Supplementary Online Content

Donohue MC, Sperling RA, Salmon DP, et al; Australian Imaging, Biomarkers, and Lifestyle Flagship Study of Ageing; the Alzheimer's Disease Neuroimaging Initiative; and the Alzheimer's Disease Cooperative Study. The Preclinical Alzheimer Cognitive Composite: measuring amyloid-related decline. *JAMA Neurol*. Published online June 2, 2014. doi:10.1001/jamaneurol.2014.803.

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix

CDR-SB and ADCS-PACC items

Based on ADNI Aβ group data, we would have 80 to 90% power to detect Δ=0.328 to 0.380 Delayed Word Recall points (110 to 127% Aβ difference), Δ=0.789 to 0.912 Logical Memory IIa points (140 to 162% Aβ difference), Δ=1.44 to 1.66 Digit-Symbol Substitution points (66.2 to 76.3% of the month 24 Aβ difference), Δ=0.312 to 0.361 MMSE points (38.1 to 44.0% of the month 24 Aβ difference), and Δ=0.151 to 0.174 CDR-SB points (45.1 to 52.3% Aβ difference).

Based on AIBL Aβ group data, we would have 80 to 90% power to detect Δ=0.485 to 0.561 List A Delayed Recall points (29.5 to 34.1% Aβ difference), Δ=0.673 to 0.778 Logical Memory IIa points (30.3 to 35.0% Aβ difference), Δ=1.67 to 1.93 Digit-Symbol Substitution points (392 to 454% Aβ difference), Δ=0.267 to 0.309 MMSE points (85.7 to 99.1% Aβ difference). CDR-SB was not available.

Based on ADCS-PI CDR-G Progressor group data, we would have 80 to 90% power to detect Δ=1.12 to 1.30 FCSRT Total Recall points (14.4 to 16.7% CDR-G difference), Δ=0.618 to 0.713 Logical Memory IIa points (14.7 to 17.0% CDR-G difference), Δ=2.20 to 2.54 Digit-Symbol Substitution points (19.5 to 22.6% CDR-G difference), Δ=1.80 to 2.08 3MSE points (28.4 to 32.8% CDR-G difference). CDR-SB was only collected in ADCS-PI after decline was observed at annual visits.

Based on ADCS-PI *APOE*-ε4 group data, we would have 80 to 90% power to detect Δ=1.23 to 1.42 FCSRT Total Recall points (127 to 147% *APOE*-ε4 difference), Δ=0.647 to 0.748 Logical Memory IIa points (52.9 to 61.2% *APOE*-ε4 difference), Δ=2.31 to 2.67 Digit-Symbol Substitution points (134 to 155% Aβ difference), Δ=2.05 to 2.37 3MSE points (81.0 to 93.5% *APOE*-ε4 difference).

Reweighting the ADCS-PACC items with logistic regression

To explore re-weighting the items of the ADCS-PACC to improve power, we fit a logistic model of AIBL Aβ status at month 36 with ADCS-PACC item change Z-scores as covariates. The regression coefficients from this model provide a weighting tuned to predict Aβ status. The resulting weights favored List A Delayed Recall and Logical Memory (55.0% List A Delayed Recall, 34.7% Logical Memory, 5.7% MMSE, and 4.6% Digit Symbol Substitution).

We see very variable results when the AIBL-derived logistic regression weights are applied across the studies. With the equal (vs. logistic regression) weighted ADCS-PACC we have 80% power to detect 33.3% (vs. 26.8%) of the AIBL Aβ group difference at month 36, 14.4% (vs. 14.9%) of the ADCS-PI CDR-G Progressor difference, and 47.9% (vs. 94.7%) of the ADCS-PI *APOE*-ε4 difference. In ADNI, the Aβ group difference in the logistics regression re-weighted ADCS-PACC is not significant at any time point. We were able to further optimize the weighting to minimize the AIBL Aβ group difference from 26.8% down to 25.3% using Nelder-Mead[1] optimization, but the solution entailed weighting Digit Symbol Substitution in the wrong direction (47.7% List A Delayed Recall, 54.1% Logical Memory, 5.8% MMSE, and -7.7% Digit Symbol Substitution).

