When Does a Difference Make a Difference? Everything You Always Wanted to Know about Effect Sizes* (*But Were Afraid to Ask)

Leslie Citrome, MD, MPH
Clinical Professor of Psychiatry and Behavioral Sciences
New York Medical College
Valhalla, New York
Faculty Disclosure

Disclosure

• The faculty have been informed of their responsibility to disclose to the audience if they will be discussing off-label or investigational use(s) of drugs, products, and/or devices (any use not approved by the US Food and Drug Administration).

• Applicable CME staff have no relationships to disclose relating to the subject matter of this activity.

• This activity has been independently reviewed for balance.
Why is This Presentation Being Done?

• Clinicians often assume that if a $P$-value < .05 then the result must be important
• Moreover, there is often the false belief that a $P$-value < .0001 denotes an even more impressive difference
• There is a general lack of understanding of effect size and how this can be used to appraise clinical trial results
• Simple to calculate measures of effect size such as number needed to treat (NNT) and number needed to harm (NNH) can help treaters place study outcomes in a more relevant clinical context
Learning Objectives

• Differentiate statistical significance and clinical significance
  – Recognize that statistical superiority of one intervention vs another does not answer the question if this superiority is clinically relevant

• Recognize the most commonly used measures of effect size
  – Understand the importance of effect size to illustrate and quantify clinical relevance of study results

• Calculate Number Needed to Treat (NNT), Number Needed to Harm (NNH), and Likelihood to be Helped or Harmed (LHH)
Outline

• What is Evidence-Based Medicine (EBM)?
• What is an effect size and how is it different from a $P$-value?
• What are Number Needed to Treat (NNT) and Number Needed to Harm (NNH)?
• Using NNT and NNH to evaluate treatment choices: The example of tardive dyskinesia
• What is Likelihood to be Helped or Harmed (LHH)?
• Using LHH to assess benefit vs risk: Returning to the example of medications approved for the treatment of tardive dyskinesia
• One more example: Pimavanserin for Parkinson’s disease psychosis
• Summary
Outline

• What is Evidence-Based Medicine (EBM)?
• What is an effect size and how is it different from a \( P \)-value?
• What are Number Needed to Treat (NNT) and Number Needed to Harm (NNH)?
• Using NNT and NNH to evaluate treatment choices: The example of tardive dyskinesia
• What is Likelihood to be Helped or Harmed (LHH)?
• Using LHH to assess benefit vs risk: Returning to the example of medications approved for the treatment of tardive dyskinesia
• One more example: Pimavanserin for Parkinson’s disease psychosis
• Summary
What is Evidence-Based Medicine?

Clinical Judgment

Relevant Scientific Evidence

Patients’ Values and Preferences

EBM

EBM: Core Features

• EBM is about process
• EBM is a philosophy
• EBM is a set of tools
• EBM is 5 steps
  (1) Formulate the question
  (2) Search for answers
  (3) Appraise the evidence
  (4) Apply the results
  (5) Assess the outcome

EBM is NOT “cookbook medicine”


Evidence Changes over Time!
Getting “Out-of-Date” Can Result in:

- Under-use of effective interventions
- Over-use of unproven interventions
- Unnecessary variations in practice
- Eminence-based vs evidence-based practice
- Reliance on LPIT (Last Patient I Treated)
Need to Learn a Process to Evaluate the Evidence That is Presented in:

- Journal articles
- CME offered by professional organizations
- Industry-sponsored lectures
- Practice guidelines
The Philosophy of EBM to the Rescue!

“Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.”

“…the integration of best research evidence with clinical expertise and patient values.”

EBM: 5 Steps to Success

Step 1: Formulate the question
What kind of patient or problem?
What intervention, treatment, diagnostic test, risk factor, or prognostic factor are you interested in?
What comparisons are you making (treatment A versus treatment B, treatment versus no treatment, etc.)?

Step 2: Search for answers
Does it work?
Has a systematic review been conducted (search Medline or the Cochrane Database)?
Are there RCTs that enrolled similar patients to yours?
If using guidelines, are they evidence-based or eminence-based?
Well formulated questions make it easier to locate an answer, if one exists.

Step 3: Appraise the evidence
Will it work in the “real world”?
Is it relevant to your question and your patient?
Is the statistically significant result clinically significant?
If effect size is not mentioned in the research report, is there sufficient information available to calculate the NNT for the categorical outcomes of interest?

Step 4: Apply the results
Is it worth it?
Is the intervention, treatment, diagnostic test, etc., important to you within the context of your clinical experience and important to the patient in terms of their preferences?

Step 5: Assess the outcome
Did you ask the right question?
Did you find answers?
Were the answers you found based on a high-quality level of evidence?
Did it make clinical sense?
Did it make a difference?
Can you quantify this?
Does the patient agree?

