Neuroimaging enrichment biomarkers for CNS diseases

Adam Schwarz
(Eli Lilly and Company)
On behalf of CAMD imaging qualification team
Special thanks to Peng Yu and Derek Hill
Outline

♦ Hippocampal volume (HV) in AD (case study of an enrichment biomarker)

♦ Overview of evidentiary considerations for biomarkers
  • General considerations
  • Mapping to HV and context of use for trial enrichment

♦ NIA-AA recommendations for clinical research in MCI due to AD

♦ Performance characteristics of HV in MCI
  • Heterogeneity of clinically-defined MCI population (differential clinical progression)
  • Supporting data from the literature
  • Test-retest
  • Sensitivity to different HV algorithms
  • Operational considerations
Hippocampal atrophy in Alzheimer’s Disease

AD = Alzheimer’s Disease. MCI = Mild Cognitive Impairment.
Brain atrophy as measured by structural MRI reflects neuropathology of AD

**Disease stage**

<table>
<thead>
<tr>
<th>Disease state</th>
<th>Volume (cc)</th>
<th>Annual % Change (APC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>aMCI</td>
<td>2</td>
<td>1.0</td>
</tr>
<tr>
<td>pAD</td>
<td>3</td>
<td>2.0</td>
</tr>
</tbody>
</table>

**Neurodegeneration**

![Graph showing neurodegeneration](image)

**Post-mortem Braak stage**

![Graph showing post-mortem Braak stage](image)

**Cognitive function**

![Graph showing cognitive function](image)
Biomarker development adapted from the framework of Pepe et al. 2001

Phase 1: Rationale for the use of BM

PA Potential leads

Phase 2: Discrimination ability of the BM

PA Identify discrimination accuracy AD/HC

Phase 3: Detection ability in early phase

PA1 Assess capacity of earliest (MCI) detection

Phase 4: BM accuracy in real world patients

PA Assess true/false referral rate in BM-diagnosed patients

Phase 5: Quantify the impact of BM-based diagnosis on relevant outcomes

PA Estimate impact on morbidity & disability

Context of use (clinical trial enrichment – MCI)

SA1 Assay definition

SA2 Ante mortem/autopsy

SA3 Covariates in HC

SA4 Covariates in AD

SA1 Detect predictive features

SA2 Practical feasibility

SA3 Compare ≠ protocols

SA4 Cost/benefit quantification

SA1 Estimate impact & costs

SA2 Compliance in ≠ settings

SA3 Estimate impact & costs

SA4 Monitor false negatives

Slide courtesy of Marina Boccardi & Giovanni Frisoni
A Prototypical Process for Creating Evidentiary Standards for Biomarkers and Diagnostics

CA Altar¹, D Amakye², D Bounos³, J Bloom⁴, G Clack⁵, R Dean⁴, V Devanarayan⁶, D Fu⁷, S Furlong⁵, L Hinman⁸, C Girman⁹, C Lathia¹⁰, L Lesko¹¹, S Madani¹², J Mayne¹³, J Meyer⁸, D Raunig¹², P Sager⁵, SA Williams¹⁴, P Wong⁸ and K Zerba¹⁵

A framework for developing evidentiary standards for qualification of biomarkers is a key need identified in the Food and Drug Administration’s Critical Path Initiative.¹ This article describes a systematic framework that was developed by Pharmaceutical Research and Manufacturers of America (PhRMA) committees and tested at a workshop in collaboration with the Food and Drug Administration and academia. With some necessary refinements, this could be applied to create an appropriately individualized evidentiary standard for any biomarker purpose.
Evidence from many studies (meta-analysis). Explicit replication part of proposed HCV analysis plan.

Standardized methods of acquisition and analysis commonly applied. 510(k)/CE-marked analysis software available. Hippocampal harmonization.

No real benchmark. Performs similarly to alternatives.

**Canonical feature of AD. Causally related to core amnestic phenotype.**

**N/A**

*non-chemical marker*

**N/A**

*outside Context of Use*

---

**Table 1:** Prototype “evidence map” — categorical description of different types of scientific evidence potentially relevant to biomarker qualification; subcategorical graded weight of evidence from least to most.