Item Response Theory analysis

To further explore re-weighting, we applied an Item Response Theory (IRT) approach[2] to all of the items of the ADNI neurological assessment battery. We trained the IRT model using data from N=322 ADNI normal control subjects with unknown Aβ status, then tested the IRT-derived latent ability on the N=97 ADNI normal control subject with known Aβ status. We derived two IRT measures of latent ability (1) based on the ADCS-PACC items alone and (2) based on the top 16 items from the complete ADNI battery (using a total information score threshold of 2.5). The top 16 items identified by this method were:

CDR-Global, ADAS-Cog Constructional Praxis, ADAS-Cog Number Cancellation, CDR-Memory, CDR-Judgment, Clock Drawing, Clock Copying, Auditory Verbal Learning Test (AVLT) 30 Minute Delay, Functional Activities Questionnaire (TV program, book, or magazine), ADAS-Cog Naming Objects and Fingers, AVLT Trial 6 Total, Digit Span Forward, ADAS-Cog Delayed Word Recall, Logical Memory Total Story Units Recalled, AVLT Total Intrusions, ADAS-Cog Word-finding Difficulty. The eFigure shows the MMRM estimated mean change from baseline for the two IRT latent ability outcomes and the eTable provides the estimates.

The 16-item latent ability score has 80 to 90% power to detect a treatment effects as small as Δ=0.435 to 0.503 points (43.5 to 50.3% of the Aβ group difference at month 36) assuming 5% two-sided alpha, and 30% dropout. The ADCS-PACC latent ability score has 80 to 90% power to detect effects as small as Δ=0.262 to 0.303 points (46.0 to 53.3% of the Aβ group difference at month 24).

We see little or no improvement over the equal weight version of the ADCS-PACC with respect to discrimination and power. With the equal (vs. IRT) weighted version of the ADCS-PACC we have 80% power to detect 42.4% (vs. 46.0%) of the ADNI Aβ group difference at month 24. The IRT re-weighting does marginally worse. The 16 item latent ability score has 80% power to detect 43.5% of the Aβ group difference at *month 36* and demonstrates a more consistent widening of the gap between ADNI Aβ groups relative to the ADCS-PACC.

eTable. MMRM estimates of IRT latent ability change from baseline for ADNI Aβ+ (n=36) vs Aβ- (n=59) normal controls.

IRT latent ability from top 16 items

Month	Group	N	Estimate	SE	p-value	p-value*	lower	upper	sigma	rho
6	Aβ–	59	-0.249	0.159	0.117		-0.560	0.062		
	Aβ+	36	-0.257	0.208	0.219		-0.665	0.152		
	Difference		-0.008	0.265	0.977	1.000	-0.527	0.511	1.213	
12	Aβ–	58	-0.167	0.188	0.376		-0.535	0.202		
	Aβ+	35	-0.347	0.247	0.161		-0.830	0.137		
	Difference		-0.180	0.313	0.564	0.950	-0.793	0.432	1.430	
24	Aβ–	53	-0.296	0.219	0.177		-0.724	0.132		
	Aβ+	34	-0.700	0.279	0.012		-1.247	-0.154		
	Difference		-0.405	0.356	0.257	0.639	-1.102	0.293	1.609	
36	Aβ–	52	-0.427	0.290	0.142		-0.996	0.142		
	Aβ+	27	-1.426	0.398	<0.001		-2.206	-0.646		
	Difference		-0.999	0.494	**0.044**	0.144	-1.967	-0.031	2.123	0.413
Area between curves			12.5	8.81	0.160					