RCT = randomized controlled trial.
1) Formulate the Question Relevant to Areas of Interest

- Clinical findings
- Etiology
- Clinical manifestations
- Differential diagnosis
- Diagnostic tests
- Prognosis
- Therapy
- Prevention

2) Search for Answers

- Does it work? Efficacy studies (RCTs) can tell us if an intervention is better than placebo
- Will it work? Effectiveness studies are usually more generalizable
- Is it worth it? Benefits vs harms? Cost?
Use Best Available Evidence

1a: Systematic review of RCTs
1b: Individual RCT with narrow CI
2a,b: Cohort studies (review, individual)
2c: Outcomes research; epidemiologic studies
3a,b: Case-control (review, individual)
4: Case series
5: Expert opinion

Anecdotal evidence ranks very low

CI = confidence interval.
Find the Best Evidence

- Textbooks may be out of date
- Journals contain much that is irrelevant
- General databases may be cluttered with less useful sources
- EBM sources are increasingly available
  - *Evidence-Based Mental Health* journal
  - Cochrane Reviews
    - Cochrane collaboration founded in 1992 for “preparing, maintaining and promoting the accessibility of systematic reviews of the effects of health care interventions”
  - American College of Physicians (ACP) Journal Club
Online Resources: Up-to-Date and Evidence-Based
Algorithms

- Time-saving summary of pre-evaluated evidence resulting in systematic, valid approach to treatment
- Examples at Psychopharmacology Algorithm Project (www.psychopharm.mobi)

Caution: Not all algorithms are evidence-based. There are many eminence-based algorithms out there!
Secondary Resources: Practice Guidelines

Caution: Not all practice guidelines are evidence-based. There are many eminence-based practice guidelines out there!
3) Appraise the Evidence: Methods

- Concealed randomization?
- Double-blind?
- All participants accounted for and analyzed in groups?
  - 80% follow-up necessary for valid results
  - ITT analysis
- Were groups comparable?
- Aside from experimental treatment, treated equally?
- Are the results statistically and clinically significant?

ITT = intention-to-treat.
4) Apply the Results

• How applicable?
  – Is my patient like those studied?
  – Is treatment consistent with my patient’s values and preferences?
  – Is treatment feasible in my practice setting?
5) Assess the Process

• Is it working?
  – Measurement-based care will help you in assessing this
    • For example, using the PHQ-9, AIMS, etc.

PHQ-9 = Patient Health Questionnaire; AIMS = Abnormal Involuntary Movement Scale.
Outline

• What is Evidence-Based Medicine (EBM)?
• **What is an effect size and how is it different from a \( P \)-value?**
• What are Number Needed to Treat (NNT) and Number Needed to Harm (NNH)?
• Using NNT and NNH to evaluate treatment choices: The example of tardive dyskinesia
• What is Likelihood to be Helped or Harmed (LHH)?
• Using LHH to assess benefit vs risk: Returning to the example of medications approved for the treatment of tardive dyskinesia
• One more example: Pimavanserin for Parkinson’s disease psychosis
• Summary
The difference in remission for a major depressive episode at 6 weeks for Drug A vs Drug B is highly statistically significant, but clinically irrelevant.


How irrelevant is this? Can we quantify this?
EBM is about Benefit and Risk: Key Concepts

- $P$-value and statistical significance
- Effect size and clinical significance
Concepts Related to Benefit / Risk: 
*P*-Value

- This gives an indication of how strong the likelihood that any difference is NOT due to chance.
- The smaller the *P*-value, the more convinced you are that something is going on that is not just random.
- This does not state anything about the size or the importance of the non-random effect.
- *P*-value is not the same as effect size.

Concepts Related to Benefit / Risk: Effect Size

• This gives an indication of how big the treatment effect is in terms of reduction in symptoms, or other outcome of interest
• The greater the effect size, the more convinced you are that the intervention will have a clinically important impact
• This does not state anything about statistical significance of the observed outcome in a clinical trial
• Effect size is not the same as $P$-value
## There are Several Measures of Effect Size

<table>
<thead>
<tr>
<th>Effect Size</th>
<th>Range of Possible Values (weakest, ie, no difference, to strongest)</th>
<th>Typical Example of a Small Effect</th>
<th>Typical Example of a Large Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relative Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Risk</td>
<td>1 to ∞</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Odds Ratio</td>
<td>1 to ∞</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Hazard Ratio</td>
<td>1 to ∞</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Relative Risk Increase</td>
<td>1 to ∞</td>
<td>&lt; 100%</td>
<td>300%</td>
</tr>
<tr>
<td><strong>Absolute Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributable Risk</td>
<td>0 to 100%</td>
<td>&lt; 10%</td>
<td>33%–50%</td>
</tr>
<tr>
<td>NNT</td>
<td>∞ to 1</td>
<td>≥ 10</td>
<td>2–3</td>
</tr>
<tr>
<td>Cohen’s $d$</td>
<td>0 to ∞</td>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Area Under the Curve</td>
<td>0.50 to 1.00, or 0.50 to 0</td>
<td>0.56</td>
<td>0.71</td>
</tr>
<tr>
<td>Success Rate Difference</td>
<td>0 to 1</td>
<td>0.11</td>
<td>0.43</td>
</tr>
</tbody>
</table>
### P-Value vs Effect Size

