Innovative Alzheimer's Disease clinical trial design in the coming age of biomarkers

By: Hampus Hillerstrom
M.S., Economics
University of St.Gallen, 2001
MBA, Business Administration
Harvard Business School, 2007
and
and the Alzheimer's Disease Neuroimaging Initiative*

Submitted to the
Harvard-MIT Division of Health Sciences and Technology
In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Health Sciences and Technology

at the
Massachusetts Institute of Technology
August 2008

© 2008 Hampus Hillerstrom. All rights reserved.
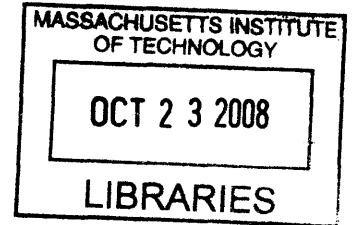
Signature of the author........... ....
Harvard-MIT Division of Health Sciences and Technology

Certified by ...   .
Joseph V. Bonventre M.D., Ph.D.
Robert H. Ebert Professor of Medicine and Health Sciences and Technology,
Harvard Medical School & Brigham & Women's Hospital
Thesis Supervisor

Certified by ....
A. Gregory Sorensen, M.D.
Co-Director, MGH-HSY A.A. Martinos Center
Associate Professor of Radiology, Harvard Medical School
Thesis Supervisor

Accepted by ....
Martha L. Gray, Ph.D.
Director
Harvard-MIT Division of Health Sciences

Abstract

Alzheimer's disease (AD) is a field with huge unmet need and only a few symptomatic treatments with limited efficacy have been made available to patients. With the testing of disease-modifying drugs in recent years, the length of AD clinical trials has tripled and the enrollment has gone up drastically. These investigational disease-modifying drugs address new targets including the amyloid beta and tau protein aggregation pathways in the brain. They have opened up a whole research field on biomarkers specific to these pathways. These biomarkers have however never been used to select a subpopulation that would enroll in clinical trials.

This thesis defines a framework for assessing any AD biomarker's quality as a selection tool for enrolling a subpopulation into an AD clinical trial. Carefully selecting the patient population with appropriate biomarkers can lead to a reduction in required enrollment in a study to show statistical significance. In turn, the decreased patient enrollment helps sponsors reduce costs and allows them to test several drugs with the same budget.

In order to test our framework in an applied and relevant setting, we established from www.clinicaltrials.gov that for disease-modifying drugs the primary endpoint is change in ADAS-cog points at 18 months and that the trials enrolled on average 337 patients per treatment group.

These disease-modifying AD trials use the inference on means statistical model. The standard deviation and the treatment effect of the primary endpoint variable (the change in ADAS-cog points at 18 months) are the main leverage factors that will influence the required enrollment (or sample size) in the trial.

In a first step, we defined the baseline values for those main variables from published information on past or ongoing trials. Using that information, we conducted a theoretical exercise showing how much you needed to affect these variables in order to reduce enrollment by a factor of 5x, an important reduction in enrollment that could potentially realistically be achieved.

In a second step, we looked in an applied setting at how well a selection of biomarkers in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database reduces the sample size by only selecting a sub population of the patients. Even with the limited data sample available on a preliminary basis from ADNI, we found that the biomarkers ABeta 1-42, ratio of Tau/ABeta 1-42, Apoe4 carriers on both genes, and average hippocampal volume show predictive power to identify change in ADAS-cog scores. When using criteria for these biomarkers to select a subpopulation we show that you can reduce the enrolled population by up to a factor 5.0x while decreasing your trial cost by up to 73% (corresponding to a $92M reduction out of $133M, the current Phase 3 costs of an 18 months diseases-modifying drug). Under the best scenario of these cost savings the sponsor can conduct pivotal trials for three drugs instead of only one.

In a last step, the biomarker and combination of biomarkers associated with the enrollment benefit and the cost savings were assessed with additional criteria such as effect on restricted labeling, necessity for longitudinal screening or additional enrollment difficulties. Even after that analysis, several of the biomarkers stood out as very strong candidates to select for subpopulations in future disease-modifying trials and save costs.

ADNI is an industry and NIH-sponsored initiative monitoring 800 normal, Mild Cognitive Impairment (MCI) and AD patients for up to three years with regular cognitive assessments and biomarker measurements.

Acknowledgements

Table of contents

Abbreviations and key definitions

Abbr.	Full terminology	Description
Tau	Tau protein	Peptide frequently found in the CSF of Alzheimer's patients
ABeta 1-42	Amyloid Beta 1-42 protein	42 amino acid long peptide frequently found in the CSF of Alzheimer's patients
AD	Alzheimer's Disease	-
ADAS-cog	Alzheimer's Disease Assessment Scale, cognitive subpart	Cognitive assessment test. Scores range from 1 to 70 (higher is more demented). These seventy points are relevant to the cognitive subpart of the test (correspond to 11 of 14 parts). Importantly, the change of ADAS-cog over time (6, 12 or 18 months) is the primary endpoint in all AD trials to date.
ADNI	Alzheimer's Disease Neuroimaging Initiative	-
Apoe4	Apolipoprotein E4	Apolipoprotein E is 299 amino acids long protein. Apoe2, ApoE3 and Apoe4 differ only by a single amino acid. Apoe4 is associated with a higher frequency of AD but is found in only 1-2% of the total population. Typically 40-65% of AD patients have at least one copy of the Apoe4 allele, but only 20-50% of the AD population has both gene copies of the Apoe4 allele associated with a much higher AD incidence.
CSF	Cerebrospinal fluid	Fluid bathing the spinal cord, the brain ventricles and the cortex.
MCI	Mild Cognitive Impairment	Variably defined but includes subjective memory or cognitive symptoms or both, objective memory or cognitive impairment or both, abnd generally unaffected activities of daily living; affected people do not meet currently accepted dementia or AD diagnostic criteria.
MMSE	Mini Mental State examination	Cognitive assessment test. Scores range from 1 to 30 (lower is more demented).
MRI	Magnetic Resonance Imaging	Imaging method used in Alzheimer's
P-tau	Phosphorylated tau	Peptide frequently found in the CSF of

181P		Alzheimer's patients
PET	Positron Emission Tomography	Imaging method used in Alzheimer's
Probable AD	-	This means that the patient has AD. He is only diagnosed with definite AD at autopsy.
Prodromal AD	-	The symptomatic pre-dementia phase of AD, generally included in the mild cognitive impairment category; this category is categorized by symptoms not severe enough to meet currently accepted diagnostic criteria for AD.
SD	Standard Deviation	In ADAS-cog points, if not indicated otherwise.
SD lever	Standard Deviation leverage factor	The standard deviation is a leverage factor to decrease sample size. See section 2.4 for further explanations.
TE	Absolute treatment effect	In ADAS-cog points, if not indicated otherwise. See formula in section 2.3.
TE %	Percent treatment effect	In %. See formula in section 2.3.
TE lever	Treatment Effect leverage factor	The treatment effect is a leverage factor to decrease sample size. See section 2.4 for further explanations.

1. Introduction

1.1. Problem

Many drugs tested in the past for the treatment of Alzheimer's Disease (AD) failed to show efficacy and the currently approved treatments for the disease only treat symptoms. There are two classes of approved symptomatic AD drugs: acetylcholinesterase inhibitors (the leading treatment in the class is Aricept donepezil) and NMDA receptor antagonists (only approved treatment is Namenda memantine). These treatments have limited efficacy and merely delay progression of the disease by up to six months so there is an enormous need for better AD treatments. With a mechanisms of action boosting the communication between neurons, these symptomatic drugs constituted the "low hanging fruit".

With the testing of disease-modifying drugs in recent years, the length of AD clinical trials has tripled to 1.5 years and the trial enrollment has gone up dramatically. Despite that several disease-modifying drugs failed in Phase 3 trials. We most recently saw the Phase 3 trial failures of Alzhemed 3APS and Flurizan MPC-7869. The Flurizan trial cost was estimated at $150 millions.[1]

The pharmaceutical companies are faced with two great difficulties when design Alzheimer trials:

1) they have difficulty in knowing before Phase 3 the effect of their drug which leads to very costly failures in the pivotal trials.
2) should the drug have only a modest effect, they are faced with the dilemma of underpowered trials or prohibitive costs.

The Alzheimer's research community has mostly focused on the first problem. They have looked at solutions from various angles:

- New categorization of patients: the MCI diagnostic category (also called prodromal AD) was coined and a new FDA-recognized primary endpoint of "time to progression" was suggested. The MCI category was later subdivided into amestic and non-amnestic sub-groups in order to further increase predictability.
- Earlier intervention trials: the diagnostic criteria of prodromal AD (new NINDS-CRADA criteria) have been revisited by a group of Alzheimer experts in 2007 with the purpose of helping design early intervention trials.[2]
- Better efficacy measures: in 2008, pharmaceutical companies such as Wyeth/Elan switched cognitive assessment scales to NTB from ADAS-cog as they hope the scale will be more sensitive as their primary outcome variable for cognitive decline.[3]
- Better screening tools: a group of researchers proposed a new screening tool called the Montreal Cognitive Assessment (MoCA) which is supposed to be a more sensitive screening tool for Mild Cognitive Impairment patients than the Mini Mental State Examination (MMSE).[4]

- Longer trials: pharmaceutical companies and the NIH have started to conduct several long AD prevention trials ranging from 2 to 10 years.
- Evaluation of biomarkers as surrogate markers: approval by the FDA needs highly validated biomarkers, but some biomarkers such as MRI brain volume measurements have shown promise in this respect.
- Concerted efforts to understand biomarker characteristics: The Alzheimer's Disease Neuroimaging Initiative (ADNI) was set up to help validate biomarkers and surrogate markers and received funding from several major pharmaceutical firms active in the Alzheimer's field.
- Innovative trial design: pharmaceutical companies such as Myriad Genetics and the Food and Drug Administration (FDA) have been suggesting other trial designs such as "Randomized Withdrawal", "Staggered Start" or "Natural History Staggered Start" but these trial designs only found very little traction in the industry in general.[5]

In addition to these efforts, the pharmaceutical companies need to explore solutions that address the second problem of prohibitive costs.

With the advent of new disease-modifying AD drugs addressing new targets (including the amyloid beta and tau protein aggregation pathways in the brain) and promising disease progression imaging tools, a number of new biomarkers have come to light. This has paved the way to the use of these biomarkers in a number of the efforts described above.

It has also created the basis for the framework proposed in this thesis of selecting a subpopulation of patients to enroll in clinical trials, for the purpose of reducing the needed enrollment and thereby of reducing costs.

A good example highlighting the need for biomarker selected subpopulation of patients is Wyeth/Elan's Phase II bapineuzumab data published in June 2008. The company did a post fact analysis showing that the treatment worked best in a subpopulation of patients who did not have the ApoE4 genetic mutation (ApoE4 non-carriers). As a result they are now conducting two separate pivotal Phase 3 trials for the two subgroups (one for ApoE4 carriers, and one for ApoE4 non-carriers).

1.2. Goal and key questions

This thesis is focused on innovative clinical trial design using biomarkers to select a subpopulation enrolling in the pivotal Phase 3 trial.

We are aiming at answering the following question:

1a. Theoretically, how good does the biomarker (or combination of biomarkers) need to be to select for a patient sub population that reduces the enrollment

needed in Phase 3 by a factor 5x, an important reduction in enrollment that could potentially realistically be achieved?"

1b. In practice and based on data from the ADNI database, how good are existing biomarkers (or combination of biomarkers) in selecting for a patient sub population to reduce the enrollment needed in Phase 3 clinical trials?
1c. What is the cost impact on patient screening and patient enrollment from selecting for a patient sub population in the Phase 3 trial?

1d. What additional considerations are important in order to assess the selected biomarkers and what are the final conclusions on the overall quality of those biomarkers?

1e. How does recent data from AD disease-modifying drugs relate to the findings in this thesis?

The thesis proposes a framework trial design that enables researchers in academia, biotech or pharmaceutical companies to do something they cannot currently do. Example of such improvements with biomarkers may include a drastic decrease in patients enrolled or enrolling only those patients that have the disease leading to higher likelihood of success.

We will focus in this thesis on how biomarkers can help strongly reduce the number of enrolled patients and thereby reduce costs. We do this in the following steps:

1. Assess the current context of disease-modifying AD clinical trials by analyzing the www.clinicaltrials.gov database and previously published data from Alzheimer trials.
2. Choose and define the relevant sample size statistical model for the primary endpoint of change in ADAS-cog at 18 months.
3. Show the theoretical impact on sample size of the variables.
4. Perform regression analysis on data from the ADNI database to find the most predictive biomarkers.
5. Stratify these predictive biomarkers and choose criteria to select subpopulations.
6. Using these biomarker selected patient subpopulations, assess how the biomarker subpopulation criteria impact sample size.
7. Using these biomarker selected patient subpopulations, assess how the biomarker subpopulation criteria impact costs.
8. Assessed the sample size and cost impact of the biomarker subpopulation criteria with additional factors such as effect on restricted labeling, on necessity for longitudinal screening or on enrollment difficulty.
9. Conclude on the overall quality of the criteria to select for a subpopulation and on their applicability for designing future disease-modifying trials.

Given the more promising nature of disease-modifying Alzheimer's drugs and the difference in clinical trial design, the analysis of the data in www.clinicaltrials.gov will not include trials of symptomatic Alzheimer treatments.

An additional goal is to conclude on practical feasibility, usefulness and cost-effectiveness of the proposed innovative clinical trial design of using a biomarker selected subpopulation.

Finally, this thesis is not meant to be:

- *a statistical exercise alone*: while we are using analyzing statistical models for this thesis, the goal is not to be abstract, but instead to use an applied perspective throughout this work. As such we have tried our best to anchor every assumption throughout the thesis in facts know from the industry, from published papers or literature and from the ADNI patient database. We wish therefore that our findings will resonate and can be directly applied by industry and academic sponsors conducting Alzheimer's trials.

- *an analysis of biomarkers as surrogate endpoint*: the FDA has clearly stated that it is currently not ready to accept biomarkers as surrogates in AD before further validation.[6]

- *an analysis of the time-to-progression from MCI to AD*: the FDA has recognized two possible primary endpoints to date - change in cognitive and functional assessment scales, and time to progression. In this thesis we will focus solely on the change in cognitive assessment scale endpoint as we can then compare our newly proposed trial design to the one currently used for disease-modifying drugs in AD. In addition, all currently marketed drugs showed improvement on the cognitive change outcome variables. It would however be very interesting to conduct the same analysis with the time to progression primary endpoint, but it is beyond the scope of this thesis.