IRT latent ability from ADCS-PACC items

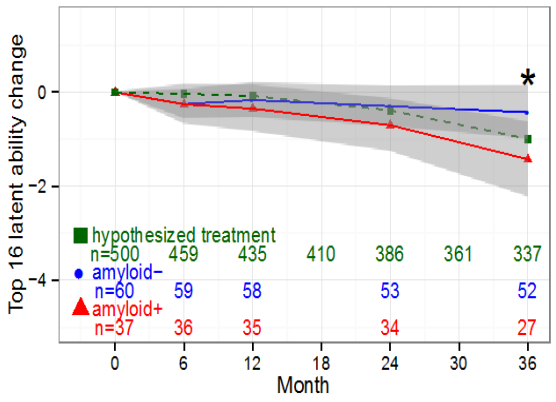
Month	Group	N	Estimate	SE	p-value	p-value*	lower	upper	sigma	rho
6	Aβ–	59	0.025	0.158	0.876		-0.286	0.335		
	Aβ+	36	-0.119	0.206	0.564		-0.523	0.285		
	Difference		-0.144	0.262	0.583	0.961	-0.656	0.369	1.214	
12	Aβ–	58	0.462	0.168	0.006		0.134	0.791		
	Aβ+	35	-0.013	0.219	0.954		-0.442	0.416		
	Difference		-0.475	0.277	0.087	0.276	-1.018	0.068	1.276	
24	Aβ–	53	0.452	0.169	0.008		0.121	0.784		
	Aβ+	34	-0.117	0.216	0.589		-0.539	0.306		
	Difference		-0.569	0.275	**0.040**	0.134	-1.108	-0.029	1.244	
36	Aβ–	52	0.199	0.174	0.255		-0.143	0.540		
	Aβ+	27	0.048	0.241	0.841		-0.424	0.521		
	Difference		-0.150	0.298	0.615	0.971	-0.735	0.435	1.272	0.364
Area between curves			12.9	6.71	0.055					

The models assume heterogeneous compound symmetric covariance structure, which allows different variance parameters (sigma) per visit, and a single correlation parameter (rho). The IRT model was trained on the N=322 ADNI normal controls with unknown Aβ status, and tested using N=97 ADNI normal controls known Aβ status. The top 16 ADNI items were selected based on a threshold of 2.5 applied to total information scores.

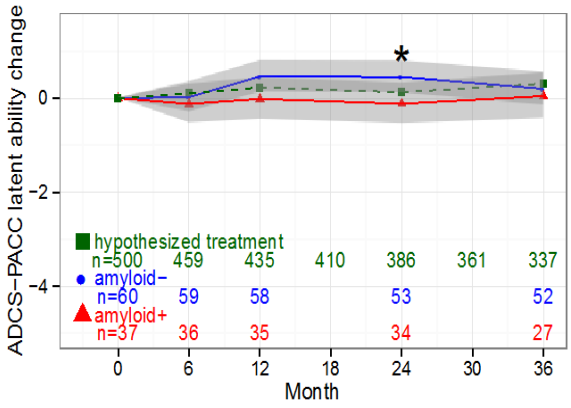
SE = Standard Error; p-value* = p-value adjusted for model-based simultaneous inference; lower = 95% confidence interval lower limit; upper = 95% confidence interval upper limit; sigma = residual standard deviation estimate at each visit; rho = estimated correlation between visits.

eFigure. MMRM estimates of IRT latent ability change from baseline.

IRT latent ability from top 16 items



IRT latent ability from ADCS-PACC items



The IRT model was trained on the N=322 ADNI normal controls with unknown Aβ status, and tested using N=97 ADNI normal controls known Aβ status. The top 16 ADNI items were selected based on a threshold of 2.5 applied to total information scores. The MMRM models assume heterogeneous compound symmetric covariance structure, which allows different variance per visit, and a single correlation parameter. Age and latent ability at baseline are included as covariates. The green dashed line indicates the hypothesized minimum treatment benefit that can be detected with 80% power, 5% alpha, and the indicated sample size and attrition. The shaded regions depict 95% confidence intervals. Asterisks (*) indicate group differences significant at $p<0.05$ level.

eReferences

1. Nelder JA, Mead R. A simplex method for function minimization. *The Computer Journal*. 1965;7(4):308-313. doi: 10.1093/comjnl/7.4.308.

2. Rizopoulos D. ltm: An R package for latent variable modeling and item response analyses. *Journal of Statistical Software*. 2006;17(5):1-25.