<table>
<thead>
<tr>
<th>P-Value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicates Statistical Significance</td>
<td>Indicates Clinical Significance</td>
</tr>
<tr>
<td>Independent of Effect Size</td>
<td>Independent of <em>P</em>-Value</td>
</tr>
</tbody>
</table>
Outline

- What is Evidence-Based Medicine (EBM)?
- What is an effect size and how is it different from a P-value?
- What are Number Needed to Treat (NNT) and Number Needed to Harm (NNH)?
- Using NNT and NNH to evaluate treatment choices: The example of tardive dyskinesia
- What is Likelihood to be Helped or Harmed (LHH)?
- Using LHH to assess benefit vs risk: Returning to the example of medications approved for the treatment of tardive dyskinesia
- One more example: Pimavanserin for Parkinson’s disease psychosis
- Summary
I have the ability to quantify the unquantifiable.
That is why they call me Dogbert the Quantifier.
WHO CALLS YOU THAT?

EIGHT PEOPLE.
Concepts Related to Benefit / Risk: Effect Size – NNT

- NNT is one measure of effect size
- It is independent of $P$-value and does not say anything about the likelihood of the difference between treatments being due to chance alone
- Helps you judge the clinical significance of a statistically significant result

NNT

- How many patients would you need to treat with Drug A instead of Drug B before you would encounter 1 extra outcome of interest, such as response

Small NNT Numbers = Bigger Difference between Drug A and Drug B
Calculating NNT is Easy

What is the NNT for an outcome for Drug A vs Drug B?

\[ f_A = \text{frequency of outcome for Drug A} \]
\[ f_B = \text{frequency of outcome for Drug B} \]

\[ \text{Attributable Risk (AR)} = f_A - f_B \]

\[ \text{NNT} = \frac{1}{\text{AR}} \]

By convention, when not presenting fractions, we round up the NNT to the next \textit{higher} whole number.

For example, Drug A results in remission 50% of the time, but Drug B results in remission 20% of the time.

\[ \text{NNT} = \frac{1}{0.50 - 0.20} = \frac{1}{0.30} = 3.33 \rightarrow \text{Round up to 4} \]
Back to Our Example: Quantifying Irrelevance

Drug A
Drug B

$NNT = 100$

$P < .0001$

$NNT = 1 / (0.315 - 0.305) = 1 / 0.01 = 100$

Patients in Remission at 6 Weeks (%)

35%
34%
33%
32%
31%
30%
An NNT of $\infty$ occurs when both interventions have the same rate for the outcome measured.

*NNT values of this magnitude are irrelevant when comparing interventions except when evaluating the utility of immunizations or when examining lethal outcomes.*

Double- and triple-digit NNT values are usually not important when comparing routine efficacy measures, but may become important regarding adverse outcomes that have long-term consequences.

Single-digit NNT values are usually important enough to see differences in routine clinical practice.

- An NNT of 9 is a small effect size; NNT of 8.96 equals Cohen’s $d$ of 0.2
- An NNT of 4 is a medium effect size; NNT of 3.6 equals Cohen’s $d$ of 0.5
- An NNT of 3 is a large effect size; NNT of 2.3 equals Cohen’s $d$ of 0.8
- An NNT of 1 can only occur if one intervention has a rate of 100% for the outcome measured and the other intervention has a rate of 0%
What is NNH?

- NNH is Number Needed to Harm
- NNH is calculated the same way as NNT
- We would use the term NNH when referring to an outcome we are trying to avoid, or to refer to a disadvantage for Drug A vs Drug B
- In calculating NNT, if it is a negative number, we can call it an NNH

What is a Clinically Important NNT?

- A large NNT of $\geq 100$ means that there is little difference between choosing Drug A or Drug B for the outcome measured.
- A small NNT of 2 would be a hugely important difference.
- NNT values $< 10$ denote a potentially useful intervention.
- NNH values $> 10$ denote a potentially tolerable intervention.
- Some NNTs may be clinically important, even if they are relatively large, for example when the outcome is death.

### Examples of NNT for Non-Psychiatric Medical Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Intervention</th>
<th>Prevented Event</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>Insulin</td>
<td>Neuropathy</td>
<td>15</td>
</tr>
<tr>
<td>Acute myocardial infarction (MI)</td>
<td>Streptokinase and aspirin</td>
<td>Death in 5 weeks</td>
<td>20</td>
</tr>
<tr>
<td>Prematurely born baby</td>
<td>Prenatal corticoid</td>
<td>Respiratory distress syndrome or prematurity</td>
<td>11</td>
</tr>
<tr>
<td>Diastolic blood pressure 115–129</td>
<td>Antihypertensive drugs for 5 years</td>
<td>Death, stroke, or MI</td>
<td>3</td>
</tr>
<tr>
<td>Diastolic blood pressure 90–109</td>
<td>Antihypertensive drugs for 5 years</td>
<td>Death, stroke, or MI</td>
<td>141</td>
</tr>
</tbody>
</table>