<table>
<thead>
<tr>
<th>Evidence type</th>
<th>Grade D</th>
<th>Grade D+/C−</th>
<th>Grade C</th>
<th>Grade C+/B−</th>
<th>Grade B</th>
<th>Grade B+/A−</th>
<th>Grade A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory on biological plausibility</td>
<td>Observed association only</td>
<td>Theory, indirect evidence of relevance of the biomarker from animals</td>
<td>As for lower grade but evidence is direct</td>
<td>Theory, indirect evidence in humans, non-causal pathway possible</td>
<td>As for lower grade, but biomarker on causal path</td>
<td>Human evidence based mathematical model of biology showing biomarker is on causal pathway</td>
<td></td>
</tr>
<tr>
<td>Interaction with pharmacologic target</td>
<td>Biomarker identifies target in <em>in vitro</em> binding</td>
<td>Biomarker identifies target in <em>in vivo</em> binding in animals</td>
<td>Biomarker identifies target in <em>in vivo</em> studies or from human tissue, no truth standard</td>
<td>Biomarker identifies target in <em>in vivo</em> studies or from human tissue, no truth standard</td>
<td>Biomarker identifies target in <em>in vivo</em> studies or from human tissue, with accepted truth standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacologic mechanistic response</td>
<td><em>In vitro</em> evidence that the drug affects the biomarker</td>
<td><em>In vitro</em> evidence that multiple members of this drug class affects the biomarker</td>
<td><em>In vitro</em> evidence that this drug affects biomarker in animals</td>
<td>As for lower grade but effect shown across drug class</td>
<td>Human evidence that this drug affects the biomarker OR animal evidence of specificity</td>
<td>Human evidence that multiple members of this drug class affect the biomarker and the effect is specific to this class/mechanism</td>
<td></td>
</tr>
<tr>
<td>Linkage to clinical outcome of a disease or toxicity</td>
<td>Biomarker epidemiologically associated with outcome without any intervention</td>
<td>Biomarker associated with change in outcome from intervention in another drug class</td>
<td>As for lower grade but in this drug class</td>
<td>As for lower grade but multiple drug classes albeit inconsistent or a minority of disease effect</td>
<td>As for lower grade but consistent linkage and explains majority of disease effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics replication, confirmation</td>
<td>Algorithm was developed from a different dataset and applied here prospectively</td>
<td>Algorithm was developed from a different dataset and applied here prospectively</td>
<td>Major sources or variation known and controlled to be less than biological signal; standardization methods applied</td>
<td>Algorithm developed from different dataset, replicated prospectively in other sets and applied prospectively here</td>
<td>All major sources of technical imprecision are known, and controlled test/assay accuracy is defined against standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy and precision (analytic validation)</td>
<td>Sources of technical variation are unknown but steps are taken to ensure consistent test application</td>
<td>Sources of technical variation are unknown but steps are taken to ensure consistent test application</td>
<td>Sources of technical variation are unknown but steps are taken to ensure consistent test application</td>
<td>Sources of technical variation are unknown but steps are taken to ensure consistent test application</td>
<td>Sources of technical variation are unknown but steps are taken to ensure consistent test application</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Not all types of evidence required all seven grades to be completed.
Biomarkers of neurodegeneration are embedded in the 2011 NIA-AA research criteria for MCI due to AD

The diagnosis of mild cognitive impairment due to Alzheimer’s disease: Recommendations from the National Institute on Aging and Alzheimer’s Association workgroup

Marilyn S. Albert\textsuperscript{a,\textordmasculine}, Steven T. DeKosky\textsuperscript{b,\textordmasculine}, Dennis Dickson\textsuperscript{d}, Bruno Dubois\textsuperscript{e}, Howard H. Feldman\textsuperscript{f}, Nick C. Fox\textsuperscript{g}, Anthony Gamst\textsuperscript{h}, David M. Holtzman\textsuperscript{i,j}, William J. Jagust\textsuperscript{k}, Ronald C. Petersen\textsuperscript{l}, Peter J. Snyder\textsuperscript{m,n}, Maria C. Carrillo\textsuperscript{o}, Bill Thies\textsuperscript{o}, Creighton H. Phelps\textsuperscript{p}

