CHALLENGES WITH DEFINING CLINICAL MEANINGFULNESS IN CNS DISEASES

Richard Keefe, Duke University Medical Center

Clinical Meaningfulness in Drug Development for Early Alzheimer’s Disease
An FDA/Alzheimer’s Disease Allies Meeting

Bethesda, MD, November 6, 2014
CONSULTANT/AD BOARD/SERVICE PROVIDER
Abbvie, Akebia, Amgen, Astellas, Asubio, AviNeuro/ChemRar, Biogen Idec, BiolineRx, Biomarin, Boehringer-Ingelheim, Eli Lilly, EnVivo/FORUM, GW Pharmaceuticals, Helicon, Lundbeck, Merck, Mitsubishi, Novartis, Otsuka, Pfizer, Roche, Shire, Sunovion, Takeda, Targacept

RESEARCH FUNDING
Allon, Department of Veteran’s Affairs, Feinstein Institute for Medical Research, GlaxoSmithKline, NIMH, Novartis, Psychogenics, Research Foundation for Mental Hygiene, Singapore Medical Research Council

FOUNDER OF NEUROCOG TRIALS, INC.
Providing rater training, data quality assurance and consultation to several pharmaceutical companies and other consortia

SHAREHOLDER
Sengenix

ROYALTIES
Brief Assessment of Cognition in Schizophrenia (BACS), MATRICS Consensus Cognitive Battery (MCCB), Virtual Reality Functional Capacity Assessment Tool (VRFCAT)
GOALS OF CLINICAL TRIALS

1. Is this drug **safe**?

2. Is this drug **efficacious**?

   ... ________?

   .... ________?

N. Will this drug produce a **clinically meaningful effect**?
“Suppose a well-done randomized clinical trial (RCT) reports a statistically significant difference between treatment (T) and control (C) groups, with $p=.05$, $p=.01$, even $p=10^{-6}$. Should these results be automatically considered of clinical significance, the basis of recommending that clinicians use T rather than C for patients like those studied? No. What would be needed in addition to infer clinical significance is the subject of this review.”

Kraemer HC, Kupfer DJ. Size of Treatment Effects and Their Importance to Clinical Research and Practice. Biol Psychiatry 2006;59:990–996
### Session 3: Defining Clinically Meaningful Effect for the Design and Interpretation of RCTs

**Chairs:**
- R Keefe PhD
- AC Leon PhD
- AC Leon PhD

**Introduction and Objectives**

8:50 AM  Consumer Perspective Discussion
- J McNulty Ab, ScB

9:20 AM  Payer Perspective Discussion
- R Epstein MD

9:50 AM  Health Care Economist Perspective Discussion
- S Reed PhD

10:05 AM  Break

**Investor Perspective Discussion**
- J Sanchez

11:10 AM  Regulatory Perspective and Discussion Panel
- T Laughren MD
- K Broich MD

### Noon

**ISCTM Annual Business Meeting Luncheon (Overflow seating ‘Chit-Chat’ Room - Salon II)**

**2:00 PM**

**Session 3 Continues: What are the Statistical Methods for Determining a Clinically Meaningful Effect?**

- AC Leon PhD

**2:15 PM**

**The Special Case of Clinically Meaningful Similarity Discussion**

**2:20 PM**

**Determining How Effective a Treatment Will be for an Individual Patient Discussion**

- H Kraemer PhD

**2:50 PM**

**3:00 PM**

**Break**

**3:20 PM**

**How to Determine Benefit in a Clinical Trial (Incorporating Risk and Benefit) Discussion**

- E Frank PhD

**3:50 PM**

**4:00 PM**

**Comments**

- K Davis MD
- S Romano MD
- H Kraemer PhD
- E Frank PhD
- R Keefe PhD
- AC Leon PhD

**Focused Discussion Panel**

**5:00 PM**

**Session Adjourns**

---

**Defining a Clinically Meaningful Effect for the Design and Interpretation of Randomized Controlled Trials**

by Richard S. Keefe, PhD; Helena C. Kraemer, PhD; Robert S. Epstein, MD, MS; Ellen Frank, PhD; Ginger Haynes, PhD; Thomas P. Laughren, MD; James McNulty, ABScS; Shelby D. Reed, PhD; Juan Sanchez, MD; and Andrew C. Leon, PhD
T-TEST
EFFECT SIZE FOR A COMPARISON OF GROUP MEANS

\[ d = \frac{\bar{X}_1 - \bar{X}_2}{s} \]