- *an analysis on how biomarkers affect the length of clinical trials*: this is both a very interesting topic and could be very valuable for industry players but it will not be analyzed in this thesis.

1.3. Disclaimers

Limited clinical-biomarker correlation:
Like in many other fields in medical diagnostics, there is no 1.0 positive correlation between any one biomarkers and the clinical symptoms. They typically vary between -0.5 and 0.5 with the norm being -0.2 to 0.2. Appendix 1 show the correlation coefficients of the biomarkers data sample that we will be using in this thesis.

Preliminary data:
The data from the ADNI database are still preliminary: the data were provided raw and are updated regularly. The underlying study is still ongoing at submission of this thesis. The results in this thesis may therefore change as new data is provided (in particular longitudinal data and ADAS-cog scores at 18 months) and a larger sample can be

obtained. Additional biomarkers not analyzed in this thesis could also be included and may lead to additional or changed findings.

The ADNI data cannot easily be combined into one data set which makes the data mining cumbersome and not automated. In section 5, we will further elaborate on the data that is not available in the ADNI database, but which would have been interesting to analyze.

2. Hypothesis

2.1. Biomarker applications in AD research and development

For the purpose of this thesis the term biomarker is defined broadly: it includes imaging markers, CSF markers, genotyping and clinical scores. The following table outlines the most important AD biomarkers monitored in AD research and development:

Biomarker	Expected change
I. Clinical	
MMSE	Decrease with time and worsening dementia.
ADAS-cog	Increase with time and worsening dementia.
II. Biomarker	
A-beta 1-42	Increase with time and worsening dementia.
Tau	Increase with time and worsening dementia.
P-tau	Increase with time and worsening dementia.
ApoE4	Genotype associated with higher risk of dementia when carrying two copies of the Apoe4 allele.
III. Imaging	
MRI volume atrophy	Hippocampal volume decrease with time and worsening dementia.

Historically biomarkers have been used in the Alzheimer's Disease field for the following applications (see Appendix 2 for additional details):

1. Identifying AD (sensitivities and specificities vary from 65-96% and 23-100% respectively);
2. Identify Prodromal AD (sensitivities vary from 74-96%);
3. Identifying converters from MCI to AD and not to other dementias (sensitivities and specificities of 91% and 100% respectively). One paper from Dubois et al. 2007 cites that at least 30% of enrolled amnestic MCI patients are false positives when measuring whether they will evolve to AD (i.e.30% of amnestic MCI patients do not later evolve to AD).[7] That is a good example of where a biomarker would be very valuable to define a sub population and thereby decrease the number of enrolled patients and increase the response to a given drug.

4. <u>Differentiating normal patients from those who will evolve to MCI</u>: Carlson et al. (2008) showed in a 79 patient study over 15 years that ventricular volume expansion measured by MRI occurred on average 2.3 years prior to clinical diagnosis of MCI.[8]
5. <u>Differentiating AD from Fronto-Temporal Dementia</u> (sensitivities and specificities of 78% and 71% respectively)
6. <u>Differentiating AD from Dementia with Lewy Bodies</u> (sensitivities and specificities vary from 86-92% and 80-81% respectively)
7. <u>Differentiating AD from Vascular Dementia</u> (sensitivities and specificities vary from 75-88% and 18-53% respectively)
8. <u>Surrogate markers as trial primary endpoints:</u> efforts to validate surrogate markers to replace cognitive assessment scales as primary endpoints have not succeeded to date with the FDA. Jack et al. (2003) showed that using MRI volumetric change as surrogate markers could decrease sample size by a factor 5.9x compared to current cognitive primary endpoints.[9]

In this thesis we explore another use of biomarkers, which consists of using biomarkers to select a patient subpopulation to enroll in the trial in order to reduce patient enrollment and trial costs.

2.2. Characteristics of a good biomarker to select a sub-population

The following characteristics make for a strong biomarker to select a sub population:
- <u>high predictive power</u>: the biomarker needs to have a strong association with the primary endpoint variable. Note that the predictive varies depending on the patient population chosen (eg. normal, MCI, mild AD, moderate Ad or severe AD). Please refer to section 4.2 for results showing this difference in predictive power.
- <u>no longitudinal data needed at the time of enrollment</u>: the need for longitudinal data would mean that enrollment can only occur after having observed a patient for 6-12 months which drastically increases screening costs and the associated logistics.
- <u>not too costly to perform</u>: the higher the screening cost, the more limited is the economic benefit of having a smaller enrolled population.
- <u>does not result in a restricted label</u>: ideally the sponsor prefers that the biomarker selected sub population results in a restricted label which limits the market opportunity. For instance, 40-70% of AD patients are ApoE4 carriers so if that biomarker is used to select a subpopulation the market opportunity could be reduced in the same proportion.
- <u>Not leading to enrollment difficulties</u>: much of the benefit of reducing the enrollment number can be offset by delays in recruiting the patients. So it is important that the biomarker does not lead to unreasonably high numbers of screened patients or to other reasons for delay.

2.3. Relevant statistical model and variables

As discussed at the end of Section 1.2, the cognitive change from baseline in ADAS-cog score is our primary endpoint. For that endpoint the inference on mean statistical model needs to be used (for more information on that, see section 3.2.1).

The inference model has the following sample size formula:

$$n = (z_{1-\beta} + z_{1-\alpha})^2 * (\sigma_T^2 + \sigma_P^2) / (\mu_T - \mu_P)^2$$

Throughout the thesis we will alternate the use of the words "sample size" and "enrollment". The former is a statistical term, whereas the latter is a clinical terminology.

The following variables influence the number of sample size:

Variable	Comments
α	Significance level, set at 5%
β (or the power: $1-\beta$)	Power, set at 90%: power for equivalence study min. 80% and for superiority studies 90% or several studies at 80%. It often depends on the quality of the Phase II data, the competitive landscape (need for an equivalence or superiority study), the unmet need and the cash resources of the sponsor. We assume a power of 90% in this thesis.
σ_P	Standard deviation of the cognitive change on ADAS-cog for the placebo group.
σ_T	Standard deviation of the cognitive change on ADAS-cog for the treatment group.
μ_P	Effect from baseline on the cognitive change on ADAS-cog for the placebo group.
μ_T	Effect from baseline on the cognitive change on ADAS-cog for the treatment group.
$(\mu_T - \mu_P)$	Absolute treatment effect (TE)

In addition, we also need the variable "average baseline ADAS-cog score". This variable is used to calculate the percent treatment effect (TE %) :

$$TE \% = [\mu_T - \mu_P] / [\text{average baseline ADAS-cog score}]$$

2.4. Leverage factor scenarios in statistical model

This thesis will focus on how a biomarker selected subpopulation can:

1) increase the treatment effect (μ_P, μ_T or $\mu_T - \mu_P$): going forward in this thesis we will call this leverage factor to decrease the sample size the **TE lever (or Treatment Effect leverage factor)**;
2) decrease the standard deviation (σ_P and/or σ_T): going forward in this thesis we will call this leverage factor to decrease the sample size the **SD lever (or Standard Deviation leverage factor)**;

and thereby lead to a decreased sample size.

There are therefore three "leverage factors scenarios" to influence the sample size:
1) SD lever only,
2) TE lever only,
3) SD & TE levers.

TE lever: any increase in treatment effect will have direct impact on the sample size.

To benefit from the TE lever there are three possible situations:
- Situation A $\mu_P\uparrow$ **and** $\mu_T\rightarrow$: increasing the placebo group change in ADAS-cog ($\mu_P\uparrow$) while keeping the treatment group change the same ($\mu_T\rightarrow$);
- Situation B $\mu_P\rightarrow$ **and** $\mu_T\downarrow$: decreasing the treatment group change in ADAS-cog ($\mu_T\downarrow$) while keeping the placebo group change the same ($\mu_P\rightarrow$), or;
- Situation C $\mu_P\downarrow$ **and** $\mu_T\downarrow\downarrow$: strongly decreasing the treatment group ($\mu_T\downarrow\downarrow$) change in ADAS-cog if the placebo group change is reduced ($\mu_P\downarrow$).

In addition there is one scenario where the TE lever cannot be used:
- Situation D $\mu_P - \mu_T$ unchanged: that means that μ_P and μ_T are affected in parallel and there is no treatment effect and therefore no TE lever to influence the sample size.

We decide to rule out the Situations B & C:
- Situation B is a plausible scenario. However, with the ADNI data for untreated patients (corresponding to μ_P and not to μ_T), we do not have a reference point by to analyze this situation in this thesis.
- Situation C is an unrealistic situation given the poor treatment effect observed by current marketed Alzheimer's drugs or drugs in development.

Therefore in this thesis we will focus on the two following situations for the treatment effect:
- Situation A where μ_T remains unchanged (effectively meaning that you want to pick a patient subpopulation that has a strong decrease in the placebo group and an unchanged treatment group), and
- Situation D where $\mu_T - \mu_P$ remains unchanged (effectively meaning that there is no increased treatment effect so no TE lever).

These two situations represent the two extremes: the first, where the treatment effect is large and independent of the patient subpopulation chosen, and the second, where there is a no treatment effect.

SD lever: any decrease in standard deviation will have direct impact on the sample size. We assume that σ_P and σ_T are equal as evidenced by published data from Aricept donepezil AD trials[10].

In summary: the three scenarios to affect sample size con be characterized as follows:
- SD lever only: situation D above, μ_T - μ_P remains unchanged, no TE lever,
- TE lever only: scenario in which the standard deviation remains unchanged, no SD lever,
- SD & TE levers: situation A above, μ_T remains unchanged, benefit from both leverage factors.

Given the fact that the SD lever is clear-cut in its effect (a lower standard deviation is always better) we do not see any value added in looking at the scenario TE lever only.

Therefore, throughout this thesis we look only at the following two leverage factor scenarios:
1) *SD lever only,*
2) *SD & TE levers.*

2.5. Values of variables in current context of disease-modifying AD trials

2.5.1. Summary of Clinicaltrials.gov data analysis

The published protocols on www.clinicaltrials.gov provided the trial duration, the enrollment number, the primary endpoints and the inclusion criteria. There was very little variability in the design (same length, similar endpoints, and similar inclusion/exclusion criteria) of disease-modifying AD trials. All were 18 months long and had on average 337 enrolled patients. There was a small difference in the patient enrollment for the individual studies and in the number of treatment arms per study.

This data was compiled from 11 disease-modifying Ph.III trials for 4 drugs (Eli Lilly's LY450139, Wyeth/Elan's bapineuzumab, Neurochem's 3APS and Myriad Genetics' MPC-7869). More detailed explanations are found in the methods section 3.1 and in Appendix 3.

2.5.2. Data analysis from published historical clinical trials

The first step is to analyze the data of previously approved drugs. In order to estimate the other variables, we rely in a first step on data published from previously approved symptomatic AD treatments (eg. Aricept donepezil):
- We need both the standard deviation of the treatment group and the standard deviation of the placebo group.

- From a look at the data published in the report "Donepezil for dementia due to Alzheimer's Disease", which summarizes the data from all donepezil trials to date, the standard deviation of the change from baseline in ADAS-cog score at week 24 for the 5mg dose was 5.81 points for the treatment arm and 5.86 points for the placebo arm.
- In order to estimate the treatment effect, we use as benchmark the published results of marketed drug Aricept which shows an improvement of -0.9 ADAS-cog points at 24 weeks (compared to +1.81 points for placebo-treated). In that study patients had an ADAS-cog mean score of 26.[11] Putting these figures together we show that Aricept had a statistically significant -0.9/26 vs +1.81/26 i.e. a 2.71/26 or an 10.4% treatment effect as measured by relative change of ADAS-cog compared to placebo.

The following table summarizes the described 6 months Aricept donepezil trial data (5 mg dose):

Variable	Value
Trial length	24 weeks
MMSE inclusion criteria	10-26 score range
Patients per group	Actual: 130, Estimated: 108
Standard deviation treatment group	5.81 ADAS-cog points
Standard deviation placebo group	5.86 ADAS-cog points
Mean baseline score	26 ADAS-cog points
Change from baseline in treatment group	-0.9 ADAS-cog points
Change from baseline in treatment group	1.81 ADAS-cog points
Calculated treatment effect	10.4%

The next step is to use the values of these variables to better estimate the values for the typical 18-month disease-modifying trial:
- These donepezil estimates are based on 24 weeks data (6 months) whereas the current disease-modifying trials are 78 weeks (18 months). So the values need to be revised for the longer, 18 month timeframe.
- Unlike the donepezil group which had an MMSE score range of 10-26 (mean around 18), the new disease-modifying trials have a 16-26 score range (mean around 21). With the donepzil mean baseline score on ADAS-cog of 26 points, the baseline score for the disease-modifying trials should be slightly lower. Based on data published in a paper from Grundman et al. (Jan 2004) which compares MMSE and ADAS-cog-scores for differently demented patients we assume here that the mean baseline ADAS-cog score of disease-modifying trials is (18+25.2)/2 = 21.6 points.[12]
- The standard deviations were almost identical between the treatment and placebo groups for donepezil. We will for now assume that it remains he case.

- With a 10.4% treatment effect for donepezil, we are looking for at least that treatment effect for the 18 months disease-modifying treatments. We will back out the treatment effect as the last variable after having estimated all the other variables.