Table 3
MCI criteria incorporating biomarkers

<table>
<thead>
<tr>
<th>Diagnostic category</th>
<th>Biomarker probability of AD etiology</th>
<th>Aβ (PET or CSF)</th>
<th>Neuronal injury (tau, FDG, sMRI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCI–core clinical criteria</td>
<td>Uninformative</td>
<td>Conflicting/indeterminant/untested</td>
<td>Conflicting/indeterminant/untested</td>
</tr>
<tr>
<td>MCI due to AD—intermediate likelihood</td>
<td>Intermediate</td>
<td>Positive</td>
<td>Untested</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Untested</td>
<td>Positive</td>
</tr>
<tr>
<td>MCI due to AD—high likelihood</td>
<td>Highest</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>MCI—unlikely due to AD</td>
<td>Lowest</td>
<td>Negative</td>
<td>Negative</td>
</tr>
</tbody>
</table>

Abbreviations: AD, Alzheimer's disease; Aβ, amyloid beta peptide; PET, positron emission tomography; CSF, cerebrospinal fluid; FDG, fluorodeoxyglucose; sMRI, structural magnetic resonance imaging.
A systematic survey of the published literature indicated strong evidence for low hippocampal volume as an enrichment biomarker in MCI.
**De novo** calculations confirmed literature findings and robustness to HCV measurement algorithm

### Table 1
Results of Coalition Against Major Diseases’ *de novo* analysis. The AUC for four different hippocampal volume quantification algorithms applied to ADNI-I data indicate the prediction by MRI hippocampal volume of clinical conversion to Alzheimer's dementia within two years.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Training, n</th>
<th>Testing, n</th>
<th>AUC based on clinical conversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEAP</td>
<td>149</td>
<td>173</td>
<td>0.7565</td>
</tr>
<tr>
<td>NeuroQuant</td>
<td>149</td>
<td>173</td>
<td>0.7516</td>
</tr>
<tr>
<td>FreeSurfer</td>
<td>148</td>
<td>171</td>
<td>0.7536</td>
</tr>
<tr>
<td>HMAPS</td>
<td>128</td>
<td>161</td>
<td>0.7290</td>
</tr>
</tbody>
</table>

Abbreviations: AUC, area under the receiver–operating characteristic curves; LEAP, Learning Embeddings for Atlas Propagation; HMAPS, Hippocampus Multi-Atlas Propagation and Segmentation.

### Table 2
AUC values reported in the Coalition Against Major Diseases literature review

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>AUC based on clinical conversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bakkour et al.</td>
<td>49</td>
<td>0.65</td>
</tr>
<tr>
<td>Devanand et al.</td>
<td>139</td>
<td>0.77</td>
</tr>
<tr>
<td>Fleisher et al.</td>
<td>129</td>
<td>0.60</td>
</tr>
<tr>
<td>Galluzzi et al.</td>
<td>90</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Abbreviation: AUC, area under the receiver–operating characteristic curves.
Robustness of automated hippocampal volumetry across magnetic resonance field strengths and repeat images

Robin Wolz\textsuperscript{a, b}, Adam J. Schwarz\textsuperscript{c}, Peng Yu\textsuperscript{c}, Patricia E. Cole\textsuperscript{c}, Daniel Rueckert\textsuperscript{b}, Clifford R. Jack, Jr.,\textsuperscript{d} David Raunig\textsuperscript{e}, Derek Hill\textsuperscript{a,}\textsuperscript{*}, for The Alzheimer’s Disease Neuroimaging Initiative
Hippocampal volume measurements are highly reliable (test-retest)

Wolz R et al. (2014) Alzheimers & Dementia 10 430
Operational considerations and practical implications for trials

