Deriving Novel Outcome Measures for Clinical Trials of Prodromal Alzheimer’s Disease:

A Statistical Perspective

Steven D. Edland, PhD
Univ. of California San Diego
sedland@ucsd.edu

ISCTM, Philadelphia
October 1, 2013
Outline

I. Optimizing an outcome measure

II. Power/Relative Efficiency Primer
   A. As a function of outcome measure
   B. As a function trial designs

I. Optimizing a composite measure

I. ADNI Pilot data limitations?
I. Optimizing performance of a single outcome measure

(Item Response Theory – IRT)

(Joint Statistical Meetings, August 2012)
Objective:

• Apply IRT methods to the Activities of Daily Living (ADL) scale

• See how IRT-rescored ADL measurements perform as outcomes for clinical trials

• “Proof of Concept”
Methods:
Item Response Theory (IRT)

• Graded Response Model (GRM), Samejima, 1969

• R package ‘ltm’, function grm()

1. Estimate IRT model parameters

2. Use IRT model to calculate “latent trait” score
Results: ADL Total vs. IRT-ADL
Results: ADL Total vs. IRT-ADL

Rates of Decline: ADL Total

Rates of Decline: IRT
## Results: Sample Size – 24 Month Trial

<table>
<thead>
<tr>
<th></th>
<th>N/arm</th>
<th></th>
<th>Relative Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ADL Total</td>
<td>IRT-ADL</td>
</tr>
<tr>
<td>Training Dataset</td>
<td></td>
<td>267</td>
<td>214</td>
</tr>
<tr>
<td>Validation Dataset</td>
<td></td>
<td>291</td>
<td>228</td>
</tr>
</tbody>
</table>

(Effect size $\Delta = 25\%$ reduction in rate of decline; two-tailed $\alpha = 0.05$; Power = 0.8; LME analysis; observations every 6 months; no dropout.)
Conclusions*

• Proof of Concept ✔
  – 20% improved efficiency by IRT rescoring
  – Validated in independent sample

• Potential Issues:
  – Need LARGE sample size for training data set
  – Range of function of training set must be representative of future targeted samples

II. Power/Relative Efficiency

Power depends on:

outcome measure

and

trial design
random effects
Power formula - mixed effect model

\[ \frac{N}{\text{Arm}} = 2\sigma^2 \left( z_{1-\alpha/2} + z_{1-\beta} \right)^2 / \Delta^2 \]

where

\[ \sigma^2 = \sigma_{\beta}^2 + \frac{\sigma_{\varepsilon}^2}{\Sigma(t - t.)^2} \]

Power formula - mixed effect model

\[ \frac{N}{\text{Arm}} = 2\sigma^2 \left( z_{1-\alpha/2} + z_{1-\beta} \right)^2 / \Delta^2 \]

where

\[ \sigma^2 = \sigma^2_\beta + \frac{\sigma^2_\varepsilon}{\Sigma(t - t.)^2} \]

Variance of random slopes
N/Arm = 2\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2 / \Delta^2

where

\sigma^2 = \sigma^2_\beta + \sigma^2_\varepsilon / \Sigma (t - t_.)^2

Residual error variance
N/Arm = \(2\sigma^2 \left( z_{1-\alpha/2} + z_{1-\beta} \right)^2 / \Delta^2\)

where

\[\sigma^2 = \sigma_\beta^2 + \sigma_\varepsilon^2 / \sum(t - t.^)^2\]

Estimate using pilot data
Power formula - mixed effect model

\[ \frac{N}{\text{Arm}} = 2\sigma^2 \left( z_{1-\alpha/2} + z_{1-\beta} \right)^2 / \Delta^2 \]

where

\[ \sigma^2 = \sigma^2_{\beta} + \sigma^2_{\varepsilon} / \sum \left( t - t. \right)^2 \]