\( s = \) pooled standard deviation for entire sample

Ratio of between groups difference/within group differences

Group difference in standard deviation units

Used for CRT sample size estimates

NUMBER NEEDED TO TREAT
NNT

\[
\text{NNT} = \frac{1}{(R_A - R_C)}
\]

WHERE:
- \( R_A \) = % responders in Active group
- \( R_C \) = % responders in Control group

EXAMPLES:
- NNT = \( \frac{1}{(50\% - 40\%)} \) = 10
- NNT = \( \frac{1}{(50\% - 10\%)} \) = 2.5
NNT’S & ES’S

Cohen’s suggestions:
0.2 = small; ex: height difference between 15 year old girl and 16 year old girl
0.5 = medium (“visible to the naked eye”); ex: IQ differences between professionals and managers
0.8 = large; ex: IQ differences between PhD’s and college freshman

Tom Laughren, former Director of Psychiatry Products Division, FDA: “if a drug effect on cognitive impairment in schizophrenia was large enough, we wouldn’t need a functional co-primary.”

Cohen’s d and its Rescaling r for Outcome Data Having Normal Distributions with Equal Variances in the Treatment and Control Groups, translated to the Equivalent Values of AUC, SRD, and NNT.

<table>
<thead>
<tr>
<th>Cohen’s d</th>
<th>r</th>
<th>AUC</th>
<th>SRD</th>
<th>NNT</th>
<th>n for 80% Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>-∞</td>
<td>-1.00</td>
<td>.000</td>
<td>-1.000</td>
<td>-1.000</td>
<td></td>
</tr>
<tr>
<td>-1.0</td>
<td>-0.447</td>
<td>.240</td>
<td>-0.521</td>
<td>-1.921</td>
<td></td>
</tr>
<tr>
<td>-0.9</td>
<td>-0.410</td>
<td>.262</td>
<td>-0.475</td>
<td>-2.103</td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>-0.371</td>
<td>.286</td>
<td>-0.428</td>
<td>-2.334</td>
<td></td>
</tr>
<tr>
<td>-0.7</td>
<td>-0.330</td>
<td>.310</td>
<td>-0.379</td>
<td>-2.636</td>
<td></td>
</tr>
<tr>
<td>-0.6</td>
<td>-0.287</td>
<td>.336</td>
<td>-0.329</td>
<td>-3.043</td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.243</td>
<td>.362</td>
<td>-0.276</td>
<td>-3.619</td>
<td></td>
</tr>
<tr>
<td>-0.4</td>
<td>-0.196</td>
<td>.389</td>
<td>-0.223</td>
<td>-4.490</td>
<td></td>
</tr>
<tr>
<td>-0.3</td>
<td>-0.148</td>
<td>.416</td>
<td>-0.168</td>
<td>-5.953</td>
<td></td>
</tr>
<tr>
<td>-0.2</td>
<td>-0.100</td>
<td>.444</td>
<td>-0.112</td>
<td>-8.892</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-0.050</td>
<td>.472</td>
<td>-0.056</td>
<td>-17.739</td>
<td></td>
</tr>
<tr>
<td>.1</td>
<td>0.050</td>
<td>.528</td>
<td>.056</td>
<td>17.739</td>
<td>1,220</td>
</tr>
<tr>
<td>.2</td>
<td>0.100</td>
<td>.556</td>
<td>.112</td>
<td>8.892</td>
<td>306</td>
</tr>
<tr>
<td>.3</td>
<td>0.148</td>
<td>.584</td>
<td>.168</td>
<td>5.953</td>
<td>139</td>
</tr>
<tr>
<td>.4</td>
<td>0.196</td>
<td>.611</td>
<td>.223</td>
<td>4.490</td>
<td>107</td>
</tr>
<tr>
<td>.5</td>
<td>0.243</td>
<td>.638</td>
<td>.276</td>
<td>3.619</td>
<td>61</td>
</tr>
<tr>
<td>.6</td>
<td>0.287</td>
<td>.664</td>
<td>.329</td>
<td>3.043</td>
<td>39</td>
</tr>
<tr>
<td>.7</td>
<td>0.330</td>
<td>.690</td>
<td>.379</td>
<td>2.636</td>
<td>26</td>
</tr>
<tr>
<td>.8</td>
<td>0.371</td>
<td>.714</td>
<td>.428</td>
<td>2.334</td>
<td>20</td>
</tr>
<tr>
<td>.9</td>
<td>0.410</td>
<td>.738</td>
<td>.475</td>
<td>2.103</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>0.447</td>
<td>.760</td>
<td>.521</td>
<td>1.921</td>
<td>13</td>
</tr>
<tr>
<td>∞</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

AUC, area under the receives operating characteristic curve; SRD, success rate difference; NNT, number needed to treat.