NNT also depends on individual baseline risk.
<table>
<thead>
<tr>
<th>NNT</th>
<th>Treatment, Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Antibiotics versus placebo for preventing recurrent urinary tract infection in non-pregnant women (Albert et al., 2004)</td>
</tr>
<tr>
<td>3</td>
<td>Colchicine versus placebo to reduce pain in acute gout (Schlesinger, Schumacher, Catton, &amp; Maxwell, 2006)</td>
</tr>
<tr>
<td>4</td>
<td>Anticonvulsants versus placebo for migraine prophylaxis (Chronicle &amp; Mulleners, 2004)</td>
</tr>
<tr>
<td>5</td>
<td>Antibiotics versus placebo for cough relief in acute bronchitis (Smucny, Fahey, Becker, &amp; Glazier, 2004)</td>
</tr>
<tr>
<td>6</td>
<td>Fluocinonide versus hydrocortisone cream for discoid lupus erythematosus (Jessop, Whitelaw, &amp; Jordaan, 2001)</td>
</tr>
<tr>
<td>7</td>
<td>Antibiotics (anti-H. pylori) versus antisecretory therapy for prevention of recurrent bleeding from peptic ulcer (Gisbert et al., 2004)</td>
</tr>
<tr>
<td>8</td>
<td>Single dose dextropropoxyphene versus placebo for postoperative pain (Collins, Edwards, Moore, &amp; McQuay, 2000)</td>
</tr>
<tr>
<td>9</td>
<td>Antifungals versus placebo for preventing oral candidiasis in patients with cancer receiving treatment (Clarkson, Worthington, &amp; Eden, 2007)</td>
</tr>
<tr>
<td>10</td>
<td>Decongestant plus antihistamine for acute otitis media in children (Flynn, Griffin, &amp; Schultz, 2004)</td>
</tr>
<tr>
<td>15</td>
<td>Carotid endarterectomy versus not for symptomatic carotid stenosis (Cina, Clase, &amp; Haynes, 2000)</td>
</tr>
<tr>
<td>20</td>
<td>Oxygen versus room air resuscitation for preventing death in infants at birth (Tan, Schulze, O'Donnell, &amp; Davis, 2005)</td>
</tr>
</tbody>
</table>
### Examples of NNT for Psychiatric Medical Conditions

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Treatment Comparison</th>
<th>Outcome Measure</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major depression</td>
<td>Antidepressant vs placebo</td>
<td>50% Reduction in HAM-D</td>
<td>3</td>
</tr>
<tr>
<td>Acute mania</td>
<td>Valproate or lithium vs placebo</td>
<td>50% Reduction in SADS-M</td>
<td>5</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>Lithium vs placebo</td>
<td>Relapse</td>
<td>3</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Antipsychotic vs placebo</td>
<td>40% Reduction in BPRS or “much improved” CGI scale</td>
<td>2–5</td>
</tr>
<tr>
<td>Panic disorder</td>
<td>SSRI vs placebo</td>
<td>Panic free</td>
<td>3–6</td>
</tr>
<tr>
<td>Social phobia</td>
<td>Paroxetine vs placebo</td>
<td>“Much improved” CGI scale</td>
<td>3</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>SSRI vs placebo</td>
<td>35% Reduction in Y-BOCS</td>
<td>4–5</td>
</tr>
<tr>
<td>Bulimia nervosa</td>
<td>Antidepressants vs placebo</td>
<td>Remission</td>
<td>9</td>
</tr>
</tbody>
</table>

BPRS = Brief Psychiatric Rating Scale; CGI = Clinical Global Impression; HAM-D = Hamilton Rating Scale for Depression; SADS-M = Schizophrenia and Affective Disorders Scale; SSRI = selective serotonin reuptake inhibitor; Y-BOCS = Yale-Brown Obsessive-Compulsive Scale.

Can We Express Statistical and Clinical Significance Together?

• We can do this for NNT by also giving the CI
  – What is the range of values of NNT within which “the truth” probably exists?
  – If this range includes “infinity” it means it can take an infinite number of patients to see a difference, ie, there is no difference
  – CI tells us about the precision of our estimate of NNT

• You can calculate it with a simple formula, or use an online calculator

Calculating 95% CIs for NNT

What is the range of values of NNT within which “the truth” probably exists?

1. Remember, NNT = 1/AR, so we first calculate the CI for AR. We will need to know the total numbers of patients who received Drug A and Drug B, call them $n_A$ and $n_B$

2. Next, calculate

$$\text{offset } = 1.96 \sqrt{\frac{f_A(1-f_A)}{n_A} + \frac{f_B(1-f_B)}{n_B}}$$

3. Next, add and subtract the offset to your AR, and you now have the upper and lower bounds of the 95% CI for the AR

4. Calculate the reciprocal of these upper and lower bounds, and you now have the 95% CI for the NNT
(Free) Resources:
http://graphpad.com/quickcalcs/NNT1.cfm
EBM Summary

• EBM goes beyond anecdotal evidence, and allows the integration of clinical research into clinical practice
• The tools of EBM include the calculation of effect size such as NNT—this tells us the clinical significance of a statistically significant result
• EBM requires us to use clinical judgment in order to weigh benefits and risk for the individual patient
Bottom Line

• EBM is an important new paradigm
• It is applicable to mental health
• It can help us
  – Explain and justify our treatment decisions
  – Increase clinical effectiveness
  – Appraise the value of treatment interventions
Limitations of Using NNT / NNH

- It is most valid to calculate from an RCT with identical conditions for all drugs under study.
- Results are only calculable for binary or dichotomous events that are either present or absent, and do not apply to continuous variables such as the value of a blood test.
- However, values with clinically significant thresholds, such as weight gain > 7% can be expressed as an NNT because then they are binary.

