Appendix e-1

The initial goal of ADNI (http://www.loni.usc.edu/ADNI) was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD.

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the

research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Appendix e-2

Table S1: Screen failure fraction (SFF), 2-year change on clinical outcome measures (mean±std), 2-year signal to noise ratio (SNR, mean/std), number of subjects that need to undergo screening (NNS), trial cost (M$), overall trial duration (years). Compared to Table 2 in the main manuscript, results for N+ enrichment are presented for a hippocampal volume cut point at the 40th percentile of the volume in healthy subjects (25th percentile in main manuscript)

	Unenriched		N+		A+		N+ then A+		A+ then N+	
	MMSE	ADAS-Cog $_{13}$	MMSE	ADAS-Cog $_{13}$	MMSE	ADAS-Cog $_{13}$	MMSE	ADAS-Cog $_{13}$	MMSE	ADAS-Cog $_{13}$
SFF	0%		24%		28%		24%->42%		28%->42%	
2y. change	-1.77±3.19	3.87±7.35	-2.10±3.37	4.75±7.67	-2.40±3.35	5.58±7.54	-2.69±3.44	6.33±7.68	-2.69±3.44	6.33±7.68
SNR	-0.55	0.53	-0.62	0.62	-0.72	0.74	-0.78	0.82	-0.78	0.82
Sample Size	816	908	644	654	489	458	410	369	410	369
NNS	2,332	2,593	2,425	2,460	1,945	1,819	2,020	1,817	2,020	1,817
Cost	74	83	62	63	58	54	50	45	53	47
Time	4.9	5.2	5	5.1	4.4	4.3	4.5	4.3	4.5	4.3

Appendix e-3

Hippocampal volume adjustment for age and head size

We built a multiple linear regression model for age and intracranial volume (ICV) using raw HVs in the 444 normal study subjects as $\mathrm{HV} = \beta_0 + \beta_1 \times \mathrm{ICV} + \beta_2 \times \mathrm{age} + \epsilon_1$ where β_i (i=1,2) are the regression coefficients to be computed from the data (HV, ICV, age) and ε represents the residual error. The model was then applied to the MCI subjects to compute adjusted HV (aHV) as $\mathrm{aHV} = \mathrm{HV} - \beta_1 \times \mathrm{ICV} - \beta_2 \times \mathrm{age} + \epsilon_1$.

Sample sizes

The hypothetical two-arm study is powered to measure a 25% difference in change in either of the clinical scales investigated (ADAS-Cog_{13}, MMSE) with 80% power and 5% significance. With mean change μ and standard deviation σ, the sample size N for two arms is calculated as follows

$$N = \frac{4\sigma^2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{(0.25\mu)^2}$$

The cut-points of the standard normal probability distribution matching the defined significance (α) and power (1-β) are $z_{1-\alpha/2} \approx 1.96$ and $z_{1-\beta} \approx 0.84$ respectively.

Subjects that need to be screened, trial cost and time

To estimate the number of subjects that need to undergo screening (NNS), we model a screen failure fraction before biomarker enrichment described as SFF'. Together with the screen failure fraction (SFF) of a given enrichment strategy, this gives

$$NNS = \frac{N}{(1 - SFF) \times (1 - SFF')}$$

Overall trial cost with for an enrichment strategy with two biomarkers B1, B2 is calculated as

$$C_T = \frac{NSS_{\mathrm{B1}}}{SFF'} \times C_s + NNS_{B1} \times C_{B1} + NNS \times C_{B2} + N \times D \times C_m$$

where $NNS_{B1} = N \times SFF_{B1}$ is the number of subjects that need to be tested with biomarker B1 with the screen failure rate SFF_{B1} (for completeness, $SFF = SFF_{B1} \times SFF_{B2}$). $C_s$is the screening cost before biomarker enrichment and includes cost for a brain MRI, genotyping, as well as cognitive and functional testing. C_{Bi} (i=1,2) are costs for obtaining biomarkers B1 and B2 (i.e. automated hippocampal volumetry and amyloid status from amyloid PET (incl.acquisition)). *D* is the duration of the trial and C_m is the annual cost to maintain a subject in the trial. Table B1 summarizes model parameters, extended from [20].