Operationalizing hippocampal volume as an enrichment biomarker for amnestic mild cognitive impairment trials: effect of algorithm, test-retest variability, and cut point on trial cost, duration, and sample size

Peng Yu, Jia Sun, Robin Wolz, Diane Stephenson, James Brewer, Nick C. Fox, Patricia E. Cole, Clifford R. Jack Jr, Derek L.G. Hill, Adam J. Schwarz, for the Coalition Against Major Diseases and the Alzheimer’s Disease Neuroimaging Initiative
Cut-point defined with respect to normative reference range

Specify aHCV cut-point based on normative reference population

Enroll only aMCI subjects with aHCV < cut-point
MCI subjects with smaller hippocampi progress more rapidly.

Enriched population (HV < 25% of normal)

All MCI subjects

Subjects excluded (HV >= 25% of normal)

Subjects with smaller HV at baseline progress more rapidly

Slower progressing subjects are excluded.
Cut-point defined with respect to normative reference range

How do the enriched trial characteristics depend on the choice of cut-point?
MCI subject selection based on low hippocampal volume results in smaller sample sizes

This improvement is not sensitive to algorithm and is maintained across a range of cut-points.
Enriched population yields smaller sample size but increased screen fail rate → implications for clinical trial operations

Enriched population yields smaller sample size but increased screen fail rate → implications for clinical trial operations

An operational recipe for the use of HCV to enrich clinical trials

**Decisions relating to trial**

**Trial MRI methodology**
- Select and standardize MRI acquisition methodology (e.g., adhering to the ADNI standard).
- Select the image QC and postprocessing methods.
- Decide which algorithm will be used to calculate HCV.
- Decide with method will be used to calculate ICV.

---

**Reference data set and cut-point**

- Process the reference vMRI scans using the same post processing methodology to be used in the trial.
- Calculate HCV values using the same algorithm to be used in the trial.
- Calculate ICV values using the same method as to be used in the trial.
- Calculate aHCV values, accounting for covariates such as age and ICV, to derive a reference distribution of aHCV values.
- Derive the aHCV cut point value to be used as an inclusion criterion.

---

**Reference data set and decision rule for inclusion**

- Select the normative reference MRI data set (e.g., ADNI healthy control subjects) from which the inclusion criterion will be defined. (The acquisition methodology must match that to be used in the trial.)
- Select a cut point for patient inclusion based on the normative reference distribution of adjusted HCVs (e.g., 10th percentile).

---

**Implementation in clinical trial**

- For each patient with MCI, calculate the aHCV from the screening MRI images.
- If the adjusted HCV is less than the selected aHCV cut point, the patient is included in the trial or proceeds in the screening cascade.
Different progression rates from person to person, and the field’s inability to predict with any precision how quickly a given person will progress, are longstanding problems in Alzheimer’s disease trials. In this instance, the fast progressors—i.e., those whose hippocampal volume and CDR-SB performance declined the most over the duration of the trial—appeared to benefit […]

Summary

♦ Evidentiary considerations and research guidelines relevant to the context of use were reviewed.

♦ Key evidentiary questions to be addressed by a putative biomarker include:
  • Heterogeneity of the clinically-defined target population
  • Strength of supporting data and robustness of findings across different studies, cohorts, geographies
  • Test-retest of the method *per se*
  • Sensitivity to technical variations
  • Operational considerations (including time and cost)

♦ Hippocampal volume (HV) provides a case study of a neuroimaging enrichment biomarker for MCI due to AD, for which the above points have been addressed.

♦ Biomarker qualification could improve chance of success, reduce number of subjects exposed to an experimental treatment that may have side effects, and reduce time/cost of trials.
• Backup
- 10\textsuperscript{th} percentile \sim 1.3 \text{ SD below normal mean}
- 25\textsuperscript{th} percentile \sim 0.6 \text{ SD below normal mean}
- 40\textsuperscript{th} percentile \sim 0.2 \text{ SD below normal mean}