Based on the above analysis, this is a summary of data for the key protocol variables in current 18 month disease- modifying trials:

Variable	Value	Comments
Trial length	78 weeks	Source: www.clinicaltrials.gov (see Appendix 3)
MMSE inclusion criteria	16-26 score range	Source: www.clinicaltrials.gov (see Appendix 3)
Patients per group	337 patients	Source: www.clinicaltrials.gov (see Appendix 3).
Standard deviation placebo group (σ_P)	9 ADAS-cog points	In our sample of 32 AD patients we have a standard deviation of the change of 7.2 points in ADAS-cog over 12 months. This is in line we the 6 points standard deviation of the change in ADAS-cog used to power the 12 months Phase 2 dimebon trial.[13] We are not looking at the S.D. of the 12 month change but we need instead the S.D. of the 18 month change. Therefore we expect the S.D. to be higher so we arbitrarily assume a S.D. of 9 points.
Standard deviation treatment group (σ_T)	9 ADAS-cog points	Assumed to be the same as for the placebo group.
Mean baseline score	21.6 ADAS-cog points	See calculations in paragraph above this table.
Change from baseline in placebo group (μ_P)	7.5 ADAS-cog points	In our sample of 32 AD patients we have a change of 5.4 points in ADAS-cog over 12 months. This is in line with the 6 points change in ADAS-cog observed in the placebo group at 12 months of the dimebon trial.[14] Another data points is the 11 points change at 18 months in the Apoe4 non-carrier group in the Phase 2 trial bapinuzemab trial.[15] Since we are not looking at the 12 month change, but instead at the 18 month change, we expect the

		change to be higher so we arbitrarily assume a change of 7.5 points.
Change from baseline in treatment group (μ_T)	4.94 ADAS-cog points	The 4.94 points were calculated as the "buffer" variable to allow the sample size to be 337 patient (the average for the current disease-modifying trials).
Treatment effect	11.9%	Calculated from the above. Please refer to section 2.3 for formula.
Drop-out rate	23%	Based on bapinuzemab's 18 month Phase 2 data (21-26%)[16]
Inclusion criteria	Age 50+, Probable AD, MMSE 16-26, caregiver available, women only if postmenopausal. And occasionally also following criteria: MRI not inconsistent with AD, living in community, education, communication level OK, stable symptomatic AD treatment.	

2.6. Innovative trial design addressing the key questions of the thesis

We have now defined all the variables in the trial design protocol of current disease-modifying drugs. We highlight in a separate column in the table below the goal of the thesis (sample size reduction), the levers to reach the goal (standard deviation and treatment effect) and the means to achieve it (addition of biomarkers as inclusion criteria).

Design parameter	Current Disease-Modifying AD trial design	Biomarker selected subpopulation trial design
Sample size per group	337 patients	**GOAL: reduce patient enrollment (i.e sample size)**
Trial duration	78 weeks	Stays the same in this thesis (see comment at the end of section 1.2).
Statistics	α – 5% power 1-β – 90% Standard deviation – 9 points on	α – stays the same power 1-β – stays the same **Standard deviation – SD lever**

	ADAS-cog scale for both treatment and placebo arms. Treatment effect – 11.9% change in decline is the aim of current disease-modifying trials	**Treatment effect – TE lever**
Primary endpoint	cognitive & functional improvements	Stays the same
Drop-out rate	23%	Stays the same
Inclusion criteria	Age 50+, Probable AD, MMSE 16-26, caregiver available, women only if postmenopausal. And occasionally also following criteria: MRI not inconsistent with AD, living in community, education, communication level OK, stable symptomatic AD treatment.	**Addition of biomarkers or combination of biomarkers to select a subpopulation.**

This is the basis from which we will try to answer the questions defined in section 1.2:

1a. Theoretically, how good does the biomarker (or combination of biomarkers) need to be to select for a patient sub population that drastically reduces the enrollment needed in Phase 3 by a factor 5x, an important reduction in enrollment that could potentially realistically be achieved?"

1b. In practice and based on data from the ADNI database, how good are existing biomarkers (or combination of biomarkers) in selecting for a patient sub population to reduce the enrollment needed in Phase 3 clinical trials?

1c. What is the cost impact on patient screening and patient enrollment from selecting for a patient sub population in the Phase 3 trial?

1d. What additional considerations are important in order to assess the selected biomarkers and what are the final conclusions on the overall quality of those biomarkers?

1e. How does recent data from AD disease-modifying drugs relate to the findings in this thesis?

Note that we will answer those questions under the two leverage factor scenarios defined under section 2.4:
1) SD lever only,
2) SD & TE levers.

3. Methods

3.1. Analysis of www.clinicaltrials.gov database

3.1.1. Background on the database

ClinicalTrials.gov offers up-to-date information for locating federally and privately supported clinical trials for a wide range of diseases and conditions. ClinicalTrials.gov currently contains 56,702 trials sponsored by the National Institutes of Health, other federal agencies, and private industry. Studies listed in the database are conducted in the 50 US States and in 155 countries. The U.S. National Institutes of Health (NIH), through its National Library of Medicine (NLM), has developed this site in collaboration with the Food and Drug Administration (FDA).

The data from www.clinicaltrials.gov puts the questions we are trying to answer in this thesis (defined in section 1.2) in an applied setting as it is based on real information.

3.1.2. Query and selection of trial protocols in database

The www.clinicaltrials.gov data was viewed online on May 11, 2008.

As discussed in the introduction, the goal is to focus on disease-modifying treatments addressing pathways such as the amyloid beta or tau tangle pathways and on Phase III trial protocols. Therefore, a first screening was done using the following query: Phase "Phase III" & Condition: "Alzheimer" & Study type "Interventional". This resulted in a total of 102 studies.

The second step was to hand pick the disease-modifying treatments only based on an analysis of the disclosed mechanism of action of the drugs. This narrowed the number of studies to 11 for 4 different drugs that target the amyloid beta pathway (bapinuzumab, 3APS, Flurizan and LY450139).

For each of the 11 studies the protocol data was summarized in a table in Appendix 34.

Importantly the number of patients enrolled in each treatment group was derived from this information.

The stated number of enrolled patients was divided by the number of disclosed treatment groups yielding a number of patients per treatment group. 2 of the 11 studies were safety studies so these studies were omitted in the calculation. The number of patients per treatment group for each study was then averaged for the 9 efficacy studies.

This resulted in <u>337 patients enrolled on average per treatment group</u>. This enrollment number is used in the section 2.5.1 and in subsequent sections.

3.2. Statistical sample size model

3.2.1. Choice of statistical sample size model

The choice of outcome variable defines the statistical model. As explained at the end of Section 1.2., the change in cognitive and functional assessment scales is the primary endpoint (and therefore outcome variable) we are focusing on in this thesis. The FDA requires improvement in both cognitive and the functional assessments but they do not define the extent of the treatment effect required.

For the sake of simplicity in our analysis we will only analyze the cognitive part, defined as the change in ADAS-cog score over time. The functional score would follow exactly the same framework of analysis and whichever sample size is higher for cognitive or functional endpoints would be the relevant sample size.

Given our choice of primary endpoint the statistical model required is the *inference on means with unequal variance* model. This is the statistical model used for all trials with the same primary endpoint.[17]

3.2.2. Equation

The sample size equation is : $n = (z_{1-\beta} + z_{1-\alpha})^2 * (\sigma_T^2 + \sigma_P^2) / (\mu_T - \mu_P)^2$

It is extensively discussed in sections 2.3 and 4.1 and appendixes 4-9 show applied calculations using this formula.

3.3. Analysis of ADNI database

3.3.1. ADNI required disclaimer

Data used in the preparation of this thesis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). 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, M.D., VA Medical Center and University of California - San Francisco. ADNI is the result of efforts of many co-investigators 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 adults, ages 55 to 90, to participate in the research -- approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years. For up-to-date information see www.adni-info.org.

The data used for this thesis was downloaded on May 5, 2008 from the ADNI website.

3.3.2. ADNI data available on May 5, 2008

The table below shows all the data collected in the ADNI study. It includes clinical, lab and imaging measurements. The biomarker data highlighted in bold were used in the biomarker regression analysis performed in section 4.2:

Biomarker	Category	Comments	Available data (n) [b]
Diagnosis	**Clinical**	Initial clinical categorization of patients as Normal, MCI or AD	**BL (819)**, M06 (779), M12 (677), M18, M24
MMSE	**Clinical**		**BL (870)**, M06 (782), M12 (705), M24
ADAS-cog	**Clinical**		**BL (819)**, M06 (782), **M12 (703)**, M24
CDR	Clinical		BL, M06, M12, M24
FAQ	Clinical		BL, M06, M12, M24
Neuropsychological Battery	Clinical		BL, M06, M12, M24
Functional assessment questionnaire	Clinical		BL, M06, M12, M24
A beta 1-42 (CSF)	**Lab**	**M12 measurements will be available around year end 2008**	**BL (415[a])**, M12 (N/A)
Tau (CSF)	**Lab**		**BL (415[a])**, M12 (N/A)
P-tau 181P (CSF)	**Lab**		**BL (415[a])**, M12 (N/A)
Tau/Ab1-42	**Lab**		**BL (415[a])**, M12 (N/A)
P-Tau181P/Ab1-42	**Lab**		**BL (415[a])**, M12 (N/A)
ApoE genotype	**Lab**		**BL (1159)**, M12 (N/A)
Family history questionnaire	Clinical		BL
GDS (depression)	Clinical		BL, M12, M24

Modified Hachinski (stroke)	Clinical		BL
MRI volume	Imaging		BL, M06, M12, M24
PET-FDG	Imaging		BL, M06, M12, M24
PET-PiB	Imaging		BL, M06, M12, M18
MRI – brain MRC	Imaging	Several different types of measurements, however lots of missing data so not relevant	BL - 394, M6 - 327, M12 - 161
MRI – Avg Jacobian	Imaging	Avg Jacobian only	BL/F – 658/18, M12 – 100
MRI – UCSD volume	Imaging	12 measurements including brain, ventricules, hippocampus, temporal, fusiform and entorhinal	BL - 336, M6 - 305, M12 – 133
MRI – UCSF sny volume	**Imaging**	**Left and right hippocampus.**	**BL – 457**, M6 - 498, **M12 – 500**
MRI – BSI	Imaging	Brain & ventricular volume.	SC/F/BL – 687/19/128, M6 - 452, M12 – 312
MRI – UAS pm voxel based morphom.	Imaging	117 measurements of all parts of the brain for each subject.	SC/F – 944/5, M6 - 344, M12 – 306
PET – UCB PET ROI (each three imaging panes)	Imaging	glucose metabolism normed to pons, voxel data, mean, SD, etc.	BL - 1416, M6 - 1311, M12 – 1112

BL: baseline visit, M06: month 6 visit; M12: month 12 visit; M18: month 18 visit; M24: month 24 visit; [a] as of 2008-01-28; [b] as of 2008-05-05.

As explained in the previous paragraph, only the highlighted data in bold was used. Several of the above data were not considered because 1) a number of them are clinical data not relevant for the purpose of this thesis, 2) some imaging biomarkers had not enough information or was difficult to analyze, 3) many of the biomarkers did not have a big enough sample of the relevant longitudinal data at 6, 12 or 18 months.

3.3.3. Data processing for final biomarkers

Incorporating these limitations, we selected the following data to include in our subsequent regression analysis (in bold in above table):
- ADAS-cog at BL and M12;
- MMSE at BL;

- MRI right hippocampal volume at BL and M12;
- MRI left hippocampal volume at BL and M12;
- Apoe genotype;
- ABeta 1-42 in CSF at BL;
- Tau in CSF at BL;
- P-Tau181P in CSF at BL.

It is a blend of clinical, lab, genotyping and imaging biomarkers.

We then processed some of the biomarkers:
- we calculated the difference between ADAS-cog scores at BL and M12 and called it "change in ADAS-cog at 12 months".
- Similarly, we took the average between right & left hippocampal volumes. And we defined two biomarkers "Average Hippocampal Volume" and " Average Hippocampal Volume change at 12 months".
- The Apoe genotype information consisted on a number of the allele on each chromosome (genotypes 2, 3 or 4). So we defined two dummy variables: the first consisting of those patients having two copies of the Aoe4 gene, i.e. the aggressive mutation. We called that variable "Both Apoe4 genes". The second biomarker is when one or the other chromosome carries the gene. We called the biomarker "At least one Apoe4 gene".

In the table below we show these processed biomarkers and we show how they compare to the biomarker characteristics defined in Section 2.2:

Biomarker	High predictive power	No longitudinal data needed	Not too costly to perform	Does nor result in restricted label	Not leading to enrollment difficulties
MMSE		√	√	√	√
A beta 1-42		√	???	???	???
Tau		√	???	???	???
P-tau 181P		√	???	???	???
Ratio Tau/Ab1-42		√	???	???	???
Ratio P-Tau181P/ Ab1-42	Depends on patient population (NL, MCI or AD)	√	???	???	???
Both apoe4 genes		√	???	X	√
At least one Apoe4 gene		√	???	X	√
MRI – avg. hippocampal volume		√	???	???	√
MRI – avg. hippocampal volume change		X	X	???	√

√: Not problematic characteristic; X: Problematic characteristic; ???: Uncertain characteristic.

From the initial 819 patients in the database we had to select out many of them in order to have a full set of data with all the biomarkers we wanted to monitor:

- We started off with a subgroup of 416 patients for which ADNI had made lab biomarker data available on January 28th, 2008;
- From those initial 416 patients, we selected out those patients who did not have all the clinical, genotyping and imaging biomarker information.
- This resulted in a sample of 156 patients: 63 normal, 61 MCI and 32 with mild AD.

This sample number should be large enough to draw meaningful conclusions in this thesis. It will however be valuable to apply the framework on bigger samples in the future.

3.3.3. Selected variables for regression analysis

a) <u>Dependant variable</u>: The most important variable is the *outcome variable* which correspond to the primary endpoints of the drug trials, i.e the *change in cognitive and functional assessment scale*. All currently marketed drugs showed improvement on these outcome variables. In addition all currently running disease-modifying trials picked this endpoint. The FDA requires improvement on both the cognitive and the functional assessments but they do not define the extent of the treatment effect required. For the sake of simplicity in our analysis we will only analyze the cognitive part going forward. It is defined as the change in ADAS-cog score over time. The functional score would however follow exactly the same model and whichever sample size is higher for cognitive or functional endpoints would be the relevant sample size.

b) <u>Independent variables</u>:

We will focus on the following biomarkers that are well recognized and are available in the ADNI database:
- Clinical: MMSE
- Lab: A beta 1-42, Tau, P-tau 181P, Ratio Tau/ABeta, Ratio PTau/ABeta.
- Genotyping: Both Apoe4 genes, At least one Apoe4 gene.
- Imaging: Average hippocampal volume, Average hippocampal volume change at 12 months.

