Preclinical Alzheimer's Cognitive Composite: Optimization & Validation

Michael Donohue, PhD

Alzheimer's Therapeutic Research Institute
University of Southern California

July 23, 2016

Preclinical Alzheimer's Cognitive Composite (PACC)

- Theory/literature driven endpoint conceived for preclinical AD clinical trials (Donohue, et al. 2014)
- Four components covering dimensions of early decline:
  1. MMSE (global; orientation to time and place)
  2. FCSRT (semantic memory)
  3. Logical Memory (episodic memory)
  4. Digit Symbol Substitution (executive function)
- PACC is correlated with self-assessment of function (Amariglio, et al. 2015)
- Primary outcome for A4 (Solanezumab, Eli Lilly) & EARLY (BACEi, Janssen)
- Similar composites are proposed for other preclinical AD studies

Preclinical Alzheimer's Cognitive Composite (PACC)

- Criticisms include:
    1. MMSE is near ceiling, and should be dropped
    2. A data-driven machine learning approach should be used to select components
    3. Weights should be optimized to increase power to detect treatment effects (or reduce sample size)
- Motivation: Explore out-of-sample performance of "optimized" versions of PACC

Standardization & Weighting

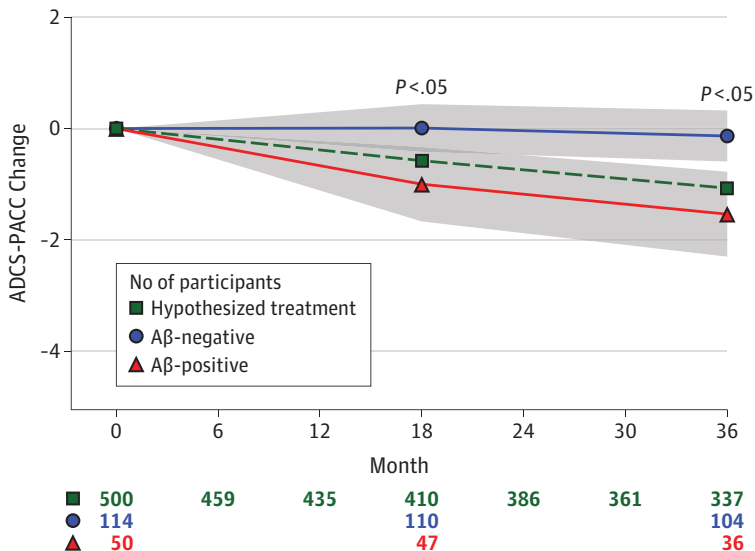Each component change score is *standardized* relative to baseline SD, to yield z-scores:

$$z_{jt} = \frac{(y_{jt} - y_{j0})}{\sigma_{j0}},$$

for component j at time t, where σ_{j0} is standard deviation of y_{j0}.

We consider *weighted sum* composites:

$$Y_t(\mathbf{w}) = z_{1t}w_1 + z_{2t}w_2 + z_{3t}w_3 + z_{4t}w_4,$$

where $\mathbf{w} = (w_1, w_2, w_3, w_4)$ is the *weight vector*
(each $w_k > 0$ and they sum to one)

Aβ group profiles and the smallest detectable effect, δ, based on AIBL with MMRM assuming 80% power, 5% two-sided α, 3 year trial, n=500 per group, and 30% attrition. (Donohue, 2014)

Optimization of $\mathbf{w}$

We can "optimize" $\mathbf{w}$ according to any *objective function*.

We explore:

1. Minimize minimum detectable δ as a percentage of $A\beta$ group difference
2. Logistic regression weights from a model to discriminate $A\beta^+$ from $A\beta^-$

Note:

1. Optimization comes at the price of simplicity & face validity
2. Available natural history data provide no info regarding treatment effects on components

Why validate?

Testing the procedure on the data that gave it birth is <u>almost certain to overestimate performance</u>, for the optimizing process . . . will have made the greatest use possible of any and all idiosyncrasies of those particular data. . . As a result, the procedure will likely work better for these data than for almost any other data that will arise in practice.

Mosteller & Tukey (1977). Data analysis and regression: a second course in statistics. p. 37

"External" validation

| A4 & EARLY | AIBL ($\hat{\mathbf{w}}$) | NA-ADNI | J-ADNI | ADCS-PI | |
PET/CSF	PET	PET/CSF		ApoEε4	CDR-G
MMSE	MMSE (6%)	MMSE		3MSE	
FCSRT	CVLT (55%)	ADAS-COG		FCSRT	
LM	LM (35%)	LM		NYU	
Digit	Digit (5%)	Digit		Digit	
δ (equal $\hat{\mathbf{w}}$)	33%	42% (year 2)	35%	48%	14%
δ (logistic $\hat{\mathbf{w}}$)	27%*	†	54%	95%	15%

* The minimum possible δ was 25%, but this required weighting
Digit Symbol in the wrong direction.