Table B1: Trial parameters

Symbol	Parameter	Value
D	Trial treatment duration	24mo
C_S	Screening cost per patient	$5,800
C_{HCV}	Additional cost for each HV measurement	$200
$C_{amyloid}$	Acquisition / reading cost for amyloid PET	$7,500
C_m	Annual maintenance cost per patient	$18,500
SFF'	Screen failure fraction before biomarker enrichment	0.7
R_S	Annual screening rate	800

Trial time is considered as the sum of the screening time and the trial observation time after randomization of the last subject:

$$T_T = \frac{NNS}{SFF' \times R_S} + D$$

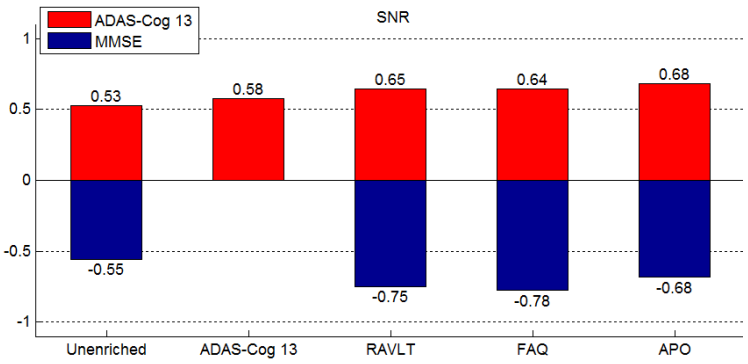
Appendix e-4

For the alternative enrichment strategies using ApoE-4 status, cognition, memory, and function at baseline, the following cut-points were used:

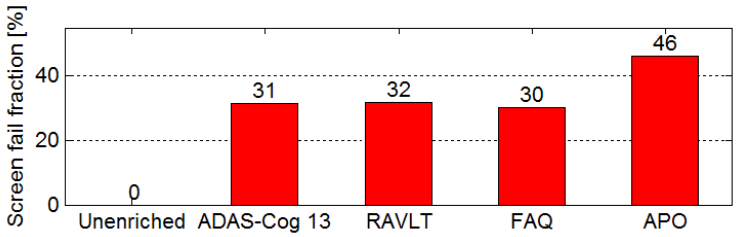
- subjects were considered positive on ApoE-4 if they carry at least one ε4 allele.
- subjects were considered positive on function with a baseline FAQ > 0.
- subjects were considered positive on Adas-Cog$_{13}$ with a baseline score > 15.
- subjects were considered positive on RAVLT with a baseline score < 35.

Cutpoints for the scales on memory, cognition and function were set in a way to get closest to 30% screen failure rate to achieve comparability to the two single-biomarker enrichment strategies (A+ / N+).

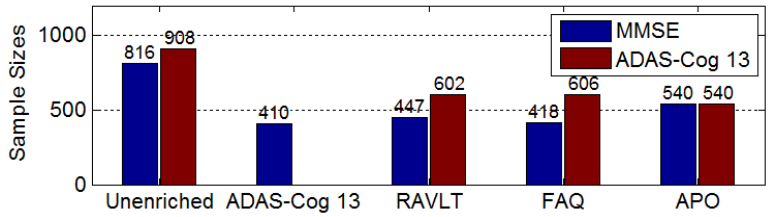
No improvement was found when screening out subjects with an MMSE larger than 28 or 29 at screen failure rates of 44% and 18% respectively.



(a): Signal to noise



(b): Screen fail fraction



(c): Sample sizes

Figure S1: "Trial characteristics with different alternative enrichment strategies". Signal to noise ratio (SNR) (a), screen fail fraction (b) and required sample sizes (c) with different enrichment strategies.

