Show Me the Evidence!
Using Number Needed to Treat

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

What is the Process of EBM?
- Quantifying benefit and risk using NNT and NNH
- Interpreting clinical trials
  - CATIE for schizophrenia
  - Placing new antidepressants for MDD into context

EBM = evidence-based medicine; NNT = number needed to treat; NNH = number needed to harm;
CATIE = Clinical Antipsychotic Trials of Intervention Effectiveness; MDD = major depressive disorder.

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

![Graph showing remission rates for Drug A and Drug B, with P < .0001](image)

How irrelevant is this? Can we quantify this?

What is Evidence-Based Medicine?

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”
Evaluation of the Quality of Data Requires Vigilance and an Organized Approach

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Getting “Out-of-Date” Can Result in: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

- Formulate the question
- Search for answers
- Appraise the evidence
- Apply the results
- Assess the outcome

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

NICE (National Institute for Clinical Excellence)

- United Kingdom’s independent organization responsible for providing national guidance on the promotion of good health and the prevention and treatment of ill health
- WWW.NICE.ORG.UK
- Evidence-based practice guidelines
- Focus on quality of evidence assessed through systematic reviews of RCTs rather than list of treatment alternatives

Online Resources: Up-to-Date and Evidence-Based

- BMJ Clinical Evidence
- Up-to-Date
- Evidence-Based Medicine
Algorithms

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

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?

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?

How Involved in EBM Should You Get?

- "Doer" uses EBM methods to formulate and answer questions, assess evidence
- "User" consults pre-appraised resources
- "Replicator" follows
  - Recommendations of EBM leaders
  - Evidence-based guidelines

Caution: Not all algorithms are evidence-based. There are many eminence-based algorithms out there!
Show Me the Evidence!

- What is the Process of EBM?
- Quantifying benefit and risk using NNT and NNH
- Interpreting clinical trials
  - CATIE for schizophrenia
  - Placing new antidepressants for MDD into context
- Summary

Evidence-Based Medicine 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 nonrandom effect
- P-value is not the same as effect size

Concepts Related to Benefit / Risk: Effect Size – Number Needed to Treat

- 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

Number Needed to Treat

- 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
  
  The smaller the NNT, the larger the differences between the 2 drugs, ie, larger numbers mean more patients needed to treat to see the difference in effect

Calculating NNT is Easy

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

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

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

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

By convention, when not presenting fractions, we round up the NNT to the next higher whole number

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

\[ \text{NNT} = \frac{1}{0.50-0.20} = \frac{1}{0.30} = 3.33 \]  
Round up to 4
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

\[ P < .0001 \]

\[ \text{NNT} = 100 \]

\[ \text{NNT} = \frac{1}{0.315 - 0.305} = \frac{1}{0.01} = 100 \]

What is NNH?

- NNH is Number Needed to Harm
- We would use 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 100 or more 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
- Some NNTs may be clinically important, even if they are relatively large, for example when the outcome is death

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%.

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</td>
<td>Antihypertensive</td>
<td>Death, stroke, or MI</td>
<td>141</td>
</tr>
</tbody>
</table>

NNT also depends on individual baseline risk.
**Examples of NNT for Psychiatric 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>Obsessive-compulsive disorder</td>
<td>Paroxetine vs placebo</td>
<td>“Much improved” CGI scale</td>
<td>3</td>
</tr>
<tr>
<td>Bulimia nervosa</td>
<td>Antidepressants vs placebo</td>
<td>35% Reduction in Y-BOCS</td>
<td>4–5</td>
</tr>
</tbody>
</table>

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

**P-Values vs NNT**

<table>
<thead>
<tr>
<th><strong>P-VALUE</strong></th>
<th><strong>NNT</strong></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 P-Value</td>
</tr>
</tbody>
</table>

**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 \( offset = \frac{1}{n_A} + \frac{1}{n_B} \)
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

**Evidence-Based Medicine 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

**Free Resources:**

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 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

Show Me the Evidence!

- What is the Process of EBM?
- Quantifying benefit and risk using NNT and NNH
- Interpreting clinical trials
  - CATIE for schizophrenia
  - Placing new antidepressants for MDD into context
- Summary

Which antipsychotic should I prescribe for my patient with schizophrenia?