3.3.4. Regression analysis
We identify biomarkers that predict cognitive change in patients. We therefore fit a multiple linear regression model to predict the outcome/dependent variable "change in ADAS-cog". We looked at several patient groups and showed which independent variables were most predictive for each group. That helped us to pick the best biomarkers that were then subsequently used to stratify the patients.

3.3.5. Patient stratification based on selected biomarkers

The patient stratification was done either using the dummy variables taking values of 0 or 1 (for genotyping biomarkers) or by defining halves and quartiles (for lab and imaging biomarkers). These were the both the most objective stratification criteria and the simplest ones that we evaluated.

3.3.6. Calculation of effect on sample size of biomarker selected subpopulation

For each of the stratified subpopulations we calculated the mean change in ADAS-cog (μ_P) and its standard deviation (σ_P). These values were then used in the sample size formula under the two leverage factor scenarios defined in section 2.4 (i.e. "SD & TE levers" and "only SD lever").

Since the ADNI database only had enough measurements for the ADAS-cog scores at BL and M12 we could not use the change in ADAS-cog at M18. So we adjusted the μ_P and σ_P values by adding the difference between the value of theses variables for the current design (values at 18 months) and the value of theses variables in the mild AD sample (values at 12 months).

When combinations of biomarkers were used, we prepared for the individual biomarkers the list of the patients we would select out ("select-out lists"). From these "select-out lists" we selected out any patients that appeared at least once in the lists of the biomarkers we want to combine. See sections 4.4,1 and 4.4.2 for examples of how this is applied.

3.4. Cost model

The cost model uses the following simplified formula:

Trial Cost	=	[Trial Cost Per Patient] x [# Enrolled Patients] x [# Treatment groups]
Screening Cost	=	[Sum of relevant biomarker testing costs + Cost of Physician visit] x [# Enrolled Patients x Screen to Enroll Ratio / (1- Subpopulation enroll reduction)] x [# Treatment groups]
Net Cost of Trial	=	Trial Cost + Screening Cost
Cost Savings	=	Net Cost of Trial (current) – Net Cost of Trial (biom. selected subpopulation)

The goal is to monitor the Cost Savings from the biomarker selected subpopulation trial design.

Using cost assumptions that were verified with industry executives, the Net Cost of current disease-modifying AD Trials was calculated.

Following that, we calculated the Cost of Trial for each scenario of biomarker selected subpopulation. We used two key results from section 4.4. The first is the "# Enrolled patients" calculated for each subpopulation. The second is the "Subpopulation enroll reduction" which defines which percent of the initial patient population is actually enrolled. These impacted both the Trial Cost and the Screening Cost.

The Screening Cost for each scenario of biomarker selected subpopulation was further affected by which biomarker the patient subpopulation was selected with, adding if relevant the individual cost of the biomarker for each screened patient.

In section 4.5, the cost model is applied and the assumptions are explained.

4. Results

4.1. Theoretical sample size calculations

The theoretical sample size calculations illustrate what effect a good biomarker needs to have to select a subpopulation that in turn will affect the variables in the sample size formula.

$$n = (z_{1-\beta} + z_{1-\alpha})^2 * \underbrace{(\sigma_T^2 + \sigma_P^2)}_{\text{SD lever}} / \underbrace{(\mu_T - \mu_P)^2}_{\text{TE lever}}$$

We have attached in Appendix 5 the calculation sheet of the statistical model the current 18 month disease modifying Alzheimer's Disease trials. Using that as the starting point we looked at what changes in the variables would allow a 5-fold decrease in enrollment.

We assume that is was powered to show statistical significance with 337 patients based on prior calculations explained in sections 3.1.2.

1) Only SD lever: The following figure shows the relationship between the number of enrolled patients and the standard deviation. In other words, the graph shows: *by how many ADAS-cog points the standard deviation needs to decrease in order to reach a 5x reduced sample size and assuming that there is no increase in treatment effect?*

In the graph below, the starting point is a 9 point standard deviation in current disease-modifying trials corresponding to 337 enrolled patients (see red vertical bar). The standard deviation needs to be reduced to approximately 4 points in order to have the desired effect of a 5-fold decrease in enrollment.

**Relationship between Enrollment and
Standard Deviation (only SD lever)**
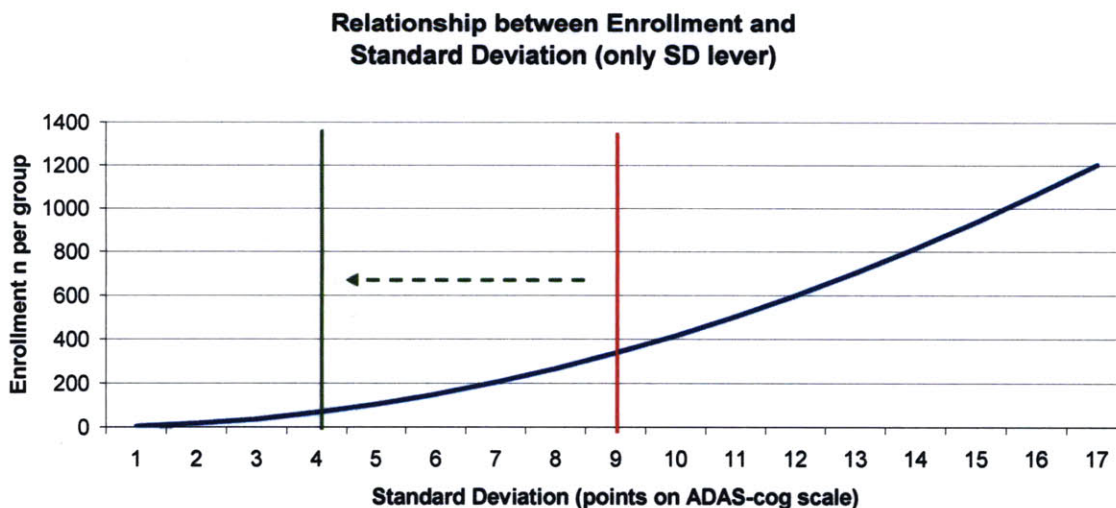
2) Only TE lever: The next figure shows the relationship between the number of enrolled patients and the treatment effect. In other words, the graph shows *by how many % points the treatment effect needs to increase in order to reach a 5-fold reduced sample size and assuming that there is no decrease in standard deviation?*

In the graph below, we assume a scenario where the change of the baseline treatment effect is 11.9% and that the treatment effect needs to be increased in order to reduce sample size. The green arrows illustrate that. As expected we have a bell shaped curve where the smaller the treatment effect the higher the number of patients you need to enroll.

Relationship between Enrollment and Treatment Effect in % (only TE lever)
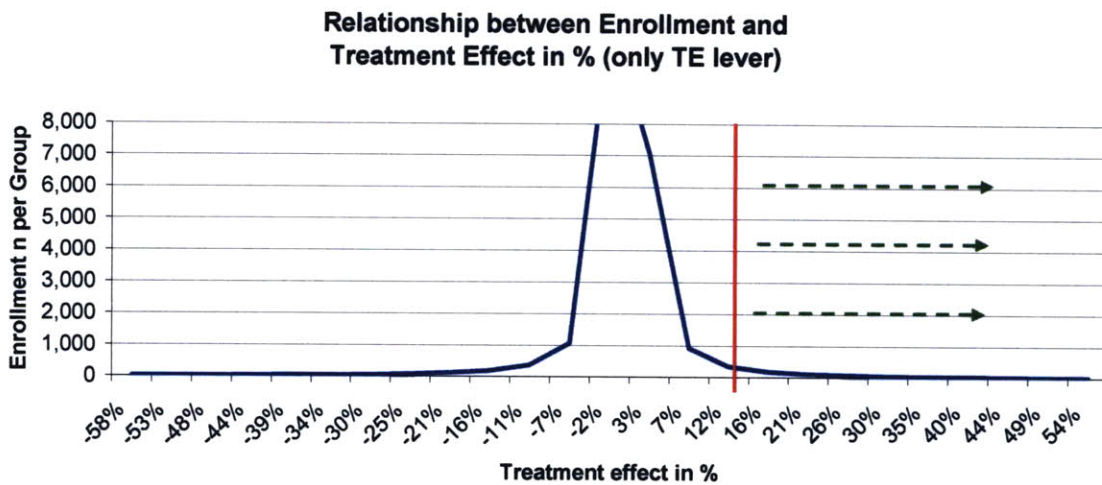
The next graph is a close up of the right side of the bell shaped curve. The starting point is the red vertical bar at a 11.9% treatment effect and we see that we need to increase it to approximately 27% in order to obtain a 5-fold decrease in enrollment.

Relationship between Enrollment and Treatment Effect in % (only TE lever)
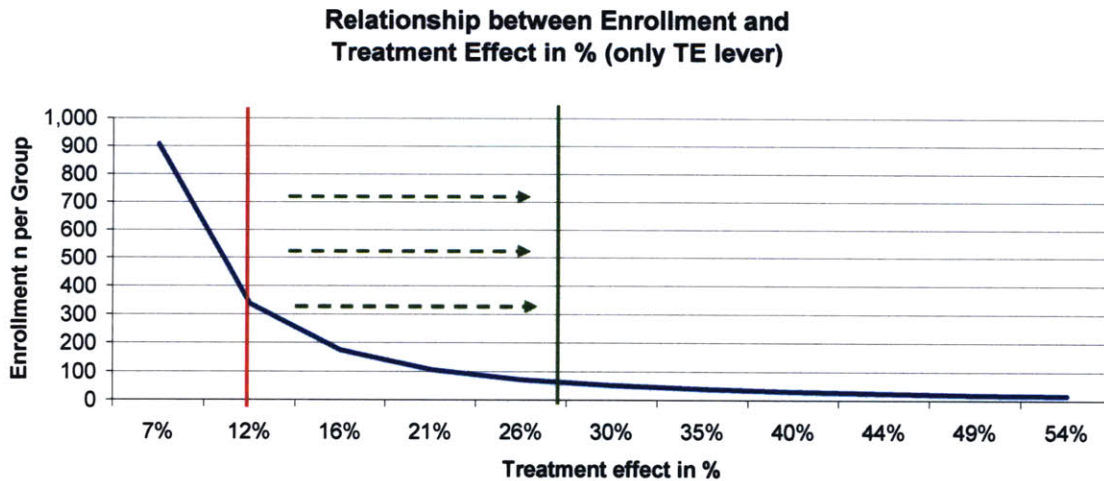
In summary, this theoretical sample size calculation analysis shows that in order to reach the stretch goal of a 5-fold decrease in enrollment you need to:
- *reduce the standard deviations from 9 to 4 ADAS-cog points; or*
- *increase your treatment effect from 11.9% to 27%.*

The table below shows different combinations of Standard Deviation and Treatment Effect that all give a 5-fold decrease in sample size and also illustrate which leverage factor scenarios they correspond to:

Standard Deviation	Treatment effect	Leverage factor scenario
4.0 ADAS-cog points	11.9%	Only SD lever
5.0 ADAS-cog points	15.0%	SD & TE levers
6.8 ADAS-cog points	20.0%	SD & TE levers
8.5 ADAS-cog points	25.0%	SD & TE levers
9.0 ADAS-cog points	26.7%	Only TE lever

4.2. Regression analysis of ADNI data to select predictive biomarkers

The goal of the regression analysis is to single out those biomarkers that have a stronger predictive power in predicting change in ADAS-cog.

In order to perform the regression analysis we picked the ADAS-cog-change in score at 12 months as the dependent variable. We would have preferred to use the ADAS-cog at 18 months but it would have drastically reduced our sample due to unavailable data.

We included the following biomarkers in our regression analysis:
- at baseline: MMSE score, CSF total tau, CSF phosphorylated tau, CSF a-beta 1-42, both Apoe4 genes, atleast one Apoe4 gene, average hippocampal volume.
- Change from baseline to month 12: average hippocampal volume.

We have a complete set of this data for 156 patients out of the 819 total population in the ADNI database.

4.2.1. Regression results

We performed the regressions on the following patient groups:
- All patients in the sample (n=156)
- Only normal patients (n=63)
- Only MCI patients (n=61)
- Only mild AD patients (n=32)

SUMMARY OUTPUT **ALL PATIENT GROUPS (n=156)**

Regression Statistics	
Multiple R	0.45
R Square	0.20
Adjusted R Square	0.14
Standard Error	4.99
Observations	156

ANOVA

	df	SS	MS	F	Significance F
Regression	10	895.02	89.50	3.60	0.00
Residual	145	3604.58	24.86		
Total	155	4499.59			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	20.09	5.25	3.83	0.00	9.72	30.47	9.72	30.47
MMSE	-0.58	0.18	-3.15	0.00	-0.95	-0.22	-0.95	-0.22
Tau	0.03	0.05	0.54	0.59	-0.07	0.12	-0.07	0.12
Ab1-42	0.00	0.01	-0.12	0.91	-0.03	0.02	-0.03	0.02
P-Tau 181P	-0.10	0.16	-0.66	0.51	-0.42	0.21	-0.42	0.21
Tau/Ab1-42	-2.03	6.21	-0.33	0.74	-14.31	10.25	-14.31	10.25
P-Tau181P/Ab1-42	10.87	21.16	0.51	0.61	-30.95	52.70	-30.95	52.70
Both APOE4 genes	1.85	1.42	1.31	0.19	-0.95	4.66	-0.95	4.66
At least one APOE4 genes	-0.64	1.02	-0.63	0.53	-2.67	1.38	-2.67	1.38
AVG HIPPO VOL	0.00	0.00	-1.25	0.21	0.00	0.00	0.00	0.00
AVG HIPPO VOL change	0.00	0.00	0.82	0.42	0.00	0.01	0.00	0.01

For the whole sample, the MMSE score, the average hippocampal volume at baseline as well as both Apoe4 genes seem predictive of the change of ADAS-cog score at 12 months. The predictive power of the MMSE score can logically be explained asboth ADAS-cog and MMSE are relatively closely related cognitive assessment scales. The model however only explains 20% of the variation (based on R squared).