NNT Summary

• Absolute differences place the data in a clinically meaningful context, relative differences can be deceiving
• The concept of NNT allows the clinician to estimate a medication’s potential relevant effect
• Examining the magnitudes of NNT (and NNH), the clinician can start to make risk–benefit decisions tailored to the individual patient’s needs or preferences

Outline

• What is Evidence-Based Medicine (EBM)?
• What is an effect size and how is it different from a $P$-value?
• What are Number Needed to Treat (NNT) and Number Needed to Harm (NNH)?
• Using NNT and NNH to evaluate treatment choices: The example of tardive dyskinesia
• What is Likelihood to be Helped or Harmed (LHH)?
• Using LHH to assess benefit vs risk: Returning to the example of medications approved for the treatment of tardive dyskinesia
• One more example: Pimavanserin for Parkinson’s disease psychosis
• Summary
Efficacy Illustrated the Usual Way: Valbenazine (KINECT 3)
Mean Change in AIMS Score (Fixed-Dose Study Design)

Intent-to-Treat Population: Included all randomized participants who had at least one post-randomization AIMS value.

*P<.05. **P<.01. ***P≤.001 for valbenazine vs placebo. *Dose that was statistically significantly different from placebo after adjusting for multiplicity.

Efficacy Illustrated the Usual Way: Deutetranbenazine (AIM-TD)

Mean Change in AIMS Score (Fixed-Dose Study Design)

1) Formulate the Question (PICO)

**PICO**

- **Patient:** Tardive dyskinesia
- **Intervention:** VMAT2 inhibitor
- **Control:** Placebo
- **Outcome:**
  - Reduction of dyskinetic movements by $\geq 50\%$ from baseline in 6 to 12 weeks
  - Avoidance of discontinuation because of an adverse event
2) Search for Answers

- Randomized placebo-controlled trials of the 2 FDA-approved VMAT2 inhibitor medications indicated for tardive dyskinesia

- Medline search reveals 2 systematic reviews that calculated NNT and NNH for the outcomes of interest
3) Appraise the Evidence

• Methods
  – Concealed randomization? Yes
  – Double-blind? Yes, with remote video-raters for the primary outcome (AIMS)
  – Were medication and placebo groups comparable? Yes
  – Aside from experimental treatment, treated equally? Yes
Valbenazine: Efficacy

Figure 1: ≥50% reduction in AIMS dyskinesia score from baseline at Weeks 2, 4, 6 from Hauser et al. NNT vs placebo and 95% CIs. AIMS, Abnormal Involuntary Movement Scale; CI, confidence interval; NNT, number needed to treat; ns, not significant; VBZ, valbenazine.

Valbenazine: Efficacy

For valbenazine 80 mg QD, NNT is 4 for response (≥ 50% reduction in AIMS from baseline) vs placebo at 6-week endpoint.
Valbenazine: Tolerability

As pooled from available data, discontinuation rates because of an adverse event were 2.9% for valbenazine-treated patients vs 1.6% for placebo-treated patients, resulting in an NNH of 76 (ns).
Valbenazine: Tolerability

As pooled from available data, discontinuation rates because of an adverse event were 2.9% for valbenazine-treated patients vs 1.6% for placebo-treated patients, resulting in an NNH of 76 (ns).

**NNH ≥ 10 for all tolerability outcomes of interest**

<table>
<thead>
<tr>
<th>Adverse event</th>
<th>Placebo (N=183)</th>
<th>Valbenazine (N=262)</th>
<th>NNH (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somnolence (somnolence, fatigue, sedation)</td>
<td>4.2%</td>
<td>10.9%</td>
<td>15 (9-52)</td>
</tr>
<tr>
<td>Anticholinergic effects (dry mouth, constipation, disturbance in attention, vision blurred, urinary retention)</td>
<td>4.9%</td>
<td>5.4%</td>
<td>200 (ns)</td>
</tr>
<tr>
<td>Balance disorders (balance disorder)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akathisia (akathisia)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vomiting</td>
<td>0.6%</td>
<td>2.6%</td>
<td>50 (ns)</td>
</tr>
<tr>
<td>Nausea</td>
<td>2.1%</td>
<td>2.3%</td>
<td>500 (ns)</td>
</tr>
<tr>
<td>Arthralgia</td>
<td>0.5%</td>
<td>2.3%</td>
<td>56 (ns)</td>
</tr>
</tbody>
</table>

CI, confidence interval; NNH, number needed to harm; ns, not significant. Numerators were not reported and were estimated using the reported percentages when calculating the NNH and 95% CIs.