1) Formulate the Question (PICO)

"Should I switch to olanzapine, quetiapine, risperidone, ziprasidone, or clozapine?"
1) Formulate the Question (PICO)

PICO
- **Patient:** Schizophrenia and switching medication is contemplated
- **Intervention:** A second-generation antipsychotic
- **Control:** Other antipsychotic
- **Outcome:**
  - Effectiveness as defined by remaining on treatment, thought to be an integration of efficacy, tolerability, and adherence
  - Avoidance of untoward effects

2) Search for Answers

- Large effectiveness trials may provide guidance
- Medline search reveals a large effectiveness trial that was randomized, mostly double-blind, and that compared multiple antipsychotics
  - **Patient:** Schizophrenia, not first-episode, not refractory, can have comorbid medical conditions, can have comorbid alcohol or substance use disorder
  - **Intervention:** Oral antipsychotic
  - **Control:** Other oral antipsychotic
  - **Outcome:**
    - Time on medication; all-cause discontinuation
    - Multiple tolerability outcomes

**CATIE Trial Design**

An effectiveness study that tested switches: time to all-cause discontinuation was the primary outcome measure

**CATIE Trial Design**

Of the 74% that discontinued Phase 1, approximately half entered Phase 2

- **Phase 1**
  - 1894 screened
  - 1493 randomized
  - 1460 after one site excluded
  - 1432 received Rx

- **Phase 2**
  - Participants who discontinued Phase 1 choose either the clozapine or the ziprasidone randomization pathways

- **Phase 3**
  - Participants who discontinued Phase 2 choose one of the following open-label treatments

- **Phase 1A:** participants with TD (N = 231) do not get randomized to perphenazine; Phase 1B: participants who fail to show response to perphenazine will be randomized to an atypical (olanzapine, quetiapine, or risperidone) before eligibility for Phase 2.

- **Phase 18:** participants with TD (N = 231) do not get randomized to perphenazine; Phase 1B: participants who fail to show response to perphenazine will be randomized to an atypical (olanzapine, quetiapine, or risperidone) before eligibility for Phase 2.


3) Appraise the Evidence

- Methods
  - Concealed randomization? Yes
  - Double-blind? Yes, except for clozapine pathway in Phase 2
  - Were groups comparable? Yes, except for the perphenazine cohort for whom TD was an exclusion criterion
  - Aside from experimental treatment, treated equally? Yes

THESE ARE THE PHASE 1 EFFECTIVENESS DATA.
ANY QUESTIONS?

Phase 1: All-Cause Discontinuation


Phase 1: Olanzapine Has Advantages


Phase 1: Risperidone and Perphenazine Have Advantages Too


We can list the NNTs and the CIs for all-cause discontinuation and for discontinuation for a specific reason.

When the CI includes "infinity," the NNT is not statistically significant.
We can list the NNHs and the CIs for adverse events. Their relative importance is greatly influenced by what the patient thinks about them.


The NNT in CATIE Phase 1

<table>
<thead>
<tr>
<th>COMPARISON (Phase 1)</th>
<th>OLZ vs RIS</th>
<th>OLZ vs QUE</th>
<th>OLZ vs ZIP</th>
<th>OLZ vs PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>D/C All cause</td>
<td>11*</td>
<td>6*</td>
<td>7*</td>
<td>5*</td>
</tr>
<tr>
<td>D/C Efficacy loss</td>
<td>4*</td>
<td>6*</td>
<td>11*</td>
<td>15*</td>
</tr>
<tr>
<td>D/C Intolerability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D/C Weight or Metabolic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rx Antidiabetic</td>
<td>-82</td>
<td>-67</td>
<td>-71</td>
<td>-61</td>
</tr>
<tr>
<td>Rx Statin</td>
<td>-81</td>
<td>-223</td>
<td>-30*</td>
<td>-57</td>
</tr>
</tbody>
</table>

*Statistically significant (95% CI did not cross from + to -). Negative numbers indicate advantage for the non-olanzapine comparator.


OLanzapine performed well in Phase 1 overall because the signal for efficacy had a larger effect size than the signal for discontinuation due to weight gain or metabolic effects.


