Novel Outcome Measures

Steven G. Potkin

University of California, Irvine
Novel Outcome Measures

• The FDA defines surrogate biomarkers as “a laboratory or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint… that replaces a distal endpoint with a more proximal one that can be measured earlier, more easily or frequently, with higher precision… as an indicator of normal biological or pathogenic processes… to develop efficient and improved understanding of drug effects, including dose and dose interval”.

• Precision: cross-lab reliability, sensitivity, and concordance among surrogates and relation to clinical disease measures
Novel outcome measures must be multi-site compatible.

- Speeds recruitment and increases generalizability
- Creates challenges in standardization and training

Examples limited to biomarkers

- Multi-site MRI – BIRN (Schizophrenia), MTA (ADHD)
- Brain and CSF markers – ADNI (AD)

Stage of illness - risk, prodromal, progression, severity
Novel outcome measures must be multi-site compatible.

**Effect size** = \( \frac{\text{mean}_1 - \text{mean}_2}{\text{pooled SD}} \)

Assume 100 subject per group 80% power alpha 0.05

We these givens measure effects size as small as .39

If pooled SD increases by 20% due to multisite variance

To get back to ES of .39 now need 148 subjects / group, a \(~50\%\) increase in sample size
BIRN Goals

• Develop the capability to analyze, as a single data set, data acquired from multiple sites.

• Develop a federated data management system to support these multi-site imaging and genetics studies.
Multi-site FMRI studies: variance issues and correction

How big a problem is it?

These are the same person’s brain in different MRI scanners across the country

From: S. Potkin, J. Turner, G. Brown
Subjects traveled around the country to be scanned at all FBIRN sites.

Unique dataset: Subject x site interactions can be measured for the first time.
## ROI – Top 10% of Activated Voxels

<table>
<thead>
<tr>
<th>Variance Source</th>
<th>Auditory</th>
<th>Hand</th>
<th>Visual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>18.8</td>
<td>18.3</td>
<td>21.8</td>
</tr>
<tr>
<td>Site</td>
<td>43.0</td>
<td>21.0</td>
<td>43.8</td>
</tr>
<tr>
<td>Day</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Run</td>
<td>0.4</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Subject X Site</td>
<td>3.6</td>
<td>14.6</td>
<td>10.5</td>
</tr>
<tr>
<td>Subject X Site+</td>
<td>20.7</td>
<td>35.2</td>
<td>20.0</td>
</tr>
<tr>
<td>Residual</td>
<td>1.5</td>
<td>4.2</td>
<td>1.5</td>
</tr>
</tbody>
</table>
Data Collection Tools

• Imaging calibration tools:
  – Stability agar phantom
  – Automated image QA

• Data formatting and metadata issues: XGEDE
Image QC Results

- SFNR Plot:
Operational Solutions

• Scanner standardization: structural and agar phantoms
• Multi-site cognitive tasks: sensitive, robust, reliable
• Physiological post-data corrections: heart rate and respiration
• Traveling Subjects
• Traveling Engineer
• Scan-day Checklist and Subject Instruction
Image QC Tool by Subject

Wiki QC Tracking Table

Cardio and Respiratory Tracking

Image QC Tracking
Impact of Calibration Methods

ANOVA Observed Effect Size

Cohen’s f

None                  Smooth               Smooth, BH      Smooth, BH
Calibrate              Calibrate, BH Screen

Remaining challenge
Working Memory in Schizophrenia

- **Sternberg task:**
  - Five: 5 6 2 8 1
  - Two: 0 9

• Five items compared to Two
DLPFC Hyperactivation / Inefficiency in Sz

Fig. 5. BOLD Signal Change in DLPFC, by Load and Hemisphere, for HV and Schizophrenic Patients. (a) Encode results by diagnostic group. (b) Probe results by diagnostic group. Error bars indicate one standard deviation.

Fig. 6. As in figure 5 but limited to the behaviorally matched sample.