**TABLE 6** Valbenazine for tardive dyskinesia: adverse events associated with the use of valbenazine (incidence of ≥2% and valbenazine incidence greater than placebo) that occurred during acute therapy (up to 6 weeks), percentage of subjects, and number needed to harm vs placebo and 95% confidence intervals, product

# Deutetrabenazine: Efficacy

## Table 4

Deutetrabenazine for tardive dyskinesia: analysis of cumulative proportion of AIMS responders from Anderson KE, et al\(^{16}\) and NNT vs placebo

<table>
<thead>
<tr>
<th>Threshold for response (reduction from baseline AIMS) (%)</th>
<th>Placebo (N = 58)</th>
<th>Deutetrabenazine 6 mg BID (N = 60)</th>
<th>Deutetrabenazine 12 mg BID (N = 49)</th>
<th>Deutetrabenazine 18 mg BID (N = 55)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rate, n (%)</td>
<td>Rate, n (%)</td>
<td>NNT (95% CI)</td>
<td>Rate, n (%)</td>
</tr>
<tr>
<td>10</td>
<td>29 (50.0)</td>
<td>37 (61.7)</td>
<td>9 (ns)</td>
<td>33 (67.3)</td>
</tr>
<tr>
<td>20</td>
<td>23 (39.7)</td>
<td>28 (46.7)</td>
<td>15 (ns)</td>
<td>29 (59.2)</td>
</tr>
<tr>
<td>30</td>
<td>18 (31.0)</td>
<td>19 (31.7)</td>
<td>159 (ns)</td>
<td>24 (49.0)</td>
</tr>
<tr>
<td>40</td>
<td>9 (15.5)</td>
<td>14 (23.3)</td>
<td>13 (ns)</td>
<td>22 (44.9)</td>
</tr>
<tr>
<td>50</td>
<td>7 (12.1)</td>
<td>8 (13.3)</td>
<td>80 (ns)</td>
<td>17 (34.7)</td>
</tr>
<tr>
<td>60</td>
<td>3 (5.2)</td>
<td>4 (6.7)</td>
<td>67 (ns)</td>
<td>10 (20.4)</td>
</tr>
<tr>
<td>70</td>
<td>1 (1.7)</td>
<td>2 (3.3)</td>
<td>63 (ns)</td>
<td>6 (12.2)</td>
</tr>
<tr>
<td>80</td>
<td>1 (1.7)</td>
<td>1 (1.7)</td>
<td>-1740(^{a}) (ns)</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td>90</td>
<td>1 (1.7)</td>
<td>0</td>
<td>-58(^{a}) (ns)</td>
<td>0</td>
</tr>
</tbody>
</table>

AIMS, abnormal involuntary movement scale; CI, confidence interval; NNT, number needed to treat; ns, not significant.

\(^{a}\)A negative value for NNT occurs where the rate of the benefit is higher for placebo than for deutetrabenazine.

Deutetrabenazine: Efficacy

For doses of deutetrabenazine 12 mg BID and 18 mg BID, NNT is 5 for response (≥ 50% reduction in AIMS from baseline) vs placebo at 12-week endpoint

<table>
<thead>
<tr>
<th>Threshold for response (reduction from baseline AIMS)</th>
<th>Placebo (N = 58)</th>
<th>Deutetrabenazine 6 mg BID (N = 60)</th>
<th>Deutetrabenazine 12 mg BID (N = 49)</th>
<th>Deutetrabenazine 18 mg BID (N = 55)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>7 (12.1)</td>
<td>8 (13.3)</td>
<td>80 (ns)</td>
<td>17 (34.7)</td>
</tr>
<tr>
<td>60</td>
<td>3 (5.2)</td>
<td>4 (6.7)</td>
<td>67 (ns)</td>
<td>10 (20.4)</td>
</tr>
<tr>
<td>70</td>
<td>1 (1.7)</td>
<td>2 (3.3)</td>
<td>63 (ns)</td>
<td>6 (12.2)</td>
</tr>
<tr>
<td>80</td>
<td>1 (1.7)</td>
<td>1 (1.7)</td>
<td>-1740(^\circ) (ns)</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td>90</td>
<td>1 (1.7)</td>
<td>0 (ns)</td>
<td>-58(^\circ) (ns)</td>
<td>0 (ns)</td>
</tr>
</tbody>
</table>

AIMS, abnormal involuntary movement scale; CI, confidence interval; NNT, number needed to treat; ns, not significant.

\(^\circ\) A negative value for NNT occurs where the rate of the benefit is higher for placebo than for deutetrabenazine.

As pooled from available data, discontinuation rates because of an adverse event were 3.6% for deutetrabenzazine-treated patients vs 3.1% for placebo-treated patients, resulting in an NNH of 189 (ns).

As pooled from available data, discontinuation rates because of an adverse event were 3.6% for deutetrabenazine-treated patients vs 3.1% for placebo-treated patients, resulting in an NNH of 189 (ns).