Kraemer HC, Kupfer DJ. Size of Treatment Effects and Their Importance to Clinical Research and Practice. Biol Psychiatry 2006;59:990–996
MULTIPLE CO-PRIMARY ENDPOINTS: MEDICAL AND STATISTICAL SOLUTIONS

A Report From the Multiple Endpoints Expert Team of the Pharmaceutical Research and Manufacturers of America

<table>
<thead>
<tr>
<th>Correlation</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>64%</td>
<td>51%</td>
<td>41%</td>
<td>14%</td>
</tr>
<tr>
<td>0.2</td>
<td>66%</td>
<td>55%</td>
<td>47%</td>
<td>25%</td>
</tr>
<tr>
<td>0.5</td>
<td>69%</td>
<td>61%</td>
<td>56%</td>
<td>40%</td>
</tr>
<tr>
<td>0.8</td>
<td>73%</td>
<td>69%</td>
<td>66%</td>
<td>58%</td>
</tr>
</tbody>
</table>

Probability of Achieving Statistical Significance on All Primary Endpoints Under the Intersection-Union Test Approach, Assuming Equal Effect Size on All Endpoints and a Sample Size That Gives 80% Power to Detect the Effect Size for an Individual Endpoint

WHAT ARE OUR HOPES?

**NUMBER NEEDED TO TREAT**

- **LARGE EFFECT SIZE**: 1
- **MEDIUM EFFECT SIZE**: 3
- **SMALL EFFECT SIZE**: 9

**COHEN’S d**

- 0.8 LARGE
- 0.5 MED.
- 0.2 SMALL
WHAT ARE SOME PRECEDENTS?

ENDPOINTS MATTER

Atorvastatin for reduction in LDL\(^1\): NNT = 1-2

Atorvastatin for stroke among high-risk, non CHD pts\(^2\): NNT = 100

1. http://www.accessdata.fda.gov/drugsatfda_docs/nda/pre96/020702_s000.pdf
2. http://www.medicine.ox.ac.uk/bandolier/booth/cardiac/statascot.html
HOW DO CNS THERAPIES COMPARE?

Risperidone versus placebo\(^1\): NNT = 3
Donepezil versus placebo for AD\(^2\): NNT = 6
Glatiramer for MS (relapse free at 1-yr)\(^3\): NNT = 16.7
Internet CBT for GAD\(^2\): NNT = 1.75

4. Freedman et al., Eur Neurol 2008;60:1–11
EFFECT SIZES AND REVENUES

PATIENT/CONSUMER PERSPECTIVE

Jim McNulty, PhD
Past President, National Alliance on Mental Illness
### RESULTS FROM FOUR FOCUS GROUPS

**Volunteers from Depression & Bipolar Alliance Affiliated Support Group and Public Mental Health System Clients with ETOH or SA in Recovery**

**N=34**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>17</td>
<td>50</td>
</tr>
<tr>
<td>Males</td>
<td>17</td>
<td>50</td>
</tr>
</tbody>
</table>

**Age Range**

- 23 – 72 yrs

**Ethnicity**

- White (26)
- African American (7)
- Indian American (2)

**Education**

- 100% Finished High School
- 26% Some college currently enrolled in bachelors level degree-granting programs
- 16% B.A or B.S.
- 14% Masters
- 3% PhD Candidate
- 1% ABD

**Employment**

- 18% Working/full time students
- 16% Unemployed

**Venue**

- Private Psychiatric hospital
- Providence, RI

**DX**

- 4 SZ, 3 SZA, 8 MDD, 19 Bipolar disorder (12 Bipolar 2 or NOS, 7 Bipolar 1 – 3 of these ETOH or SA), Serious Anxiety cited as problem for 24.

MOST HAD LOW EXPECTATIONS FOR Rx

DBSA SG HAD INTERESTING SPLIT: Under 35 has higher expectations (cure or substantial symptom reduction equivalent to cure – cure not defined)

- Anxiety most prominent complaint; depression 2nd
- anergia, anhedonia, isolating behavior;
- manic and hypomanic symptoms not seen as a problem by most;
- substance use/ETOH identified as a problem;
- “paranoid” thinking mentioned by 5 participants;
- “psychosis” not mentioned as a problem;
- bad decision-making identified as a serious problem;
- lack of work, income seen as a problem;
- sleep and circadian issues serious concern.

