for 1 extra patient to respond.

NNH (Number needed to harm)

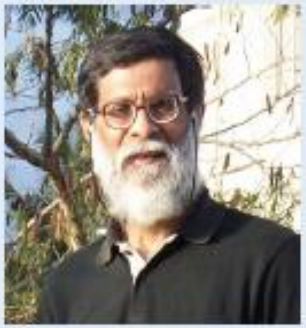
- Concept is the same as for NNT.
- Outcome is a harm variable
  - E.g. All cause discontinuation
  - E.g. Discontinuation due to AEs
  - E.g. Risk of a specific AE
- Variable is dichotomized (e.g. AE present/absent)

LHH

Clinical and Practical Psychopharmacology

Likelihood of Being Helped or Harmed as a Measure of Clinical Outcomes in Psychopharmacology

Chittaranjan Andrade, MD



Each month in his online column, Dr Andrade considers theoretical and practical ideas in clinical psychopharmacology with a view to update the knowledge and skills of medical practitioners who treat patients with psychiatric conditions.

Department of Psychopharmacology, National Institute of Mental Health and Neurosciences, Bangalore, India (candrade@psychiatrist.com).

ABSTRACT

The likelihood of being helped or harmed (LHH) ratio is an indirect measure of effect size. It tells the reader how much as likely a patient is to benefit from a treatment as to suffer from an adverse outcome with that treatment; larger values for LHH indicate more favorable treatment outcomes. The numerator for LHH is usually a measure of

Introduction

In clinical psychopharmacology, useful measures of effect size include statistics such as standardized mean deviation, relative risk, odds ratio, number needed to treat (NNT), and number needed to harm (NNH). The likelihood of being helped or harmed (LHH) is one among these statistics that is less well known and therefore less used (or perhaps it is less used and therefore less well known). Previous articles in this column addressed certain of these measures of effect size and related subjects.[1–4] The present article considers the LHH.

Likelihood of Being Helped or Harmed: Concept

Conceptually, the LHH is the ratio of the probability of benefit to the probability of harm. In the context of a randomized controlled trial (RCT) that compares active drug with placebo, the probability of benefit is operationalized as the risk difference (between drug and placebo) for a favorable outcome, and the probability of harm is operationalized as the risk difference (between drug and placebo) for an unfavorable outcome. The favorable outcome is usually treatment response, and the unfavorable outcome is usually treatment discontinuation.

In the context referred to above, if the LHH is greater than 1, the patient

LIKELIHOOD OF BEING HELPED OR HARMED (LHH)

- Concept:
  - Ratio of the probability of benefit to the probability of harm
- Operationalization (e.g. for a Drug vs Placebo RCT)
  - (Risk difference for a favorable outcome) divided by (risk difference for an unfavorable outcome)
- Favorable outcomes:
  - E.g. treatment response or treatment remission
- Unfavorable outcomes:
  - E.g. drop out due to AEs or all cause drop out

LHH

- If LHH >1, probability of benefit is greater than probability of harm.
  - If LHH <1, probability of harm is greater.
- Alternate method of calculation:
  - LHH = NNH/NNT

LHH: WORKED EXAMPLE: 1

- Response to venlafaxine: 28/40, or 70.0%
- Response to placebo: 15/36, or 41.7%
- Risk difference for benefit: 70.0-41.7 = 28.3%
- NNT = 100/28.3, or 3.5
- Drop out with venlafaxine: 14/40, or 35.0%
- Drop out with placebo: 8/36, or 22.2%
- Risk difference for harm: 35.0-22.2 = 12.8%
- NNH = 100/12.8, or 7.8

LHH: WORKED EXAMPLE: 2

- LHH = 28.3/12.8 = 2.2
- LHH = 7.8/3.5 = 2.2
- Interpretation:
  - For every 2 (extra) patients who respond to venlafaxine, 1 (extra) patient will drop out
  - “Extra” because the effect of placebo is subtracted
- Note: LHH is sensitive to drug, dose, duration of treatment, etc. and should be viewed in this context.

LHH: USES AND LIMITATIONS

- Can guide patients, clinicians about risk-benefit trade-off
- Can be compared across studies for different drugs/doses provided that the studies are similar in nature (e.g. sample characteristics, study duration etc.)
- No information about absolute rates of benefit or harm.

CONFIDENCE INTERVALS

Clinical and Practical **Psychopharmacology**

A Primer on Confidence Intervals in Psychopharmacology

Chittaranjan Andrade, MD

Each month in his online column, Dr Andrade considers theoretical and practical ideas in clinical psychopharmacology with a view to update the knowledge and skills of medical practitioners who treat patients with psychiatric conditions.

Department of Psychopharmacology, National Institute of Mental Health and Neurosciences, Bangalore, India (candrade@psychiatrist.com).

ABSTRACT

Research papers and research summaries

Clinical Question

You are reading a meta-analysis on antipsychotic augmentation with modafinil or armodafinil for the treatment of negative symptoms of schizophrenia.[1] You observe that, on the Positive and Negative Syndrome Scale, negative subscale (PANSS-N), augmentation with either of these drugs was superior to augmentation with placebo by a mean of 0.27 points. Which of the following statements is more informative?