4) Apply the Results

- Is my patient like those studied?
  - Psychiatrically stable?
  - If on an antipsychotic medication, is on a stable dose?
- Is treatment consistent with my patient’s values and preferences?
- Is treatment feasible in my practice setting?
  - Formulary?
  - Cost?
  - Access to patient assistance program?
5) Assess the Process

- Is it working?
- Will it continue to work?
Valbenazine: 
Long-Term Data is Reassuring

Of the 163 participants included in the analyses, 149 completed the Week 8 visit and 103 completed Week 48.

No new safety signals or concerns emerged in this long-term study.

Sustained improvements were found in adults with TD who received once-daily valbenazine for up to 48 weeks, based on clinician- and patient-rated measures.

Factor SA, et al. Effects of Long-Term Valbenazine on Tardive Dyskinesia and Patient-Reported Outcomes: Results from the KINECT 4 Study. Presented at: 70th Annual Meeting of the American Academy of Neurology; April 21–27, 2018; Los Angeles, CA.
Deutetrabenazine: Long-Term Data is Reassuring

2-year (Week 106) open-label response rates are reported in this interim analysis. Of 343 patients enrolled in the extension study, 232 previously received deutetrabenazine and 111 previously received placebo.

The mean total daily dose of deutetrabenazine at Week 80 was 38.6 (1.13) mg for all patients.

No new safety signals or concerns emerged in this long-term study.
Long-Term Side Effects

• Good news, in long-term studies in TD with both the VMAT2 therapies (deutetrabenazine and valbenazine), no new or unexpected side effects emerged

• But…no controlled studies

• No direct comparison trials between the 2 options

Outline

• What is Evidence-Based Medicine (EBM)?
• What is an effect size and how is it different from a \( P \)-value?
• What are Number Needed to Treat (NNT) and Number Needed to Harm (NNH)?
• Using NNT and NNH to evaluate treatment choices: The example of tardive dyskinesia
• What is Likelihood to be Helped or Harmed (LHH)?
• Using LHH to assess benefit vs risk: Returning to the example of medications approved for the treatment of tardive dyskinesia
• One more example: Pimavanserin for Parkinson’s disease psychosis
• Summary
LHH: Likelihood to be Helped or Harmed

- LHH ratios (likelihood of being helped or harmed) can be a valid and useful way of synthesizing data regarding benefits and risks
- LHH = \( \frac{1}{NNT} / \frac{1}{NNH} = \frac{NNH}{NNT} \)

LHH: Likelihood to be Helped or Harmed

• Example:
  – NNT to prevent 1 additional case of stroke by using warfarin for a person with atrial fibrillation was found to be 39
  – The patient could be told that he has a 1 in 39 chance of being helped by warfarin therapy with a stroke being prevented
  – However, the NNH for 1 additional event of major hemorrhage (including gastrointestinal bleeding) is 333
  – The patient can be told that if he were to receive warfarin, the likelihood of him having a major hemorrhage is 1 in 333
LHH: Likelihood to be Helped or Harmed

• Example (cont’d):
  – The LHH in this example would be LHH = (1/NNT) : (1/NNH) = 1/39 : 1/333 = 9 to 1 in favor of warfarin
  – This translates to “warfarin treatment is 9× as likely to help you as to harm you”
• An LHH > 1 would mean the likelihood to be helped is greater than the likelihood to be harmed; for an LHH < 1, the reverse is true
LHH:
Likelihood to be Helped or Harmed

• Benefit–risk can thus be quantified by calculating LHH, provided that the benefit and harm being considered are relevant to the particular individual and are logically matched in terms of expected time course and consequences

• Remember that benefits and risks can take on greatly differing degrees of importance or relevance depending on the subjective point of view of the patient and clinician, baseline risks, and severity of the underlying illness

Outline

• What is Evidence-Based Medicine (EBM)?
• What is an effect size and how is it different from a $P$-value?
• What are Number Needed to Treat (NNT) and Number Needed to Harm (NNH)?
• Using NNT and NNH to evaluate treatment choices: The example of tardive dyskinesia
• What is Likelihood to be Helped or Harmed (LHH)?
• Using LHH to assess benefit vs risk: Returning to the example of medications approved for the treatment of tardive dyskinesia
• One more example: Pimavanserin for Parkinson’s disease psychosis
• Summary
Deutetrabenazine vs Valbenazine: NNT in Fixed-Dose Trials

**Figure 1** Reduction (≥50%) in AIMS dyskinesia score from baseline to end-point, NNT vs placebo and 95% CIs, for the Phase III fixed-dose studies of deutetrabenazine and valbenazine; data for deutetrabenazine from Table 3, data for valbenazine from Figure 1 in Citrome. 