† The AIBL-optimized PACC was not significantly different at any visit in
ADNI, while the original was significant only at year 2.

5×3-fold cross-validation
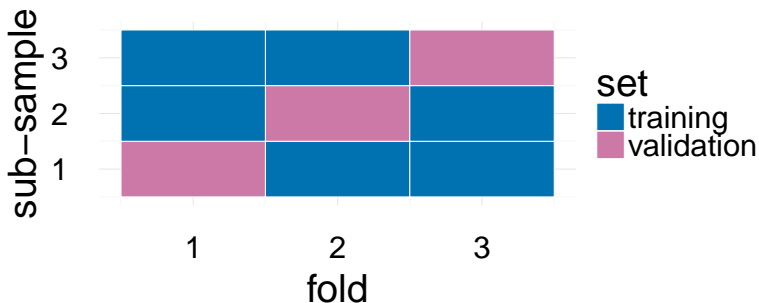
We split the sample up into 3 non-overlapping sub-samples of size $n/3$ which take turns as the validation set (red), leaving the remaining sample (blue) for training set. In our case, weights are optimized on each training set and out-of-sample perfomance (e.g. δ) is assessed on validation sets. This was repeated 5 times with random permutations of the data.

Median/range of optimized weights across 15 training sets
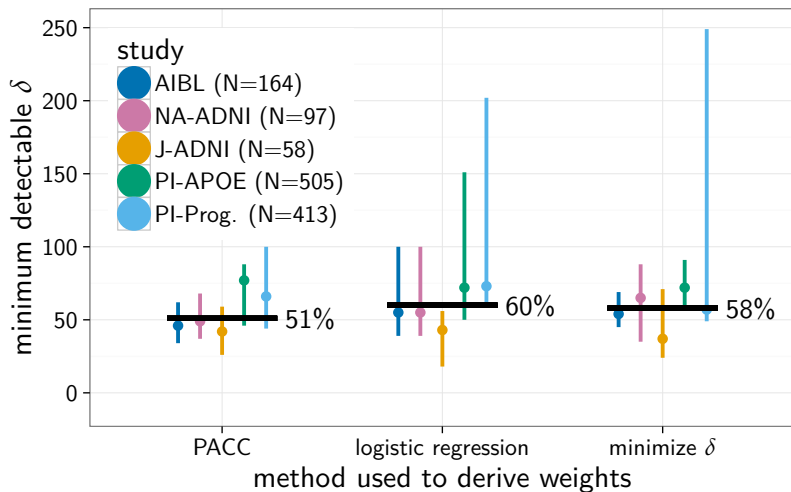

Weights derived by logistic regression
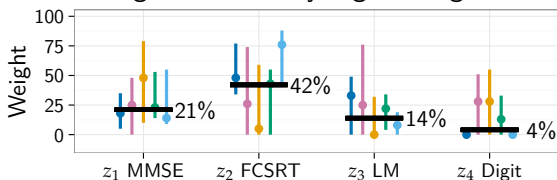
Weights derived by minimum δ
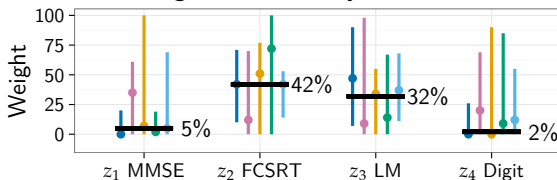
study
- AIBL (N=164)
- NA-ADNI (N=97)
- J-ADNI (N=58)
- PI-APOE (N=505)
- PI-Prog. (N=413)

Median/range of minimum δ across 15 validation sets

Limitations of validation

Limitations of "external" validation:

- expensive to collect new data
- existing data is never ideally matched
  (populations or components)

Limitations of cross-validation:

- mere simulation of real-world replication
- sub-samples may not be sufficient size for
  training and/or validation steps

Limitations of optimization

- If sample size is insufficient for cross-validation, then it is insufficient for optimization
- Natural history data provide no info regarding treatment effects on components
- Optimization comes at the price of simplicity & face validity
- Optimization should only be considered if there is a convincing rationale and it can be validated.

Summary

- Both MMSE and Digit Symbol were consistently down-weighted by optimization, however down-weighting did not reliably improve composite performance.

- MMSE has good face-validity as a global assessment and has demonstrated sensitivity to preclinical decline (Amieva, et al 2008).

- Digit Symbol has good face validity as a measure of executive function.

- Component weight optimization does not (yet?) yield reliable improvements in power to detect treatment effects in preclinical AD clinical trials.

Acknowledgements

- Chung-Kai Sun, Rema Raman, Philip Insel, Niklas Mattson, Beth Mormino, Reisa Sperling, Paul Aisen
- Study teams and participants of the ADCS-PI, NA-ADNI, J-ADNI, and AIBL studies.

Poster P3-034 on Tuesday