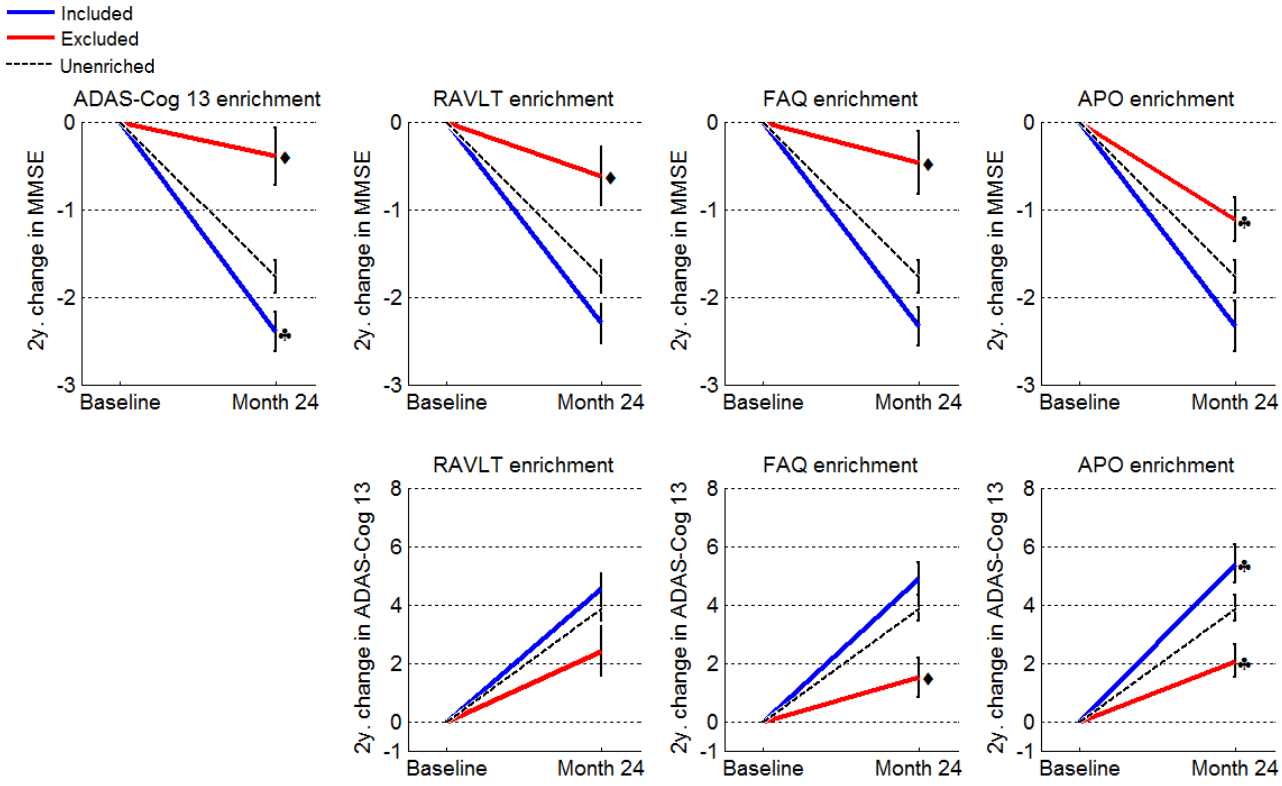


Figure S2: "Time course graphs for different alternative enrichment strategies". Presented is the change in MMSE (top row) and ADAS-Cog $_{13}$ (bottom row) for the unenriched sample (dashed black line), the enriched sample (solid blue line) and the screened out sample (solid red line). Whiskers present standard error. Significance of difference in included and excluded groups to unenriched sample is reported with ♣: $p<0.05$ and ♦: $p<0.01$.