Determined by Study Design
N/Arm as a function of design (ADAS-cog, MCI, effect size = 33%)
Effect of Instrument/Outcome on Power

\[ N/\text{Arm} = \sigma^2 \left( z_{1-\alpha/2} + z_{1-\beta} \right)^2 / \Delta^2 \]

where

\[ \sigma^2 = \sigma_\beta^2 + \sigma_\varepsilon^2 / \Sigma (t - t.)^2 \]

Determined by Outcome Measure & Study Target Population
## Effect of Instrument/Outcome on Power

<table>
<thead>
<tr>
<th></th>
<th>Mean Slope</th>
<th>$\sigma_\beta$</th>
<th>$\sigma_\varepsilon$</th>
<th>N/Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instrument 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Instrument 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Instrument 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sample Size, *Prevention Trial* with Biannual Sampling, 2 or 3 Year Followup, Effect Size = 50% Reduction in Mean Slope, Power = 90%

<table>
<thead>
<tr>
<th></th>
<th>Mean Slope</th>
<th>$\sigma_\beta$</th>
<th>$\sigma_\varepsilon$</th>
<th>N/Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2Yr</td>
</tr>
<tr>
<td>Word List Delayed Rec.</td>
<td>-.17</td>
<td>0.20</td>
<td>1.27</td>
<td>1985</td>
</tr>
<tr>
<td>WMSR LM I</td>
<td>-.73</td>
<td>1.18</td>
<td>2.44</td>
<td>595</td>
</tr>
<tr>
<td>WMSR LM II</td>
<td>-.89</td>
<td>1.20</td>
<td>2.48</td>
<td>415</td>
</tr>
</tbody>
</table>
“Relative Efficiency” Defined

\[
\frac{\text{N/Arm test } #1}{\text{N/Arm test } #2} = \frac{\left[ \sigma_1^2 + \sigma_1^2/\Sigma(t - t.)^2 \right] / \beta_1^2}{\left[ \sigma_2^2 + \sigma_2^2/\Sigma(t - t.)^2 \right] / \beta_2^2}
\]
“Relative Efficiency” Defined

\[
\frac{\text{N/Arm test } #1}{\text{N/Arm test } #2} = \frac{\left[ \sigma_1 \beta^2 + \sigma_1 \epsilon^2 / \Sigma (t - t.)^2 \right] / \beta_1^2}{\left[ \sigma_2 \beta^2 + \sigma_2 \epsilon^2 / \Sigma (t - t.)^2 \right] / \beta_2^2} = \frac{(\text{MSDR}_1)^{-2}}{(\text{MSDR}_2)^{-2}}
\]
III. Optimizing performance of a composite outcome measure

American Statistical Association
FDA/Industry Statistical Workshop

September 18, 2013
Background:

• For Alzheimer Tx trials, FDA requires co-primary outcomes:
  – Functional (e.g., ADL or CDR-sob)
  – Cognitive (e.g., ADAS-cog)

• MCI/Prodromal AD trial alternatives:
  – Single primary = Functional (CDR-sob)
  – Single “composite” scale
E.g., Composite scale:

\[ w \times \text{ADAS-cog} + (1 - w) \times \text{CDR-sob} \]

- Choice of \( w \) determines power of composite
- Why not choose the optimal \( w \)?
Example 1: single primary

- gantenerumab (Hoffman-La Roche)*
  - 2 year trial
  - Prodromal disease ~ MCI
  - Phase 3
  - N = 770
  - Single primary = CDR-sob

- Experiment:
  - $\text{CDR-sob vs } w^{*}\text{CDR-sob} + (1 - w)^{*}\text{ADAS-cog}$

*see clinicaltrials.gov
Results:

• Trained on ADNI late MCI cohort, n=398

• Relative efficiency (optimal : single)
  – 12 month trial, RE = .83
  – 18 month trial, RE = .88
  – 24 month trial, RE = .91
Example 2: composite