SUMMARY OUTPUT **NORMAL PATIENT GROUP (n=63)**

Regression Statistics	
Multiple R	0.37
R Square	0.13
Adjusted R Square	-0.03
Standard Error	3.41
Observations	63

ANOVA

	df	SS	MS	F	Significance F
Regression	10	93.63	9.36	0.81	0.62
Residual	52	603.93	11.61		
Total	62	697.57			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-13.75	16.01	-0.86	0.39	-45.88	18.37	-45.88	18.37
MMSE	0.36	0.52	0.69	0.49	-0.69	1.41	-0.69	1.41
Tau	0.07	0.07	0.96	0.34	-0.07	0.20	-0.07	0.20
Ab1-42	0.00	0.02	0.17	0.87	-0.03	0.03	-0.03	0.03
P-Tau 181P	-0.14	0.18	-0.80	0.43	-0.50	0.22	-0.50	0.22
Tau/Ab1-42	-6.06	10.91	-0.55	0.58	-27.96	15.85	-27.96	15.85
P-Tau181P/Ab1-42	10.15	29.35	0.35	0.73	-48.74	69.03	-48.74	69.03
Both APOE4 genes	-3.98	3.74	-1.07	0.29	-11.48	3.52	-11.48	3.52
At least one APOE4 genes	0.64	1.20	0.53	0.60	-1.78	3.06	-1.78	3.06
AVG HIPPO VOL	0.00	0.00	0.62	0.54	0.00	0.00	0.00	0.00
AVG HIPPO VOL change	0.00	0.00	1.17	0.25	0.00	0.01	0.00	0.01

None of the biomarkers were predictive in the normal group and the R squared was very low (13%).

SUMMARY OUTPUT **MCI PATIENT GROUP (n=61)**

Regression Statistics	
Multiple R	0.39
R Square	0.15
Adjusted R Square	-0.02
Standard Error	4.91
Observations	61

ANOVA

	df	SS	MS	F	Significance F
Regression	10	213.52	21.35	0.88	0.55
Residual	50	1207.56	24.15		
Total	60	1421.07			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	23.45	10.49	2.24	0.03	2.38	44.53	2.38	44.53
MMSE	-0.74	0.39	-1.89	0.06	-1.53	0.05	-1.53	0.05
Tau	-0.08	0.08	-0.98	0.33	-0.24	0.08	-0.24	0.08
Ab1-42	0.00	0.02	-0.15	0.88	-0.05	0.04	-0.05	0.04
P-Tau 181P	0.31	0.29	1.07	0.29	-0.27	0.88	-0.27	0.88
Tau/Ab1-42	9.68	10.09	0.96	0.34	-10.59	29.94	-10.59	29.94
P-Tau181P/Ab1-42	-34.61	37.37	-0.93	0.36	-109.66	40.44	-109.66	40.44
Both APOE4 genes	-0.66	2.26	-0.29	0.77	-5.19	3.88	-5.19	3.88
At least one APOE4 genes	-0.40	1.65	-0.24	0.81	-3.71	2.92	-3.71	2.92
AVG HIPPO VOL	0.00	0.00	-0.95	0.34	-0.01	0.00	-0.01	0.00
AVG HIPPO VOL change	0.00	0.00	0.64	0.52	-0.01	0.01	-0.01	0.01

In the MCI group, only the MMSE variable showed predictive power. R Squared was only 15%.

SUMMARY OUTPUT **AD PATIENT GROUP (n=32)**

Regression Statistics	
Multiple R	0.69
R Square	0.47
Adjusted R Square	0.22
Standard Error	6.38
Observations	32

ANOVA

	df	SS	MS	F	Significance F
Regression	10	758.55	75.85	1.87	0.11
Residual	21	853.87	40.66		
Total	31	1612.42			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	18.17	23.95	0.76	0.46	-31.65	67.98	-31.65	67.98
MMSE	0.96	0.90	1.07	0.30	-0.90	2.82	-0.90	2.82
Tau	0.38	0.22	1.71	0.10	-0.08	0.85	-0.08	0.85
Ab1-42	-0.18	0.08	-2.25	0.04	-0.35	-0.01	-0.35	-0.01
P-Tau 181P	-0.99	0.64	-1.55	0.14	-2.31	0.34	-2.31	0.34
Tau/Ab1-42	-45.16	29.21	-1.55	0.14	-105.91	15.59	-105.91	15.59
P-Tau181P/Ab1-42	112.09	81.48	1.38	0.18	-57.35	281.53	-57.35	281.53
Both APOE4 genes	0.51	3.24	0.16	0.88	-6.23	7.25	-6.23	7.25
At least one APOE4 genes	-4.48	3.21	-1.39	0.18	-11.15	2.20	-11.15	2.20
AVG HIPPO VOL	0.00	0.00	-0.51	0.61	-0.01	0.01	-0.01	0.01
AVG HIPPO VOL change	0.01	0.01	0.60	0.55	-0.02	0.03	-0.02	0.03

Both the normal and MCI patient regressions were poorly predictive. In contrast, in the AD group the model has a 47% R squared and the A-beta 1-42 at baseline was

statistically significant. The biomarkers Tau, Phospho tau, tau/a-beta ratios and at least one Apoe4 gene were also close to significance (t values above 1.3).

A preliminary conclusion is that the biomarkers analyzed are more appropriate to explain ADAS-cog changes for the AD group and not for the Normal and MCI group. Looking closer at individual independent variables for the AD patient group and their predictive power and R squared to predict ADAS-cog change at 12M, we observe the following regression outcomes (sorted by predictive power):

Independent variable	P value	R Squared
Abeta 1-42	0.014	18.4%
Avg hippocampal volume	0.086	9.5%
Both Apoe4 genes	0.105	8.5%
Tau/Abeta 1-42	0.126	7.6%
P-tau 181P/ Abeta 1-42	0.457	1.9%
Tau	0.476	1.7%
Avg hippocampal volume change at 12M	0.583	1.0%
MMSE BL	0.959	0.0%
P-tau 181P	0.923	0.0%
At least one Apoe4 gene	0.802	0.0%
ALL VARIABLES TOGETHER		**48.6%**

Based on this breakdown the four biomarkers with the strongest predictive power are Abeta 1-42 (18.4%), Average hippocambal volume (9.5%), Both Apoe4 genes (8.5%) and the ratio Tau/A-beta1-42 (7.6%). Together they predict 44% of the change in ADAS-cog. They are highlighted in grey background in the above table.

For the stratification in section 4.3 we will only use these four biomarkers.

4.2.2. Analysis of regression equation

For the AD patient group the linear multiple regression equation is:

Predicted ADAS-cog change = 18.174 + 0.96*MMSE + 0.38*Tau – 0.18*Ab1-42 – 0.99*Ptau181P – 45.16*Ab1-42 + 112.09*PTau181P/Ab1-42 + 0.51*BothApoe4Genes – 4.48*AtLeastOneApoe4Gene – 0.0022*AcgHippoVol + 0.0067*AvgHippoVolChange

This equation is predicting the change in ADAS-cog for mild AD patients as those found in the sample we have analyzed (n=32).

Effectively it means that for the group of mild AD patients the average of the difference in points between the change in ADAS-cog points predicted by the equation and the actual change in ADAS-cog points is zero. The standard deviation of this difference in points is 5.2 points.

To support that this equation is specific for the AD patient group, we also applied the formula on normal and MCI groups and as expected the formula was much poorer at predicting the outcome variable:
- when applied to the independent sample of <u>normal patients</u> in the ADNI database (n=63), the average is <u>-3.3 points</u> and the standard deviation <u>9.9 points</u>.
- when applied to the independent sample of <u>MCI patients</u> (n=61), the average is <u>-6.4 points</u> and the standard deviation <u>8.3 points</u>. Data is not shown but can be provided on request.

This is promising as the predictive equation clearly delineates AD, MCI and normal patients. However it has several caveats:
- although the mean is zero for the mild AD group, the standard deviation of the ideal equation should also be zero if the model is 100% predictive.
- it would be even better to have an independent sample of AD patients on which to apply the equation, as it would show how well it holds against a sample of patients with similar symptoms.
- There are mild, moderate and severe AD patients. So it would also be worthwhile to see if the same equation is predictive when applied to a group of moderate or severe AD patients. This is beyond the scope of this thesis.

4.3. Stratification of patient subpopulation based on selected biomarkers

When analyzing our 32 mild AD patient group data we saw that the mean and standard deviation of the change in ADAS-cog at 12 months were 5.4 points and 7.2 points.

Now we want to define a sub population from the 32 AD patient sample that either has a smaller standard deviation or a larger mean. As you recall from our analysis in section 4.1, a smaller standard deviation or a larger treatment effect (helped by a large mean change in the placebo group as we are trying to define it in this paragraph) both reduce the sample size needed to power your clinical trial.

In section 4.2.1., we selected out four biomarkers that help explain a change in ADAS-cog at 12 months:
- both Apoe4 genes,
- Abeta 1-42,
- ratio Tau/A-beta1-42, and
- average hippocampal volume.

In the next steps we are stratifying our patient group based on those biomarkers.

Note that we will analyze the data under the two leverage factors scenarios defined under section 2.4:
1) SD & TE levers, where μ_T remains unchanged,
2) Only SD lever, where $\mu_T - \mu_P$ remains unchanged.

4.3.1. Stratification for "SD & TE levers" scenario

Under this scenario, any stratification leading to a treatment effect below 10.4% was considered "Not Applicable" because it would mean that the drug is inferior in efficacy than marketed drug Aricept donepezil.

A) Analysis of effect of "both Apoe4 genes" biomarker:
In our 32 patient sample, 10 out of 32 patients (31.3%) have both Apoe4 mutations. If we select only those patients having both Apoe4 genes, the mean and standard deviations are 8.5 points and 8.7 points. If we select those patients not having both Apoe4 genes, the mean and standard deviations are 4.0 points and 6.1 points (see Appendix 6 for calculations).

If you input this into our theoretical model, the following impact on treatment effect and sample size can be observed:

	SD & TE levers scenario: assumes μ_T remains unchanged			
	μ_P	σ_P	Treatment effect in %	Corresponding Sample size
Baseline	5.4 points	7.2 points	11.9%	337 patients
Both Apoe4 genes (n=10)	8.5 points	8.7 points	26.2%	**94 patients**
Not both Apoe4 genes (n=22)	4.0 points	6.1 points	5.4% (N/A)	N/A

To illustrate, the stratification that yielded a 5.4% treatment effect was "not applicable" as the treatment effect is below the treatment effect of currently approved drugs (10.4%). Any drug with a poorer treatment effect than 10.4% would not be approved by the FDA. This same reasoning will also be applied in the subsequent tables for the stratification of the other biomarkers.

B) Analysis of effect of "A-beta 1-42baseline" biomarker:
In our 32 AD patient sample, the A-beta concentration in CSF ranges from 81.6 to 283.8 with an average of 145.5 and a median of 137.8. The four quartiles are: 129.7, 137.8, 160.7 and 283.8. We split the patients into these four quartiles and looked at their respective averages and standard deviations for the ADAS-cog change at month 12 (see Appendix 7 for calculations).

If you input this into our theoretical model, the following impact on treatment effect and sample size can be observed:

	SD & TE levers scenario: assumes μ_T remains unchanged			
	μ_P	σ_P	Treatment effect in %	Corresponding Sample size
Baseline	5.4 points	7.2 points	11.9%	337 patients
Quartile 1 (values 81.6 – 129.6, n=8)	8.6 points	9.9 points	26.7%	**113 patients**
Quartile 2 (values 129.7 – 137.8, n=8)	6.9 points	3.9 points	18.8%	**54 patients**
Quartile 3 (values 137.9 – 160.7, n=8)	6.1 points	7.2 points	15.1%	**208 patients**
Quartile 4 (values 160.8 – 283.8, n=8)	0.1 points	4.3 points	-12.7% (N/A)	**N/A**
Half 1 (values 81.6 – 137.8, n=16)	7.8 points	7.3 points	23.0%	**92 patients**
Half 2 (values 137.9 – 283.8, n=16)	3.1 points	6.5 points	1.2% (N/A)	**N/A**

C) Analysis of effect of "the ratio Tau/A-beta 1-42" biomarker:

The quartiles for the biomarker "ratio tau/Abeta" are 0.58, 0.87, 1.07 and 2.59.
We split the patients into these four quartiles and looked at their respective averages and standard deviations for the ADAS-cog change at month 12 (see Appendix 8 for calculations).

If you input this into our theoretical model, the following impact on treatment effect and sample size can be observed:

	SD & TE levers scenario: assumes μ_T remains unchanged			
	μ_P	σ_P	Treatment effect in %	Corresponding Sample size
Baseline	5.4 points	7.2 points	11.9%	337 patients
Quartile 1 (values 0.22 – 0.56, n=8)	1.1 points	4.5 points	-8.1% (N/A)	**N/A**
Quartile 2 (values 0.60 – 0.84, n=8)	5.1 points	7.2 points	10.5%	**432 patients**
Quartile 3 (values 0.85 – 1.05, n=8)	9.3 points	10.3 points	29.9%	**96 patients**
Quartile 4 (values 1.14 – 2.59, n=8)	6.2 points	3.5 points	15.6%	**68 patients**
Half 1 (values 0.22-0.84, n=16)	3.1 points	6.1 points	1.2% (N/A)	**N/A**
Half 2 (values 0.85-2.59, n=16)	7.8 points	7.6 points	23.0%	**98 patients**

D) Analysis of effect of "the average hippocampal volume" biomarker:
The quartiles for the biomarker "average hippocampal volume" are 1375, 1604, 1789 and 2370. We split the patients into these four quartiles and looked at their respective

averages and standard deviations for the ADAS-cog change at month 12 (see Appendix 9 for calculations).

If you input this into our theoretical model, the following impact on treatment effect and sample size can be observed:

	SD & TE levers scenario: assumes μ_T remains unchanged			
	μ_P	σ_P	Treatment effect in %	Corresponding Sample size
Baseline	5.4 points	7.2 points	11.9%	337 patients
Quartile 1 (values 1082 – 1375, n=8)	9.4 points	9.8 points	30.4%	**85 patients**
Quartile 2 (values 1376 – 1604, n=8)	3.8 points	5.9 points	4.4% (N/A)	N/A
Quartile 3 (values 1605 – 1789, n=8)	3.1 points	2.9 points	1.2% (N/A)	N/A
Quartile 4 (values 1790 – 2370, n=8)	5.4 points	8.0 points	11.9%	**400 patients**
Half 1 (values 1082-1604, n=16)	6.6 points	8.3 points	17.4%	**151 patients**
Half 2 (values 1605-2370, n=16)	4.3 points	5.9 points	6.8% (N/A)	N/A

4.3.2. Stratification for "Only SD lever" scenario

Please note also that, when working under this assumption, any stratification leading to a treatment effect below 10.4% was considered "Not Applicable" because it would mean that the drug is inferior in efficacy than marketed drug Aricept donepezil.