Deutetrabenazine vs Valbenazine: NNT in Fixed-Dose Trials

- NNT for response for either medication was ~5
- NNH vs placebo for discontinuation because of an adverse effect in the fixed dose studies for either medication was ~100
- Thus LHH for response vs discontinuation because of an adverse effect is ~20
Outline

• What is Evidence-Based Medicine (EBM)?
• What is an effect size and how is it different from a $P$-value?
• What are Number Needed to Treat (NNT) and Number Needed to Harm (NNH)?
• Using NNT and NNH to evaluate treatment choices: The example of tardive dyskinesia
• What is Likelihood to be Helped or Harmed (LHH)?
• Using LHH to assess benefit vs risk: Returning to the example of medications approved for the treatment of tardive dyskinesia
• One more example: Pimavanserin for Parkinson’s disease psychosis
• Summary
Pimavanserin 34 mg vs Placebo
ACP-103-020

PIM = pimavanserin; PBO = placebo; SAPS-PD = Scale for the Assessment of Positive Symptoms for Parkinson’s Disease Psychosis.
Pimavanserin 34 mg vs Placebo
ACP-103-020

These are the results of the primary outcome measure, but is this clinically relevant? Can we quantify this?

• LHH = NNH/NNT and answers the question of how often would one encounter a benefit vs a harm
• A useful definition of response is a ≥ 3-point decrease from baseline on SAPS-PD (a 2.33-point change on the SAPS-PD corresponds to a clinically meaningful 1-unit change in the CGI-I scale)
  – This was observed for 68.4% of patients receiving PIM 34 mg and in 43.3% of patients receiving placebo in the pivotal efficacy trial, for an NNT of 4 (95% CI 3–9)
• The NNT is 4 for a ≥ 3-point decrease from baseline on SAPS-PD for PIM 34 mg/day vs placebo and NNH is 21 for the overall tolerability metric of discontinuation because of adverse event from all pooled data for PIM 34 mg/day vs placebo
• The resulting LHH is 21/4 = 5.25
Likelihood to be Helped or Harmed (LHH)

Number Needed to Treat or Harm vs Placebo

- Number needed to treat for 3 point decrease from Baseline on SAPS-PD: 4
- Number needed to harm for discontinuation because of an adverse event: 21

LHH = 5.25

PIM 34 mg/day is about 5× more likely to result in response (≥ 3-point decrease from baseline on SAPS-PD) than discontinuation because of an adverse event.
Are LHH Values > 5 Typical? Not Always … New Antidepressants

- NNT for response vs placebo calculated
  - Response defined as ≥ 50% reduction from baseline on the MADRS or HAM-D
- NNH for poor tolerability vs placebo calculated
  - The tolerability outcome of interest was discontinuation because of an adverse event
- LHH calculated to contrast efficacy vs tolerability

MADRS = Montgomery–Åsberg Depression Rating Scale.
• NNT values < 10 denote reasonable efficacy
• NNH values > 10 reasonable tolerability
• Lower values for NNT and higher values for NNH are desirable
• The ratio of NNH to NNT is the LHH; higher values of LHH are desirable

Are LHH Values > 5 Typical? Not Always…

<table>
<thead>
<tr>
<th>Drug</th>
<th>NNT</th>
<th>NNH</th>
<th>LHH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duloxetine</td>
<td>5.7</td>
<td>24.5</td>
<td>4.3</td>
</tr>
<tr>
<td>Escitalopram</td>
<td>6.7</td>
<td>30.7</td>
<td>4.6</td>
</tr>
<tr>
<td>Levomilnacipran</td>
<td>9.8</td>
<td>18.2</td>
<td>1.8</td>
</tr>
<tr>
<td>Sertraline</td>
<td>5.3</td>
<td>5.3</td>
<td>1.2</td>
</tr>
<tr>
<td>Venlafaxine</td>
<td>5.7</td>
<td>7.8</td>
<td>1.4</td>
</tr>
<tr>
<td>Vilazodone</td>
<td>8.0</td>
<td>26.1</td>
<td>3.3</td>
</tr>
<tr>
<td>Vortioxetine</td>
<td>8.4</td>
<td>42.7</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Outline

• What is Evidence-Based Medicine (EBM)?
• What is an effect size and how is it different from a P-value?
• What are Number Needed to Treat (NNT) and Number Needed to Harm (NNH)?
• Using NNT and NNH to evaluate treatment choices: The example of tardive dyskinesia
• What is Likelihood to be Helped or Harmed (LHH)?
• Using LHH to assess benefit vs risk: Returning to the example of medications approved for the treatment of tardive dyskinesia
• One more example: Pimavanserin for Parkinson’s disease psychosis
• Summary
Summary

• NNT and NNH can be easily calculated
• NNT and NNH must be interpreted in a clinical context
  – Every patient has individual patterns of treatment and a different set of preferences
• LHH can help when discussing trade-offs between benefit and risk, provided both are meaningful to the clinician and patient, and are comparable in terms of time course
Practical Take-Aways

• Statistically significant study results can be clinically meaningless if the effect size is small

• Effect sizes such as Number Needed to Treat (NNT) and Number Needed to Harm (NNH) are easy to calculate and can help quantify the clinical relevance of statistically significant results

• Likelihood to be helped or harmed (LHH), calculated as the ratio of NNH to NNT, can be used to illustrate trade-offs between benefits and harms