• solanezumab (the ADCS “A4” trial)*
  – 3 year trial
  – Clinically normal, biomarker positive for AD
  – Phase 3
  – N = 1,000
  – \( w_1 \cdot \text{MMSE} + w_2 \cdot \text{Word Recall} + w_3 \cdot \text{Logical Memory} + w_4 \cdot \text{WAIS-R Digit Symbol} \)
  – \( w_i = 1/SD_{0i} \)

• Experiment:
  – “A4” weights versus optimal weights

*see slideshare.net/alzforum/arf-donohue-2013
Results:

- Trained on ADNI late MCI cohort, n=398
- Relative efficiency (optimal: “A4“ weights)
  - 36 month trial, RE = .64
Results:

• Trained on ADNI late MCI cohort*, n=398

• Relative efficiency (optimal: “A4“ weights)
  – 36 month trial, RE = .64

*Note:
- These pilot data not relevant to A4 trial
- (A relevant demonstration of optimal weighting nonetheless)
Conclusions*

• Proof of Concept ✔
  – 17% plus improved efficiency
  – Why not use optimal composite?

• Potential Issues:
  – Perhaps less heuristic than simple sum?
  – training set must be representative of future targeted samples

IV. Evidence of learning or placebo effects: Relevance to clinical trials of mild cognitive impairment (MCI)

Steven D. Edland
University of California, San Diego

Alzheimer’s Association International Conference (AAIC), Boston, July 16, 2013
Experiment 1

Compare Sample Size Estimates Informed by

ADNI (MCI cohort)

v.

ADCS (MCI clinical trial)
## Results

<table>
<thead>
<tr>
<th>Pilot data:</th>
<th>18 month change</th>
<th>(var.)</th>
<th>N/Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADNI MCI</td>
<td>3.02</td>
<td>39.7</td>
<td>1084</td>
</tr>
<tr>
<td>ADCS MCI</td>
<td>1.42</td>
<td>31.1</td>
<td>3894</td>
</tr>
</tbody>
</table>

18 month trial
\[ \Delta = 25\% \text{ slowing of progression} \]
\[ \alpha = 0.05; \beta = 0.80 \]
ADNI MCI:
3.02 pts / 18 months
ADNI MCI: 3.02 pts / 18 months

ADCS MCI: 1.42 pts / 19 months
ADNI: No learning effect

Learning/placebo effect?
Exp 2: Single blind Run-in?

- A month single-blind run-in phase may:
  - “learn up”
  - “washout” placebo effects

- Thought experiment:
  - 18 month treatment trial
  Versus
  - 15 month treatment trial with 3 month single-blind run-in phase
  - Train on ADCS MCI trial data
ADCS MCI:
1.42 pts / 19 months
ADCS MCI: 1.86 pts / 15 months
## Results

**18 month treatment:**

<table>
<thead>
<tr>
<th>Pilot data:</th>
<th>18 month change</th>
<th>(var.)</th>
<th>N/Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADCS MCI (0, 6, 12, 18)</td>
<td>1.42</td>
<td>31.1</td>
<td>3894</td>
</tr>
</tbody>
</table>

**3 month run in, 15 month treatment:**

<table>
<thead>
<tr>
<th>Pilot data:</th>
<th>15 month change</th>
<th>(var.)</th>
<th>N/Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADCS MCI (3, 6, 12, 18)</td>
<td>1.86</td>
<td>24.5</td>
<td>1783</td>
</tr>
</tbody>
</table>
Conclusions

1) ADNI MCI cohort
   - A great laboratory to explore *relative efficiency* of different designs and endpoints for clinical trials
   - Cautions however for powering trials with endpoints subject to learning / placebo effects

2) Single blind run-in phase for MCI trials (????)
   - Viable based on statistical considerations
The ADNI Experience

- Design, analysis plan, power formulas not always specified in sufficient detail
- Parameter estimates used in power calculations not always reported
- 6-fold range in projected sample size for given trial design