1. Armodafinil was superior to placebo on the PANSS-N; the mean difference was 0.27 points ($P = .02$).

2. Armodafinil was superior to placebo on the PANSS-N; the mean difference was 0.27 (95% confidence interval [CI], 0.04–0.50) points.

Introduction

Consider a randomized controlled trial (RCT) in which an antidepressant drug elicits a 63% response rate in patients with major

CONFIDENCE INTERVALS (C.I.)

- Descriptive statistics are approximations based on our sample.
- We wish to know what the population value is.
- 95% CI are the values somewhere between which we are 95% certain that the population value lies.
  - In 2.5% of cases, each, the population value will lie below or above the interval.
  - This is much more informative than a P value for difference between means!

CONFIDENCE INTERVALS

- 95% CI are not the same as SD. They are derived, for e.g., from the SD and sample size.
  - The larger the sample size, the narrower (i.e., more precise) the CI.
  - Sample size will need to be quadrupled for the CI to be narrowed to half (Altman and Bland, BMJ 2014)
  - [But larger sample size will not change the SD unless there was grievous sampling error in the smaller sample (Altman and Bland, BMJ 2005)]

Narrow vs wide CI

- Risk: Chance of cure with a new drug:
  - Chance of cure in the study, 50% (95% CI, 3% to 98%)
- Risk: Risk of teratogenicity with valproate:
  - Risk observed in the study, 6% (95% CI, 3% to 11%)
- Mean: Age of onset of schizophrenia:
  - Mean, 25.1 (95% CI, 16.8-33.4)
- Wide CI can indicate many things:
  - Small sample size, and hence inability to be accurate
  - A subgroup effect (E.g. Gender & age of psychosis onset)
  - Applies to all statistical parameters (mean, RR, NNT etc.)

Overlapping and non-overlapping CI

- Mean: Age of onset of schizophrenia
  - In men : Mean, 23.3 (95% CI, 20.0-26.6)
  - In women : Mean, 27.5 (95% CI, 25.5-29.5)

  - In men : Mean, 23.3 (95% CI, 22.0-24.6)
  - In women : Mean, 27.5 (95% CI, 25.5-29.5)

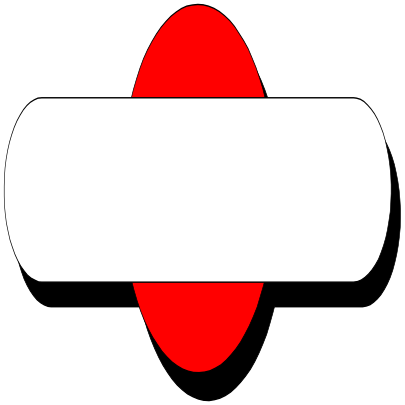
Non-overlapping CI

- If the 95% CI do not overlap, the groups differ at P<0.01 level.
- If the 95% CI overlap by less than about half of the CI, the groups differ at the P<0.05 level.
- If CI overlap to a large extent (e.g. >50%), the difference is unlikely to be significant.
- Applies to all statistical parameters (means, proportions)

CI: Examining significance

- Proportion: 65% (52% to 78%)
- Difference between means: 4.4 (3.1-5.7)
- Difference between means: (-4.1 to 9.6)
  - (-0.1 to 9.6)
- Difference between proportions: 6% (3% to 9%)
- Difference between proportions: (-8% to 13%)
  - (-1% to 18%)
- Red ranges are not significant at P<0.05 level.
  - But look at where the bulk of the values lie.
  - Applies to all statistical parameters

CI: Examining significance with RRs



- RR of 1.0 = identical risk.
- If 95% confidence intervals include 1.0, the risk is not significantly different from the control group.
- E.g. RR=0.7; 95% CI=0.5-0.9
- E.g. RR=0.7; 95% CI=0.3-1.1
- E.g. RR=1.6, 95% CI=0.9-2.3
- E.g. RR=1.6, 95% CI=1.1-2.3

CI: The importance of the boundaries

- SSRIs were associated with a 3.1% risk of major teratogenic malformations (95% CI, 2.0-4.5).

- Note the upper bound value of the 95% CI.
  - This study suggests that we can be 95% certain that the teratogenic risk with SSRIs is 4.5% or less.
  - Applies to all statistical parameters.

CI: Terms

- The range of values in the CI comprise the confidence interval.
  - 95% CI, 5.0-11.5; range=6.5
- The lower and upper values are called the confidence limits.
  - 5.0: lower bound/limit; 11.5: upper bound/limit
- Do not confuse CI with confidence level, which is 1-alpha
  - 1-0.05 = 0.95, or 95%

CI: OTHER NOTES

- The confidence limits are not necessarily symmetrically distributed around the estimate.
  - E.g. NNH=14 (95% CI, 7-45)

- 95% CI are usual
  - But 90%, 99%, or other confidence intervals can also be computed.
  - 99% CI are wider than 95% CI

PARTING NOTES

- Always interpret results using BOTH absolute and relative estimates
  - E.g. Absolute difference between groups and SMD
  - E.g. Absolute risk difference and RR

  - Examples

FOR E-MAIL UPDATES ON RESEARCH IN PSYCHIATRY

- Synergytimes
- Write to andradec@gmail.com

ENFIN...

- That’s it, folks;
  thanks for listening!