A) Analysis of effect of "both Apoe4 genes" biomarker:
In our 32 patient sample, 10 out of 32 patients (31.3%) have both Apoe4 mutations. If we select only those patients having both Apoe4 genes, the mean and standard deviations are 8.5 points and 8.7 points. If we select those patients not having both Apoe4 genes, the mean and standard deviations are 4.0 points and 6.1 points (see Appendix 6 for calculations).

If you input this into our theoretical model, the following impact on treatment effect and sample size can be observed:

	Only SD lever scenario: assumes μ_T - μ_P remains unchanged			
	μ_P	σ_P	Treatment effect in %	Corresponding Sample size
Baseline	5.4 points	7.2 points	11.9%	337 patients
Both Apoe4 genes (n=10)	8.5 points	8.7 points	11.9%	**459 patients**
Not both Apoe4 genes (n=22)	4.0 points	6.1 points	11.9%	**260 patients**

Watching the recent Phase 2 data published by Wyeth/Elan on July 29th 2008 for their disease-modifying drug bapineuzumab (see section 5.4 for more details on that), the drug performed better on Apoe4 non carriers than on Apoe4 carriers (we assume that Apoe4 carriers means they have both Apoe4 genes).

That is consistent with our analysis above that Apoe4 carriers need more patients to power a trial.

B) Analysis of effect of "A-beta 1-42baseline" biomarker:
In our 32 AD patient sample, the A-beta concentration in CSF ranges from 81.6 to 283.8 with an average of 145.5 and a median of 137.8. The four quartiles are: 129.7, 137.8, 160.7 and 283.8. We split the patients into these four quartiles and looked at their respective averages and standard deviations for the ADAS-cog change at month 12 (see Appendix 7 for calculations).

If you input this into our theoretical model, the following impact on treatment effect and sample size can be observed:

	Only SD lever scenario: assumes $\mu_T - \mu_P$ remains unchanged			
	μ_P	σ_P	Treatment effect in %	Corresponding Sample size
Baseline	5.4 points	7.2 points	11.9%	337 patients
Quartile 1 (values 81.6 – 129.7, n=9)	8.6 points	9.9 points	11.9%	**351 patients**
Quartile 2 (values 129.8 – 137.8, n=7)	6.9 points	3.9 points	11.9%	**256 patients**
Quartile 3 (values 137.9 – 160.7, n=8)	6.1 points	7.2 points	11.9%	**337 patients**
Quartile 4 (values 160.8 – 283.8, n=8)	0.1 points	4.3 points	11.9%	**155 patients**
Half 1 (values 81.6 – 137.8, n=16)	7.8 points	7.3 points	11.9%	**345 patients**
Half 2 (values 137.9 – 283.8, n=16)	3.1 points	6.5 points	11.9%	**287 patients**

C) Analysis of effect of "the ratio Tau/A-beta 1-42" biomarker:

The quartiles for the biomarker "ratio tau/Abeta" are 0.58, 0.87, 1.07 and 2.59.
We split the patients into these four quartiles and looked at their respective averages and standard deviations for the ADAS-cog change at month 12 (see Appendix 8 for calculations).

If you input this into our theoretical model, the following impact on treatment effect and sample size can be observed:

	Only SD lever scenario: assumes $\mu_T - \mu_P$ remains unchanged			
	μ_P	σ_P	Treatment effect in %	Corresponding Sample size
Baseline	5.4 points	7.2 points	11.9%	337 patients
Quartile 1 (values 0.22 – 0.56, n=8)	1.1 points	4.5 points	11.9%	**255 patients**
Quartile 2 (values 0.57 – 0.87, n=8)	5.1 points	7.2 points	11.9%	**312 patients**
Quartile 3 (values 0.88 – 1.07, n=8)	9.3 points	10.3 points	11.9%	**609 patients**
Quartile 4 (values 1.08 – 2.59, n=8)	6.2 points	3.5 points	11.9%	**117 patients**
Half 1 (values 0.22-0.84, n=16)	3.1 points	6.1 points	11.9%	**260 patients**
Half 2 (values 0.85-2.59, n=16)	7.8 points	7.6 points	11.9%	**368 patients**

D) Analysis of effect of "the average hippocampal volume" biomarker:
The quartiles for the biomarker "average hippocampal volume" are 1375, 1604, 1789 and 2370. We split the patients into these four quartiles and looked at their respective averages and standard deviations for the ADAS-cog change at month 12 (see Appendix 9 for calculations).

If you input this into our theoretical model, the following impact on treatment effect and sample size can be observed:

	Only SD lever scenario: assumes $\mu_T - \mu_P$ remains unchanged			
	μ_P	σ_P	Treatment effect in %	Corresponding Sample size
Baseline	5.4 points	7.2 points	11.9%	337 patients
Quartile 1 (values 1082 – 1375, n=8)	9.4 points	9.8 points	11.9%	**560 patients**
Quartile 2 (values 1376 – 1604, n=8)	3.8 points	5.9 points	11.9%	**247 patients**
Quartile 3 (values 1605 – 1789, n=8)	3.1 points	2.9 points	11.9%	**92 patients**
Quartile 4 (values 1790 – 2370, n=8)	5.4 points	8.0 points	11.9%	**400 patients**
Half 1 (values 1082-1604, n=16)	6.6 points	8.3 points	11.9%	**424 patients**
Half 2 (values 1605-2370, n=16)	4.3 points	5.9 points	11.9%	**247 patients**

4.4. Sample size impact of biomarker selected patient subpopulation

In the previous section we stratified our patient population based on dummy variables or quartiles and halves.

The final step is now to use a rationale for selecting a subpopulation. We will observe different criteria for selecting the subpopulation depending on the basic assumptions we take.

Note that we will do that analysis separately for the two leverage factors scenarios defined under section 2.4:
> *1) SD & TE levers, where μ_T remains unchanged,*
> *2) Only SD lever, where μ_T - μ_P remains unchanged.*

4.4.1. Biomarker subpopulation selection criteria for "SD & TE levers" scenario

The following table summarizes the findings from section 4.3.1:

Biomarker	Biomarker subpopulation selection criteria	Patients included out of 32 AD patients	Corresponding sample size
Apoe4 genotype	Carrier of both Apoe4 genes	10	94 patients
ABeta 1-42 CSF concentration	**Concentration < 161**	**24**	**150 patients**
Ratio Tau/ABeta1-42	**Ratio > 0.84**	**16**	**98 patients**
Avg. Hippocampal Volume	Volume < 1604	16	151 patients

From a quick glance at this table we see that the criteria for the ABeta 1-42 biomarker has a very strong profile, reducing the subpopulation by only 25% from the initial 32 and the sample size down by a factor of 2.2x. Also very promising is the Ratio Tau/Abeta biomarker which reduces the subpopulation by 50% but the sample size by a factor of 3.4x.

We then looked at the combination of these superior biomarkers to see if they yield an even stronger profile:

Biomarker / combination of biomarkers	Biomarker subpopulation selection criteria	Selected out
ABeta 1-42 CSF concentration	Concentration < 161	404, 535, 543, 547, 784, 836, 1109, 1171 (total 8)
Ratio Tau/ABeta1-42	Ratio > 0.84	366, 372, 404, 426, 535, 547, 619, 753, 784, 836, 850, 891, 1041, 1082, 1109, 1171 (total 16)
Combination ABeta and Ratio tau/Abeta biomarkers		**366, 372, 404, 426, 535, 543, 547, 619, 753, 784, 836, 850, 891, 1041, 1082, 1109, 1171 (total 17 selected out)**

Using the subpopulation of 15 patients we get to a mean ADAS-cog change of 8 and a standard deviation of 7.8. This subpopulation of 15 patients (down 53%) reduced the required sample size to 94 patients (down from 337 with whole population) or a factor of 3.6x.

4.4.2. Biomarker subpopulation selection criteria for "Only SD lever" scenario

The following table summarizes the findings from section 4.3.2:

Biomarker	Biomarker subpopulation selection criteria	Patients included out of 32 AD patients	Corresponding sample size
Apoe4 genotype	**Non-carrier of both Apoe4 genes**	**22**	**260 patients**
ABeta 1-42 CSF concentration	Concentration > 138	16	287 patients
Ratio Tau/ABeta1-42	Ratio < 0.56 or > 1.08	16	176 patients
Avg. Hippocampal Volume	**Volume > 1376 and < 1789**	**16**	**151 patients**

From a quick glance at this table we see that the criteria for the Non-carrier of both Apoe4 genes has a promising profile, reducing the subpopulation by only 31% from the initial 32 and the sample size down by a factor of 1.3x. Also very promising is the Average Hippocampal Volume biomarker which reduces the subpopulation by 50% but the sample size by a factor of 2.2x.

We then looked at the actual data to define the subpopulation for the combination of biomarkers:

Biomarker / combination of biomarkers	Biomarker subpopulation selection criteria	Selected out
Apoe4 genotype	Non-carrier of both Apoe4 genes	474, 565, 577, 606, 619, 627, 733, 753, 1041, 1082
Ratio Tau/ABeta1-42	Ratio < 0.56 or > 1.08	366, 404, 431, 474, 517, 535, 565, 724, 753, 754, 850, 852, 1041, 1082, 1109, 1221
Avg. Hippocampal Volume	Volume > 1376 and < 1789	366, 372, 404, 426, 517, 535, 565, 577, 606, 724, 733, 814,852, 891, 1041, 1171
Combination Apoe4 non carrier and Ratio Tau/ABeta		**366, 404, 431, 474, 517, 535, 565, 577, 606, 619, 627, 724, 733, 753, 754, 850, 852, 1041, 1082, 1109, 1221 (total 21 patients selected out)**
Combination Apoe4 non carrier and Avg hippocampal volume		**366, 372, 404, 426, 474, 517, 535, 565, 577, 606, 619, 627, 724, 733, 753, 814, 852, 891, 1041, 1082, 1171 (total 21 patients selected out)**

Using the subpopulation of 11 patients with the combination of Apoe4/Ratio biomarker we get to a mean ADAS-cog change of 2.3 and a standard deviation of 4.5. This subpopulation of 11 patients (down <u>65%</u>) reduced the required sample size to 166 patients (down from 337 with whole population) or a factor of <u>2.0x</u>.

Using the subpopulation of 11 patients with the combination of Apoe4/Hippocampakl Volume biomarker we get to a mean ADAS-cog change of 2.2 and a standard deviation of 4.0. This subpopulation of 11 patients (down <u>65%</u>) reduced the required sample size to 142 patients (down from 337 with whole population) or a factor of <u>2.4x</u>.

4.4.3. Biomarker subpopulation selection criteria for "Universal biomarkers"

The following table summarizes the findings from section 4.3.2:

Biomarker	Biomarker subpopulation selection criteria	Patients included out of 32 AD patients	Corresponding sample size
Ratio Tau/ABeta1-42	**Ratio > 1.08**	**8**	**117 patients**
Avg. Hippocampal Volume	Volume > 1376 and < 1604	8	247 patients

Given that the groups that overlapped were small (n=8) we cannot attempt to do a combination of biomarkers in this instance as the patient number would become too

small. So we select only the ratio Tau/ABeta as the best "universal" biomarker to select a subpopulation.

Using the subpopulation of 8 patients (down by <u>75%</u> from the initial 32 AD patients) with a ratio above 1.08, we reduce the required sample size to 117 patients (down from 337 with whole population) or a factor of <u>2.9x</u>.

4.4.4. Summary of sample size impact of biomarker selected patient subpopulation

The following table summarizes our findings:

Leverage factors scenario	Biomarker subpopulation selection criteria	Subpopulation reduction	Required Sample Size	Sample Size reduction
SD & TE levers	ABeta 1-42 < 161	25%	150 pts	2.2 x
	Ratio Tau/ABeta > 0.84	50%	98 pts	3.4 x
	<u>Combi 1</u>: ABeta 1-42 < 161+ Ratio Tau/ABeta > 0.84	54%	94 pts	3.6 x
Only SD lever	Apoe4 Non carriers	31%	260 pts	1.3 x
	Hippocampal Volume between > 1376 & < 1789	50%	151 pts	2.2 x
	<u>Combi 2</u>: Apoe4 non carriers + Ratio Tau/ABeta < 0.56 or > 1.08	65%	166 pts	2.0 x
	<u>Combi 3</u>: Apoe4 non carriers + Hippocampal Volume Volume > 1376 and < 1789	65%	142 pts	2.4 x
Universal	Ratio Tau/ABeta > 1.08 (SD & TE levers)	75%	68 pts	5.0 x
	Ratio Tau/ABeta > 1.08 (only SD lever)	75%	117 pts	2.9 x

In conclusion, we see here that the biomarker or combination of biomarkers a sponsor would use to select a subpopulation to enroll in the clinical trial varies depending on the leverage factor scenario you fall under.

The universal biomarker (the ratio Tau/Abeta) is also very promising but it requires a large reduction in the subpopulation.

4.5. Cost impact of biomarker selected patient subpopulation

In this section we propose a simple model evaluating the extra cost of screening more patients and compare it with the benefit of enrolling less patients in the trial.