Phase II results - Reliability

Average Hemodynamic Time Series Across ROI

ACG  STG  Thalamus

Visit 1 = Blue  Visit 2 = Red

Initial analyses showing Auditory Oddball reliability of BOLD signal across visits

Monetary Incentive Delay Task

Knutsen et al. 2001
Blunted NAcc activation in drug-free schizophrenics during reward anticipation


- Several lines of evidence indicate that the BOLD signal (activation of the Nacc) is generated via activation of D1 receptor in NAcc (go-pathway) by DA release during reward anticipation

Low activation of the left ventral striatum by reward cues was correlated with increased severity of negative symptoms

Juckel et al., Neuroimage, 2006
MTA Dashboard provides complete study tracking across data federation.
Multimodal Treatment study of ADHD
14-month RCT with 12 year follow-up

Inhibition (NoGo>Go)
LNCG>ADHD

GRF Cluster Thresholded
p>.05
z=2.3 4.0

Behavior
ADHD FPs>LNCG FPs

p(ADHD>LNCG)<.05

Mean FPs per Run

NoGo-Go Task
Standardisation of hippocampal volumetry
Developing a global gold standard for Alzheimer's research
“One shot” hippocampal atrophy for enrichment
EMA qualification opinion

17 November 2011
EMA/CHMP/SAWP/809208/2011
Committee for Medicinal Products for Human Use (CHMP)
Qualification opinion of low hippocampal volume (atrophy) by MRI for use in clinical trials for regulatory purpose - in pre-dementia stage of Alzheimer’s disease

16 February 2012
EMA/CHMP/SAWP/892998/2011
Committee for Medicinal Products for Human Use (CHMP)
Qualification opinion of Alzheimer’s disease novel methodologies/biomarkers for PET amyloid imaging (positive/negative) as a biomarker for enrichment, for use in regulatory clinical trials in predementia Alzheimer’s disease

16 February 2012
EMA/CHMP/SAWP/893622/2011
Committee for Medicinal Products for Human Use (CHMP)
Qualification opinion of Alzheimer’s disease novel methodologies/biomarkers for the use of CSF AB 1-42 and t-tau and/or PET-amyloid imaging (positive/negative) as biomarkers for enrichment, for use in regulatory clinical trials in mild and moderate Alzheimer’s disease
Which Good Clinical Practice and specific highest international standards?

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Med border</th>
<th>Lat border</th>
<th>Inf border</th>
<th>Norm. hippo vol (cm³)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Left</td>
</tr>
<tr>
<td>Watson et al.</td>
<td>Mesial edge of temporal lobe</td>
<td>Temp horn of lat ventr</td>
<td>Incl subicular complex &amp; uncal cleft w/ border separating subicular complex</td>
<td>4.903</td>
</tr>
<tr>
<td>Zipursky et al.</td>
<td>Regional outline at choroidal fissure</td>
<td>Not mentioned</td>
<td>hippocampal tissue and parahippocampal gyrus white matter</td>
<td>1.990</td>
</tr>
</tbody>
</table>

2.5-fold difference

Harmonized Protocol
Results

Figure 2. Stability of local and harmonized protocols among 14 naïve tracers. All comparisons between local and harmonized are significant at $p<0.015$.

<table>
<thead>
<tr>
<th></th>
<th>Local</th>
<th>Harmonized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L consistency</td>
<td>0.86</td>
<td>0.97</td>
</tr>
<tr>
<td>R consistency</td>
<td>0.85</td>
<td>0.97</td>
</tr>
<tr>
<td>L absolute</td>
<td>0.44</td>
<td>0.88</td>
</tr>
<tr>
<td>R absolute</td>
<td>0.43</td>
<td>0.89</td>
</tr>
</tbody>
</table>

95% C.I.

<table>
<thead>
<tr>
<th></th>
<th>sup.</th>
<th>inf.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>0.78</td>
<td>0.93</td>
</tr>
<tr>
<td>R</td>
<td>0.76</td>
<td>0.93</td>
</tr>
<tr>
<td>consistency</td>
<td>0.23</td>
<td>0.67</td>
</tr>
<tr>
<td>absolute</td>
<td>0.22</td>
<td>0.66</td>
</tr>
</tbody>
</table>