Table S1: Screen failure fraction (SFF), 2-year change on clinical outcome measures (mean±std), 2-year signal to noise ratio (SNR, mean/std), number of subjects that need to undergo screening (NNS), trial cost (M$), overall trial duration (years). NE: not evaluated

	Unenriched		ADAS-Cog 13		RAVLT		FAQ		APO	
	MMSE	ADAS-Cog $_{13}$	MMSE	ADAS-Cog $_{13}$	MMSE	ADAS-Cog $_{13}$	MMSE	ADAS-Cog $_{13}$	MMSE	ADAS-Cog $_{13}$
SFF	0%		31%		32%		30%		46%	
2y. change	-1.77±3.19	3.87±7.35	-2.40±3.07	NE	-2.30±3.08	4.55±7.05	-2.34±3.01	4.89±7.61	-2.33±3.42	5.41±7.95
SNR	-0.55	0.53	-0.78	NE	-0.75	0.64	-0.78	0.64	-0.68	0.68
Sample Size	816	908	410	NE	447	602	418	606	540	540
NNS	2,332	2,593	1,706	NE	1,873	2,521	1,712	2,486	2,875	2,876
Trial Cost	74	83	40	NE	44	59	41	59	57	57
Trial Time	4.9	5.2	4.1	NE	4.3	5.2	4.1	5.1	5.6	5.6

Appendix e-5

The presented results assume a 24-month trial period. Table S-6 shows for comparison the impact of biomarker enrichment in a hypothetical trial with a duration of 6 and 12 months. Averaged across both endpoints (MMSE, ADAS-Cog_{13}), SNR improves by 21%, 34% and 53% when enriching with N+, A+ and A+N+ respectively and when considering a 6-month trial. When considering a 12-month trial, SNR improvements are 26%, 40% and 66% respectively, giving a comparable relative improvement to the values presented in this paper for a 24-month trial.

Table S-6: Enrichment performance over 6 months and 12 months. NNS: number of subjects needed to screen

6 Months						
	Signal to noise ratio		**Sample sizes**		**NNS**	
	ADAS-Cog_{13}	**MMSE**	**ADAS-Cog_{13}**	**MMSE**	**ADAS-Cog_{13}**	**ADAS-Cog_{13}**
Unenriched	0.32	-0.26	2,428	3,795	6,937	10,844
N+	0.43	-0.28	1,336	3,234	5,330	12,937
A+	0.45	-0.33	1,229	2,295	4,901	9,087
A+N+	0.55	-0.35	825	2013	4,266	10,327
12 Months						
Unenriched	0.35	-0.31	2,086	2,577	5,959	7,363
N+	0.44	-0.39	1,311	1,660	5,210	6,598
A+	0.46	-0.46	1,168	1,198	4,640	4,759
A+N+	0.54	-0.55	866	816	4,460	4,205

Appendix e-6

In the operational consideration of price, the presented work assumes an analysis of amyloid positivity (A+) measured from PET imaging but cut points from CSF Aβ were used to define A+ subject in the on the ADNI 1 database as limited amyloid-PET imaging is available. To analyse the effect different measurements have on the included patient population, Table S5 presents a comparison of PET and CSF-derived cut points on the ADNI 2 database where both measurements were acquired. Good concordance can be observed between both measurements confirming previous experiments [31] [34].

Table S5: Comparison of amyloid PET and CSF Aβ for defining Amyloid positivity (A+). Values for two-year MMSE and ADAS-Cog are shown together with baseline PET SUVR and baseline CSF Aβ concentration. For each measurement, values that show same color-coding are not significantly different from each other.

	PET+	**CSF+**	**CSF+/PET+**	**CSF-/PET+**	**CSF+/PET-**	**CSF-/PET-**
N (%)	83 (68%)	88 (72%)	78 (64%)	5 (4%)	10 (8%)	29 (25%)
2 y. change MMSE	-2.65±2.73	-2.64±2.66	-2.74±2.76	-1.00±1.41	-1.78±1.39	-0.10±2.18
2 y. change ADAS-Cog	5.44±6.90	5.19±6.81	5.74±6.91	0.25±3.95	0.56±3.47	1.86±3.85
Baseline PET	1.40±0.17	1.37±0.20	1.41±0.17	1.17±0.07	1.03±0.06	1.01±0.05
Baseline CSF Aβ	138±29.0	136±26.2	134±25.6	202±9.4	153±26.6	236±32.2