In a first step we define the cost drivers for the screening as well as of the 18 month trial:

Trial Cost = [Trial Cost Per Patient] x [# Enrolled Patients] x [# Treatment groups]

Screening Cost = [Sum of relevant biomarker testing costs + Cost of Physician visit] x [# Enrolled Patients x Screen to Enroll Ratio / (1- Subpopulation enroll reduction)] x [# Treatment groups]

Net Cost of Trial = Trial Cost + Screening Cost

Cost Savings = Net Cost of Trial (current) − Net Cost of Trial (biom. selected subpopulation)

4.5.1. Cost model assumptions

Based on two interviews with a clinical development executives with extensive pharmaceutical industry experience in the AD field, we have made the following cost assumptions:
- Apoe4 genotyping cost: $100 per patient during screening;
- CSF lumbar puncture costs: $1,000 per patient during screening (fully loaded cost, includes procedure and analysis of ABeta 1-42 & Tau biomarkers).
- Brain volumetric MRI scan: $2,000 per patient during screening (fully loaded cost, includes external over-read at lab).
- Physician visit & cognitive assessment costs: $1,500 per patient at screening (fully loaded cost, includes institutional overhead @40% and takes into account site start-up fees and IRB fees).
- 4x ratio of number of patients screened to number of patients enrolled.
- 18-month clinical trial cost: $60,000 per patient (fully loaded cost, includes external costs for treatment administration, cognitive assessment and biomarker measurements, but also internal costs such as project management).
- 6 treatment arms (2 required Phase 3 pivotal trials with 3 groups each: placebo, dose 1 and dose 2).
- Patients enrolled: as defined for each subpopulation selected with biomarkers or combinations of biomarkers (see table in section 4.4.4).
- Subpopulation reduction: as defined for each subpopulation selected with biomarkers or combinations of biomarkers (see last table in section 4.4.4). This assumption effectively increases the number of patients that need to be screened.

4.5.2. Cost impact calculations

The table on the next page highlights the corresponding calculations. It shows that the fully loaded cost of current disease-modifying AD trials is $133M. This number is in line with the recently published cost of the Flurizan 18 month trial.

Depending on the assumptions you take, you can save anywhere between $25M and $92M off that amount.

Interestingly, the biggest saving occurs when you drastically selects out 75% of patients you would normally enroll (see universal biomarker ratio Tau/ABeta 1-42).

Also, under several of the biomarker selected subpopulation groups, you end up screening less patients than the number of patients screened in current disease-modifying drug trials: screened number per treatment group is 8,088 patients for current trials according to our assumptions, while under biomarker selected subpopulation groups the number of screened patients varies between 4,704 (-42%) and 11,383 (+41%).

Interestingly, we observe that even though we impose more stringent selection criteria (i.e screening criteria) on the biomarker selected subpopulation trial, we can end up screening fewer patients than in the current disease-modifying trial.

	Current disease-modifying trials	Leverage factor scenario: SD & TE levers			Leverage factor scenario: Only SD lever				Universal biomarker	
		ABeta 1-42 < 161	Ratio Tau/ABeta > 0.84	Combi 1: ABeta + Ratio	Apoe4 Non carriers	Hippocampal Volume between > 1376 & < 1789	Combi 2: Apoe4 + Ratio	Combi 3: Apoe4 + Hippocampal Volume	Ratio Tau/ABeta > 1.08 (µT unchanged)	Ratio Tau/ABeta > 1.08 (µT − µP unchanged)
Trial cost										
Trial cost per patient	$ 60,000	$ 60,000	$ 60,000	$ 60,000	$ 60,000	$ 60,000	$ 60,000	$ 60,000	$ 60,000	$ 60,000
# enrolled patients	337	150	98	94	260	151	166	142	68	117
# treatment groups	6	6	6	6	6	6	6	6	6	6
Total # enrolled patients	2,022	900	588	564	1,560	906	996	852	408	702
TOTAL	$ 121,320,000	$ 54,000,000	$ 35,280,000	$ 33,840,000	$ 93,600,000	$ 54,360,000	$ 59,760,000	$ 51,120,000	$ 24,480,000	$ 42,120,000
Screening cost										
- Apoe4 genotyping					$ 100		$ 100	$ 100		
- CSF tap & biochemistry		$ 1,000	$ 1,000	$ 1,000			$ 1,000		$ 1,000	$ 1,000
- MRI scan						$ 2,000		$ 2,000		
- Visit & cognitive assessment	$ 1,500	$ 1,500	$ 1,500	$ 1,500	$ 1,500	$ 1,500	$ 1,500	$ 1,500	$ 1,500	$ 1,500
Screening cost per patient	$ 1,500	$ 2,500	$ 2,500	$ 2,500	$ 1,600	$ 3,500	$ 2,600	$ 3,600	$ 2,500	$ 2,500
- # enrolled patients	337	150	98	94	260	151	166	142	68	117
- Screen to enroll ratio	4	4	4	4	4	4	4	4	4	4
- Subpopulation enroll reduction	0%	25%	50%	54%	31%	50%	65%	65%	75%	75%
# screened patients	1,348	800	784	817	1,507	1,208	1,897	1,623	1,088	1,872
# treatment groups	6	6	6	6	6	6	6	6	6	6
Total # screened patients	8,088	4,800	4,704	4,904	9,043	7,248	11,383	9,737	6,528	11,232
TOTAL	$ 12,132,000	$ 12,000,000	$ 11,760,000	$ 12,260,870	$ 14,469,565	$ 25,368,000	$ 29,595,429	$ 35,053,714	$ 16,320,000	$ 28,080,000
NET COST OF TRIAL	$ 133,452,000	$ 66,000,000	$ 47,040,000	$ 46,100,870	$ 108,069,565	$ 79,728,000	$ 89,355,429	$ 86,173,714	$ 40,800,000	$ 70,200,000
SAVINGS	N/A	$ (67,452,000)	$ (86,412,000)	$ (87,351,130)	$ (25,382,435)	$ (53,724,000)	$ (44,096,571)	$ (47,278,286)	$ (92,652,000)	$ (63,252,000)

5. Conclusions and Discussion

In order to conclude on our findings in this thesis we go through a stepwise process:

- Conclude on the quality of the biomarkers to reduce enrollment and save costs (section 5.1);

- Explain additional factors that influence the overall quality and practical use of the selected biomarkers (section 5.2);

- Summarize the overall characteristics of the biomarkers we used in this thesis (section 5.3);

- Put the findings in the context of recently published data (section 5.4);

- Conclude overall on the overall quality of the biomarkers to select a subpopulation of patients and summarize other general take-aways from the thesis (section 5.5);

- Provide an outlook and discussion from the thesis (section 5.6).

5.1. Summary of sample size and cost impact findings

In this thesis we have taken the trial design used by current sponsors developing disease-modifying drugs for the treatment of Alzheimer's Disease, and we show how good a biomarker needs to be to select a sub-population to drastically reduce patient enrollment in pivotal clinical trials.

First, we showed what was theoretically needed in order to reduce the required enrollment in the trial by a factor 5x. We calculated in the table below a set of values for the standard deviation and treatment effect that lead to a 5-fold reduction in enrollment (from 337 patients to 67 patients):

Standard Deviation	Treatment effect	Leverage factor scenario
4.0 ADAS-cog points	11.9%	Only SD lever
5.0 ADAS-cog points	15.0%	SD & TE levers
6.8 ADAS-cog points	20.0%	SD & TE levers
8.5 ADAS-cog points	25.0%	SD & TE levers
9.0 ADAS-cog points	26.7%	Only TE lever

Second, we analyzed the biomarker data collected in the ADNI patient database and looked at how a biomarker selected population can impact required enrollment and trial costs in an applied setting.

Our main finding is that if you select a subpopulation among mild AD patients you can strongly reduce the number of enrolled patients and the cost of the trial as shown below:

Leverage factor scenario	Biomarker subpopulation selection criteria	Sub-population reduction	# Enrolled/ # Screened (per group)	Reduction in # enrolled	Total trial cost	Cost savings
Current design	-	0%	337pts / 8,088 pts	-	$133.5M	-
SD & TE levers	ABeta 1-42 < 161	25%	150 pts / 4,800 pts	2.2 x	$66.0M	$67.5M
	Ratio Tau/ABeta > 0.84	50%	98 pts / 4,704 pts	3.4 x	$47.0M	$86.4M
	Combi 1: ABeta + Ratio	54%	94 pts / 4,904 pts	3.6 x	$46.1M	$87.4M
Only SD lever	Apoe4 Non carriers	31%	260 pts / 9,043 pts	1.3 x	$108.1M	$25.4M
	Hippocampal Volume between > 1376 & < 1789	50%	151 pts / 7,248 pts	2.2 x	$79.7M	$53.7M
	Combi 2: Apoe4 + Ratio	65%	166 pts / 11,383 pts	2.0 x	$89.4M	$44.1M
	Combi 3: Apoe4 + Hippocampal Volume	65%	142 pts / 9,737 pts	2.4 x	$86.2M	$47.3M
Universal biomarker	Ratio Tau/ABeta > 1.08 (SD & TE levers)	75%	68 pts / 6,528 pts	5.0 x	$40.8M	$92.7M
	Ratio Tau/ABeta > 1.08 (only SD lever)	75%	117 pts / 11,232 pts	2.9 x	$70.2M	$63.3M

To our knowledge, this is the first time the cost impact of using a biomarker to select a subpopulation in a trial has been calculated.

The above table also illustrates a number of interesting findings:
- The number of enrolled patients can be reduced by a factor of 1.3 – 5.0 x.
- In some scenarios, despite having a biomarker selected subpopulation, you end up screening fewer patients than for current disease-modifying trials. It means that the reduction in enrolled patients over compensates the need for an increased number of screened patients.
- The cost savings vary from $25M to $93M based on a cost for current trials of $133M. This is a very significant cost saving and could in the best scenario it allows the sponsor to test three drugs in pivotal trials instead of only one.
- The biomarker selected subpopulations vary based on the leverage factor scenario ("SD & TE levers" or "only SD lever") that you assume for your treatment. Because of that, the sponsor needs to find out before the Phase 3 trial "what type of drug you are developing". We believe this framework gives tools to the sponsors to make such assessments early. Since we do not know anything initially about the sample size equation variables μ_T and σ_T for our drug, we need to use the Phase 1 and 2 trials to get informed about our drug's effect on those variables. Therefore, if possible, the sponsor can use AD patients in Phase 1 trials

you should do so. And it is crucial to do extensive testing in Phase 2 trials that will get a hint of the drug's effect (and the variables μ_T and σ_T).

5.2. Additional considerations pertaining to biomarker selected subpopulations

In the previous section we showed the impact of a biomarker selected sub population on both patient enrollment and trial cost.

Here we also discuss additional considerations necessary to evaluate the quality of biomarker or combinations of biomarkers introduced above. Such considerations include:
 a. the risk of having a restricted label reducing the market opportunity;
 b. the additional enrollment difficulty due to the to an increase in the number of screened patients;
 c. the use of longitudinal biomarkers can make the screening process very lengthy and complicated.

a. Restricted label issue
One of the potential limitations of using biomarkers to select a subpopulation to enroll in the trial is that the approval of the drug will be made only with a restricted label. Depending on the biomarker criteria used to select a subpopulation, the market opportunity could be reduced by up to 75%. From preliminary discussion with a regulatory specialist, we do not expect a restricted label for selection with brain volume or CSF protein concentration criteria. That would be good news for the sponsors but needs to be confirmed by the FDA for these biomarkers. Only the Apoe4 genotyping criteria would lead to a label restriction given the binary nature of genotyping results.

Note also that one of the current inclusion criteria for the pivotal trial of Eli Lilly's drug LY450139 is "*A magnetic resonance imaging (MRI) or computerized tomography (CT)scan in the last 2 years with no findings inconsistent with a diagnosis of Alzheimer's disease*". We do not expect that to limit the label and we think that you can easily make the case that using the average hippocampal volume biomarker in that context would not result in a restricted label.

Should the label be restricted, the sponsor could partially offset the downside by setting a higher treatment price with the argument that the therapy is targeted.

b. Enrollment difficulty could lead to delay in enrollment
In 2007, there were around 120 AD drugs in human clinical trials making the enrollment of AD patients difficult and time-consuming. The quest for "naïve" patients is very competitive and some sponsors even conducted their trials in foreign countries where there are no ethical issues with not giving the patients the current symptomatic treatments.[18]

Therefore it is important to realize that an increased enrollment difficulty can result in a longer enrollment period. Such a delay in the trial is very costly.

The delay in enrollment can partially be offset by starting the screening of the patients "at risk", i.e. before the Phase 2 data becomes available and the sponsors decides to go ahead with the Phase 3 trial.

c. Practical feasibility of longitudinal screening:
In this analysis, the subpopulation could be selected based on baseline data (and not longitudinal data) thereby avoiding more costly and complicated screening of the subpopulation in the trial. This is a great advantage over longitudinal biomarkers to select a subpopulation.

Interestingly we showed that the change in average hippocampal volume was not predictive of change in ADAS-cog at M12. However this result could have been flawed because we selected the MRI measurement concurrent with the ADAS-cog measurement (i.e. both at BL and M12 and running the regression with that data), instead of measuring them in a sequential fashion (i.e. the imaging at BL and M12 and then ADAS-cog at M12 and M24/M30).
Hence, it is cumbersome to use changes in biomarkers over time because they would require longitudinal screening before enrollment. One way around this issue is the company initiates screening of a patient population at least 12 months before the start of the study. This has the downside of putting money at risk before a peek at the Phase 2 results and it needs to be carefully evaluated from a practical feasibility standpoint. At a time of an increasing number disease-modifying therapies and a better understanding of the molecular mechanisms of the disease, this may however be a risk worth taking.

5.3. Review of the selected biomarkers characteristics

As shown in the table below, on the whole the suggested biomarkers or combination of biomarkers to select a subpopulation to enroll in the trial compares favorably to the characteristics defined in section 2.2:

Criteria of strong biomarker	Assessment	Comments
High predictive power.	Good	47% R squared is good and one of the four selected biomarkers even had a p value below 0.05
No longitudinal data needed at the time of enrollment	Excellent	No longitudinal data needed. In fact, we did select out the only longitudinal biomarkers (average hippocampal volume change at 12 months) as it was poorly predictive.
Not too costly	Good	Several of the biomarker tests, especially the MRI scan an the CSF lumbar puncture, are relatively expensive. However, these screening costs are dwarfed in the total picture of trial savings. If you use a combination of biomarkers you could envision a stepwise use of these tests. For instance you would start with the cheap Apoe4 genetic test followed later by the other tests if patient is "still in the game".

Does not result in a restricted label	Depends	Apoe4 genotyping might be the exception here and would lead to a decreased market size due to the restricted label. For the other biomarkers, if is unclear on how the FDA would decide on this matter.			
Not leading to enrollment difficulty	Good	Most probably they will not lead to enrollment difficulty as we will probably not use the inferior biomarkers needing a very large pool of screened patients. We do not think that a lumbar puncture (CSF tap) are too problematic as a screening procedure, given the benefit for the patient of knowing his CSF protein concentrations for further treatment should he not be enrolled in the given trial he was initially screened for.			