But...Quetiapine Looks (A Lot) Better in Phase 1B

All-Cause Discontinuation and Number Needed to Treat

<table>
<thead>
<tr>
<th>Phase 1B patients had failed perphenazine; they were re-randomized and had to proceed through this Phase prior to being eligible for Phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNT 4</td>
</tr>
<tr>
<td>50%</td>
</tr>
</tbody>
</table>

OLanzapine performed well in Phase 1 overall because the signal for efficacy had a larger effect size than the signal for discontinuation due to weight gain or metabolic effects.


Quetiapine performed well in Phase 1B overall because the signal for efficacy had a larger effect size than the signal for discontinuation due to weight gain or metabolic effects.


Phase 2E: Clozapine Pathway Results

All-Cause Discontinuation and Number Needed to Treat

<table>
<thead>
<tr>
<th>Phase 2E: Clozapine Pathway Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNT 4</td>
</tr>
<tr>
<td>71%</td>
</tr>
</tbody>
</table>

Phase 2T: Ziprasidone Pathway Results

All-Cause Discontinuation and Number Needed to Treat

<table>
<thead>
<tr>
<th>Phase 2T: Ziprasidone Pathway Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNT 6</td>
</tr>
<tr>
<td>77%</td>
</tr>
</tbody>
</table>

Olanzapine performed well in Phase 2T overall because the signal for efficacy had a larger effect size than the signal for discontinuation due to weight gain or metabolic effects.


Quetiapine performed well in Phase 2T overall because the signal for efficacy had a larger effect size than the signal for discontinuation due to weight gain or metabolic effects.


Ziprasidone performed well in Phase 2T overall because the signal for efficacy had a larger effect size than the signal for discontinuation due to weight gain or metabolic effects.


*Statistically significant (95% CI did not cross from + to -). Negative numbers indicate advantage for the non-ziprasidone comparator.


Phase 2T: Ziprasidone Pathway Results

All-Cause Discontinuation and Number Needed to Treat

<table>
<thead>
<tr>
<th>Phase 2T: Ziprasidone Pathway Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNT 6</td>
</tr>
<tr>
<td>77%</td>
</tr>
</tbody>
</table>

Olanzapine performed well in Phase 2T overall because the signal for efficacy had a larger effect size than the signal for discontinuation due to weight gain or metabolic effects.


Quetiapine performed well in Phase 2T overall because the signal for efficacy had a larger effect size than the signal for discontinuation due to weight gain or metabolic effects.


Ziprasidone performed well in Phase 2T overall because the signal for efficacy had a larger effect size than the signal for discontinuation due to weight gain or metabolic effects.


*Statistically significant (95% CI did not cross from + to -). Negative numbers indicate advantage for the non-ziprasidone comparator.

What was Ziprasidone’s Principal Advantage?

![Graph showing weight loss and gain in patients with phase 1 results.]

* N = 61, statistical significance not calculated, only NNT relative to ZIP shown.


4) Apply the Results

- Is my patient like those studied?
  - Ambulatory patient, non-treatment refractory?
  - Not schizoaffective
  - Not first-episode
- Is treatment consistent with my patient’s values and preferences?
- Is treatment feasible in my practice setting?
  - Formulary?
  - Cost?

How Does This Apply to My Patient?

- Switches offer both opportunity and risk
- Where you end depends on where you start
  - Did the patient fail a “tight” D₂ binding agent?
  - Did the patient fail because of efficacy or tolerability?
  - Is weight gain greater than 7% the predominant concern?
  - Is risk for hospitalization the predominant concern?

Show Me the Evidence!

- What is the Process of EBM?
- Quantifying benefit and risk using NNT and NNH
- Interpreting clinical trials
  - CATIE for schizophrenia
  - Placing new antidepressants for MDD into context
- 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 = (1/NNT) / (1/NNH) = NNH/NNT

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 (continued):
  – 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 times as likely to help you as to harm you”
• An LHH greater than 1 would mean the likelihood to be helped is greater than the likelihood to be harmed; for an LHH less than 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

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

New Antidepressants

• 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

Show Me the Evidence!

• What is the Process of EBM?
• Quantifying benefit and risk using NNT and NNH
• Interpreting clinical trials
  – CATIE for schizophrenia
  – Placing new antidepressants for MDD into context
• 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