EADC-ADNI Hippocampal Harmonization Working Group, In preparation

Courtesy of P Bellac
Kaplan–Meier time to conversion to AD survival curves for ADNI subjects who had a diagnosis of mild cognitive impairment at their baseline visit. The small vertical lines are, the survival curves are shown for MCI subjects with CSF Aβ_{1-42} concentrations above or below the threshold value of 192 pg/mL at their baseline. In b, CSF t-tau/Aβ_{1-42} ratio values above or below the threshold value of 0.39

Novel Outcome Measures

Original Hypothetical AD Model:
Jack et al, 2010
Original Hypothetical AD Model: Jack et al, 2010

Revised AD Model: Jack et al, 2013
Dynamic Biomarker Changes in ADNI

Hypothetical longitudinal changes of biomarkers for a 75-year-old person at different disease stages

Biomarkers vs time (months) based on longitudinal data

Lo et al Archives of Neurology 2011
%CV values for CSF $\text{Aβ}_{1-42}$, t-tau & p-tau$_{181}$

Bar plots for the total, between-center and within-center %CV values derived for each CSF pool for a $\text{Aβ}_{1-42}$, b t-tau and c p-tau$_{181}$

Shaw et al Acta Neuropathol 2011
Reliability of Blinded CSF samples across 84 laboratories

Alzheimer’s Association international QC program

Kang et al Clin Chem 2013
Reliability of Blinded CSF samples across 84 laboratories

Alzheimer’s Association international QC program

![Graph showing contribution to variation for Aβ42, T-tau, and P-tau in ELISA INNOTEST](image)

Kang et al Clin Chem 2013
Regression plots of concentrations, measured in never previously thawed CSF aliquots from 118 ADNI subjects, utilizing 2–3 subjects randomly selected from each of 38 analytical runs. For each randomly selected subject, a second never previously thawed aliquot was included in the run following analysis of the first never previously thawed aliquot. In plots d–f, the % difference between the test and retest values are plotted versus the average value for each test/retest pair of concentrations. The shaded area around each linear regression line is the 95% confidence interval for the regression line. In plots d–f, the dotted lines are the 95% confidence intervals for the mean difference lines (solid lines).
Agreement between florbetapir and CSF Aβ

Both positive
Both negative
Florbetapir cortical retention ratio
CSF Aβ

1-42

Normal
EMCI
LMCI
AD

Landau et al, Annals of Neurology In press
Disagreement between florbetapir and CSF Aβ

Florbetapir cortical retention ratio

CSF Aβ1-42

Florbetapir+ / CSF –
N=13
(11 EMCI, 2 NC)
6/13 ApoE4+

Florbetapir- / CSF+
N=7
(1 NC, 1 EMCI, 4 LMCI, 1 AD)
0/7 ApoE4+

Landau et al, Annals of Neurology In press
Disagreement between florbetapir and CSF $\text{A}\beta$

Landau et al, Annals of Neurology In press
Relatively Poor Agreement Between Florbetapir and Tau

\[ \kappa = .42 \]

Landau et al, Annals of Neurology In press
CSF change by florbetapir status

Total N=94

Florbetapir -

Florbetapir +

Normal | EMCI | LMCI | AD

CSF Aβ+

CSF Aβ-

Time relative to florbetapir scan

-8 -6 -4 -2 0 2 -8 -6 -4 -2 0 2 -8 -6 -4 -2 0 2
Florbetapir and FDG in ADNI (N = 910)

r = -0.372
p < 0.0001

Discordant
AD 16/146 (11%)
LMCI 61/212 (29%)
EMCI 92/289 (32%)
Normal 51/263 (19%)

Concordant
AD 109/146 (74%)
LMCI 73/212 (34%)
EMCI 43/289 (15%)
Normal 23/263 (9%)

Courtesy of Bill Jagust
Original Hypothetical AD Model: Jack et al, 2010

Revised AD Model: Jack et al, 2013
Conclusion Biomarker Outcomes

Multi-site collection of surrogate biomarkers is required for a timely sufficient n, generalizability, and regulatory requirements.

Standardization of collection protocol, paradigm details, quantitative measures, and dynamic automatic QC is necessary. Don’t miss these details.

Training and ongoing monitoring for research procedures and personnel at all levels is required to prevent “garbage in, garbage out”.