Most were hopeful that new medications “being developed” would help, also more “new” interventions like TMS.

RESULTS

- Patients/consumers define **clinically meaningful response** as having practical relevance to being able to function in life (NOT ADL’s), not whether the intervention completely alleviated the condition. Quality of life and getting on with life are what is important.

- When the Hamilton scales, YMRS, BPRS and so on were explained, **no participants felt that this method of measuring illness/function would be helpful** in measuring clinically meaningful effect for them as individuals – although they could see benefit for larger scale assessments (e.g. research, clinical trials).

---

RESULTS

- **Loneliness & anxiety** were major concerns for the majority of all groups, regardless of diagnosis.

- Lay people, even those with a vital stake in the outcome of research **do not understand most of the concepts underlying research** – null hypothesis, NNT, meta-studies, why placebos are used or that a placebo has a robust response.

- **Much concern about side-effects**, growing awareness of impact of shortened life-span among people living with mental illness – also, more awareness of more FDA post-approval actions, though not necessarily understanding or the significance.

BRAIN TUMOR CLINICAL ENDPOINTS WORKSHOP

OCTOBER 14, 2014

- Traditional endpoint is mortality
- Clear feedback that patients and families value cognition
- Group is working toward inclusion of cognitive endpoints in their trials
- Value of cognition is increasingly recognized
  - Brain tumors
  - Schizophrenia
  - MDD
  - Bipolar disorder
PAYER PERSPECTIVE

Robert S. Epstein, MD, MS
Chief R&D Officer, Medco
President, UBC
MINIMALLY CLINICALLY MEANINGFUL DIFFERENCE

“The smallest difference...which patients perceive as beneficial, and which would mandate, in the absence of troublesome side effects and cost, a change in the patient’s management.”

Epstein, R.S. Payers Perspective: Defining Clinically meaningful Effect for the Design and Interpretation of RCTs.
PAYER SIGNIFICANCE

“The smallest clinical difference which would mandate reimbursement for a particular technology.”

PARTICULAR CHALLENGES IN THIS THINKING

- Multiple payer perspectives – no ‘single’ answer
- May be in conflict with either statistical or clinical significance (i.e. an additional hurdle)

Epstein, Teagarden, 2012
Epstein, R.S. Payers Perspective: Defining Clinically meaningful Effect for the Design and Interpretation of RCTs.
ROLE OF BRIDGING STUDY
A FORM OF ANCHOR-BASED INTERPRETATION

▪ Correlates clinical difference/change with total direct and indirect medical care costs

▪ Defines the minimal change clinically that would be associated with meaningful cost differences
FOR EXAMPLE
RELATING COMPLIANCE TO HEALTHCARE COSTS

TOTAL HEALTHCARE COST
Per Patient per Year

ADHERENCE LEVEL (%)

<table>
<thead>
<tr>
<th>Adherence Level (%)</th>
<th>Drug Cost</th>
<th>Medical Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-9</td>
<td>$8,950</td>
<td>$7,180</td>
</tr>
<tr>
<td>20-39</td>
<td>$7,180</td>
<td>$6,600</td>
</tr>
<tr>
<td>40-59</td>
<td>$6,600</td>
<td>$6,350</td>
</tr>
<tr>
<td>60-79</td>
<td>$6,350</td>
<td>$4,660</td>
</tr>
<tr>
<td>80-100</td>
<td>$4,660</td>
<td></td>
</tr>
</tbody>
</table>

Epstein, R.S. Payers Perspective: Defining Clinically meaningful Effect for the Design and Interpretation of RCTs.
FOR EXAMPLE

ALZHEIMER’S DISEASE DEPENDENCE SCALE
RELATE DEPENDENCE SCALE TO COSTS

## ANNUAL HEALTHCARE COSTS

**BY MMSE DISEASE-SEVERITY**

<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Country</th>
<th>Type</th>
<th>Mild</th>
<th>Mod.</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>60,730</td>
<td>93,959</td>
<td>184,081</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SEK</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21-25</td>
<td>15-20</td>
<td>10-14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>93,959</td>
<td>184,081</td>
<td>226,876</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10-14</td>
<td>10-14</td>
<td>&lt;10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>226,876</td>
<td>226,876</td>
<td>374,962</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolstenholme (1998)</td>
<td>UK</td>
<td>Direct</td>
<td>&lt;20</td>
<td>15-20</td>
<td>10-14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8312</td>
<td>11,643</td>
<td>15,681</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pounds</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5040</td>
<td>7704</td>
<td>&lt;11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Euro</td>
<td></td>
<td>14,788</td>
</tr>
</tbody>
</table>

Epstein, R.S. Payers Perspective: Defining Clinically meaningful Effect for the Design and Interpretation of RCTs.
SHOULD STUDIES BE POWERED TO DETECT ECONOMIC OUTCOMES?

**IN FAVOR**
- What payers want to see
- Randomization handles baseline differences
- Goes beyond relating clinical outcome to economic value epidemiologically or in modeling

**AGAINST**
- Too much variability in costing
- Too many protocol mandated visits, etc., to tease out differences
- Setting is not “real world” anyway

---

Epstein, R.S. Payers Perspective: Defining Clinically meaningful Effect for the Design and Interpretation of RCTs.
USING THE ‘BRIDGES’

BEFORE STUDY INITIATED

- Examine statistical power to detect “X” difference in clinical measure
- Model what economic savings would translate into
  - Challenge: Is it enough? To whom?

AFTER STUDY COMPLETION

- Economic Models can be constructed
  - “X” change in clinical measure = “y” change in expected dollars
  - Can be input into:
    - Cost minimization models
    - Cost benefit Models
    - Cost effectiveness models (cost/LYS)
    - Cost utility (cost/QALY)

Epstein, R.S. Payers Perspective: Defining Clinically meaningful Effect for the Design and Interpretation of RCTs.
REGULATORY PERSPECTIVE

Thomas Laughren, MD
Director, Division of Psychiatry Products
Food and Drug Administration
EFFICACY REQUIREMENT
IN FOOD, DRUG, AND COSMETIC ACT

USC TITLE 21, SEC 505(d)

“A new drug must have “substantial evidence” that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the proposed labeling…

…substantial evidence means evidence coming from adequate and well-controlled investigations.”

NO REQUIREMENT…
• for a minimum or clinically meaningful effect size for a new drug
• that a new drug has to be better than, or even as good as, other drugs in the class
APPROACHES
TO EFFICACY DETERMINATION IN PSYCHIATRIC DRUG DEVELOPMENT

FOCUS ON RESPONDER vs NON-RESPONDER

- Compare drug and placebo on proportion of “responders”
- Need clinically meaningful measure of “response”
- Easier in some area, e.g., if mortality is the endpoint
- Not so easy in psychiatry: no clear definition of “response”
  - May rely on percentage reduction for standard rating scale (e.g., 50% reduction on HAMD total score), but no universal agreement on such arbitrary definitions
  - Could use “remission,” but few drugs would win if this were the standard
- So, not usual in psychiatric drug development to focus on “responders” as primary outcome.
- Can a certain magnitude of cognitive improvement define a responder in AD/MCI trials?

Laughren, T. FDA Perspective on Defining a Clinically Meaningful Effect for Psychiatric Drug Trials.
APPROACHES
TO EFFICACY DETERMINATION IN PSYCHIATRIC DRUG DEVELOPMENT

FOCUS ON CHANGE FROM BASELINE FOR ILLNESS SEVERITY MEASURE

- Compare drug and placebo on change from baseline to endpoint on a standard measure of illness severity (e.g., PANSS total score).
- Most common approach in psychiatric drug development programs
- Rarely try to set a standard for a minimum required “effect size”
  - Counter Example: NICE and MDD criterion
- Typically effect sizes for psychiatric drugs are quite modest, however one measures “effect size”
EFFECT SIZE
OBSERVED IN REGISTRATION TRIALS

- Effect size measure: difference between drug and placebo in change from baseline to endpoint in HAMD-17
- Estimates vary depending on trial sample and methodology
- FDA estimate for both US and Non-US sites: about 2.5 HAMD-17 units
  - E.g., in US sites, decrease of about -10.5 for drug vs -8.0 decrease for placebo
NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE
Part Of The NHS In UK

"Effect size has to be at least 3 HAMD-17 units to be considered clinically significant."

POSITIVE TRIALS
13 out of 13

AVERAGE DIFFERENCE IN RELAPSE RATE BETWEEN DRUG AND PLACEBO GROUPS

\[20\%\]
Range: 10%-33%

AVERAGE DECREASE IN RELAPSE RATE IN DRUG GROUP:

\[52\%\]
Range 28%-90%
Thank you.