Based on the analysis in sections 2.2 and 3.3.3 we conducted below an analysis of the characteristics of the biomarkers to select a subpopulation of mild AD patients.

Biomarker	High predictive power	No longitud-inal data needed	Not too costly to perform	Does not result in restricted label	Not leading to enrollment difficulty
MMSE	NO	√	√	√	√
A beta 1-42	YES	√	OK compared to total cost	Need FDA confirm.	√
Tau	NO	√	OK compared to total cost	Need FDA confirm.	√
P-tau 181P	NO	√	OK compared to total cost	Need FDA confirm.	√
Ratio Tau/Ab1-42	YES	√	OK compared to total cost	Need FDA confirm.	√
Ratio P-Tau181P/ Ab1-42	NO	√	OK compared to total cost	Need FDA confirm.	√
Both apoe4 genes	YES	√	√	X	√
At least one Apoe4 gene	NO	√	√	X	√
MRI – avg. hippocampal volume	YES	√	OK compared to total cost	Need FDA confirm.	√
MRI – avg. hippocampal volume change	NO	X	OK compared to total cost	Need FDA confirm.	√

√: Not problematic characteristic; X: Problematic characteristic; ???: Uncertain characteristic.
In red we highlighted the poor characteristics of inferior biomarkers and with a grey background we show the superior biomarkers.

5.4. Findings in the context of recent data of AD disease-modifying drugs

In July 2008, the data from bapineuzumab's 18 months Phase 2 were presented at an investor conference call and at the ICAD Conference in Chicago. These are the first data set released with an analysis for different subpopulations. Specifically they made a modified intent to treat analysis of the subpopulations of Apoe4 carriers and Apoe4 non-carriers. Although it was not specified, we assume that they mean by Apoe4 carriers, that both chromosomes carry the Apoe4 allele (corresponding the "Both Apoe4 genes" biomarker in this thesis). We have however not verified it with the company.

The data presented showed statistical significance in the Apoe4 non-carrier group but not in the Apoe4 carrier group. This would hint that bapineuzumab is a treatment falling under the leverage factor scenario "only SD lever". Based on our findings described in section 4.3.2., Wyeth/Elan would have needed 1.8 x more patients to show statistical significance in the Apoe4 carrier group compared to the non-carrier group (according to the required enrollments for these subgroups: 1.8 x =459 patients in Apoe4 carrier group / 260 patients in Apoe4 non-carrier group).

In addition, in the bapineuzumab trial there were 5 groups (4 doses plus placebo) and 229 patients (i.e approximately 40-50 patients per group). As a result, even though the data was not statistically significant, it was not powered to show statistical significance (the average number of patient per group was shown to be powered on overage with 337 patients, see section 3.1.2.).

Finally with the real worry of having a restricted label for bapineuzumab when using the Apoe4 biomarker to select a subpopulation, the sponsor could instead pick their subpopulation with the average hippocampal volume biomarker: that will probably not lead to a restricted label.

As a caution, note that the patient population in the bapineuzumab trial had MMSE scores in the 16-26 range corresponding to mild to moderate patients, while the MMSE scores of the mild AD patients in our ADNI sample range from 20-26 points (average 22.9 points). The framework would need to be tested on mild to moderate patients in order to verify this conclusion.

5.5. Final conclusions

The following table incorporates the findings under section 5.1 and the additional considerations in sections 5.2 and 5.3:

Leverage factor scenario	Biomarker subpopulation selection criteria	# Screened / Risk of delay	Cost savings	Limited label	Longitu-dinal biomarker	Overall quality of biomarker/ combination
Current design	-	8,088 pts / -	-	-	-	-
SD & TE levers	ABeta 1-42 < 161	4,800 pts / Benefit	$67.5M	Probably not	NO	√√
	Ratio Tau/ABeta > 0.84	4,704 pts / Benefit	$86.4M	Probably not	NO	√√√
	Combi 1: ABeta + Ratio	4,904 pts / Benefit	$87.4M	Probably not	NO	√√√√
Only SD lever	Apoe4 Non carriers	9,043 pts / Risk	$25.4M	YES	NO	√√
	Hippocampal Volume between > 1376 & < 1789	7,248pts / Benefit	$53.7M	Probably not	NO	√√√√
	Combi 2: Apoe4 + Ratio	11,383 pts / Risk	$44.1M	YES	NO	√√
	Combi 3: Apoe4 + Hippocampal Volume	9,737 pts / Risk	$47.3M	YES	NO	√√
Universal biomarker	Ratio Tau/ABeta > 1.08 (SD & TE levers)	6,528 pts / Benefit	$92.7M	Probably not	NO	√√√√
	Ratio Tau/ABeta > 1.08 (Only SD lever)	11,232 pts / Risk	$63.3M	Probably not	NO	√√√

From the above graph we conclude that:
- under the "SD & TE levers" scenario, the biomarker subpopulation selection criteria "Ratio Tau/ABeta > 1.08" is the most promising when coming down to cost savings.
- under the same scenario if you expect difficulties with enrollment we recommend the combination 1 biomarker criteria: "ABeta 1-42 < 161 and Ratio Tau/ABeta > 0.84".
- under the "only SD lever" scenario, the biomarker subpopulation selection criteria "Hippocampal Volume between > 1376 & < 1789" is the most promising both in terms of cost savings and due to the very small risk of having a limited label.

In addition, we summarize here our general take-aways from this thesis:

- The ABeta biomarker, the ratio Tau/ABeta biomarker and the average hippocampal volume biomarker have the most superior characteristics to select a patient subpopulation to enroll in 18 month disease-modifying AD trials.

- The ratio Tau/ABeta is the best biomarker to use across the board to select subpopulation independently of the general assumptions taken.

- Pharmaceutical companies should collect biomarker data relevant to their treatment already in Phase 1 and Phase 2. That will help them understand under which leverage factor scenario they work and be able to successfully use biomarkers to select a subpopulation that will enroll in Phase 3.

- From an analysis of the recent bapineuzumab data from Wyeth/Elan, it is likely that the drug (and possibly other amyloid beta targeting drugs) is falls under the "only SD lever" scenario. In addition, Wyeth/Elan's use of the Apoe4 biomarkers to select a subpopulation may lead to a restricted label. This thesis suggests that Wyeth/Elan's worry could potentially be solved by using the average hippocampal volume biomarker instead of the Apoe4 biomarker. Since Phase 3 studies are already under way with subpopulations based on the Apoe4 biomarker, Wyeth/Elan now need to wait and see if they get approval for a restricted market. If that is the case, they could potentially in the future expand their market by running a Phase IV study based on patients selected with the average hippocampal biomarker.

- These conclusions were based on the analysis of a small sample of only 32 mild AD patients. As a result it would be important to get a bigger sample and verify that the recommendations and conclusions still hold true.

5.6. Discussion and outlook

Risk of reducing enrollment?
In this thesis, we outlined the benefits of reducing the sample size in terms of cost savings and potentially also on enrollment time. However, it is important to mention the risk to industry sponsors of reducing enrollment: the risk is that the sponsor rejects a drug with modest effect (eg. Aricept donepezil had a treatment effect of 10.4% at 6 months) that can have several billion dollars in annual sales.

The flip side of this argument is that the industry sponsor can still decide to enroll a large number of patients in the trial and, in such a situation, the proposed biomarker selected subpopulation would only contribute to a greater chance of showing statistical significance in the trial or to the sponsor deciding to target a higher treatment effect.

Picking other primary endpoints such as time-to-progression:
Further studies should be conducted to apply this framework of biomarker selected subpopulations to different primary endpoints. High on that list is the time-to-progression endpoint, which has been already accepted by the FDA as a primary endpoint for MCI trials. It is even more the case in the current environment where the sponsors and the scientific community believe that earlier intervention would be beneficial for the patient and would give the sponsor a higher likelihood of showing effect with their drugs.

Limitations of current data in the ADNI database:
The ADNI database is a very extensive database and a great tool to analyze biomarkers. However I would mention the following limitations of the database :
- For the biomarker data we wish to analyze we only have a limited number of patients who were actually measured at 18 months. So we picked the ADAS-cog change at 12 months instead of at 18 months and we adjusted it accordingly (see section).
- We were missing longitudinal biomarker data for the CSF biomarkers Tau, P-Tau and ABeta 1-42. Ideally you would like to have it at 6 and 12 months and have the opportunity to analyse their predictive power in the regression;
- the PET imaging data would also be nice to have at those same intervals., By analyzing the available PET data we did not find enough overlap with the other biomarker data so it would have resulted in a too small sample to analyze;
- the database only has mild AD patients. So it would be interesting to also have moderate AD patients and severe AD patients to further benchmark our findings.
- with the 10 biomarkers selected in our regression, we only get a patient population of 32 mild AD patients from the initial 200 mild AD patients in the database. As a result we get a good R squared value (47%) but a relatively poor adjusted R squared (22%) due to our small sample. So if possible we would rather have 100-300 patients having the data for all 10 biomarkers.

Appendices

Appendix 1 – Correlation coefficients of biomarkers

All (n=156)

	MMSE	Tau	Ab1-42	P-Tau 181P	Tau/Ab1-42	P-Tau181P/Ab1-42	Both APOE4 genes	At least one APOE4 genes	AVG HIPPO VOL	AVG HIPPO VOL change	ADAS 11 change
MMSE	1.00										
Tau	(0.37)	1.00									
Ab1-42	0.38	(0.45)	1.00								
P-Tau 181P	(0.32)	0.75	(0.53)	1.00							
Tau/Ab1-42	(0.37)	0.91	(0.67)	0.73	1.00						
P-Tau181P/Ab1-42	(0.35)	0.75	(0.71)	0.93	0.85	1.00					
Both APOE4 genes	(0.28)	0.22	(0.36)	0.18	0.32	0.29	1.00				
At least one APOE4 genes	(0.31)	0.36	(0.56)	0.37	0.45	0.45	0.39	1.00			
AVG HIPPO VOL	0.51	(0.35)	0.35	(0.28)	(0.37)	(0.31)	(0.20)	(0.29)	1.00		
AVG HIPPO VOL change	0.11	(0.03)	0.17	(0.14)	(0.07)	(0.15)	0.02	0.03	(0.11)	1.00	
ADAS 11 change	(0.40)	0.22	(0.21)	0.14	0.24	0.19	0.23	0.15	(0.31)	0.04	1.00

AD (n=32)

	MMSE	Tau	Ab1-42	P-Tau 181P	Tau/Ab1-42	P-Tau181P/Ab1-42	Both APOE4 genes	At least one APOE4 genes	AVG HIPPO VOL	AVG HIPPO VOL change	ADAS 11 change
MMSE	1.00										
Tau	(0.08)	1.00									
Ab1-42	0.06	(0.10)	1.00								
P-Tau 181P	0.02	0.71	(0.29)	1.00							
Tau/Ab1-42	(0.08)	0.89	(0.49)	0.72	1.00						
P-Tau181P/Ab1-42	(0.04)	0.64	(0.54)	0.94	0.80	1.00					
Both APOE4 genes	0.17	0.21	(0.39)	0.19	0.36	0.31	1.00				
At least one APOE4 genes	0.21	0.20	(0.34)	0.28	0.27	0.30	0.36	1.00			
AVG HIPPO VOL	(0.05)	(0.25)	0.00	(0.08)	(0.21)	(0.11)	(0.10)	(0.04)	1.00		
AVG HIPPO VOL change	0.14	(0.04)	0.13	(0.18)	0.01	(0.12)	0.26	(0.02)	(0.24)	1.00	
ADAS 11 change	(0.01)	0.13	(0.43)	(0.02)	0.28	0.14	0.29	(0.05)	(0.31)	0.10	1.00

NL (N=63)

	MMSE	Tau	Ab1-42	P-Tau 181P	Tau/Ab1-42	P-Tau181P/Ab1-42	Both APOE4 genes	At least one APOE4 genes	AVG HIPPO VOL	AVG HIPPO VOL change	ADAS 11 change
MMSE	1.00										
Tau	0.03	1.00									
Ab1-42	(0.17)	(0.37)	1.00								
P-Tau 181P	0.04	0.74	(0.36)	1.00							
Tau/Ab1-42	0.11	0.84	(0.71)	0.68	1.00						
P-Tau181P/Ab1-42	0.09	0.69	(0.68)	0.86	0.88	1.00					
Both APOE4 genes	0.12	0.00	(0.19)	(0.10)	0.08	(0.02)	1.00				
At least one APOE4 genes	0.04	0.11	(0.43)	0.03	0.29	0.22	0.24	1.00			
AVG HIPPO VOL	0.06	0.04	0.05	(0.05)	0.07	(0.00)	0.01	0.06	1.00		
AVG HIPPO VOL change	(0.02)	0.07	0.17	0.01	(0.06)	(0.08)	0.12	0.07	(0.31)	1.00	
ADAS 11 change	0.04	0.01	0.17	(0.14)	(0.10)	(0.18)	(0.10)	0.01	0.06	0.18	1.00

MCI (n=61)

	MMSE	Tau	Ab1-42	P-Tau 181P	Tau/Ab1-42	P-Tau181P/Ab1-42	Both APOE4 genes	At least one APOE4 genes	AVG HIPPO VOL	AVG HIPPO VOL change	ADAS 11 change
MMSE	1.00										
Tau	(0.18)	1.00									
Ab1-42	0.16	(0.44)	1.00								
P-Tau 181P	0.03	0.70	(0.54)	1.00							
Tau/Ab1-42	(0.19)	0.92	(0.63)	0.65	1.00						
P-Tau181P/Ab1-42	(0.04)	0.73	(0.71)	0.93	0.81	1.00					
Both APOE4 genes	(0.08)	0.07	(0.31)	0.00	0.21	0.14	1.00				
At least one APOE4 genes	0.09	0.30	(0.53)	0.32	0.39	0.40	0.32	1.00			
AVG HIPPO VOL	0.24	(0.25)	0.23	(0.04)	(0.29)	(0.11)	(0.01)	(0.15)	1.00		
AVG HIPPO VOL change	(0.00)	(0.01)	0.14	(0.19)	(0.04)	(0.19)	(0.09)	0.14	(0.07)	1.00	
ADAS 11 change	(0.29)	0.15	(0.12)	0.12	0.16	0.13	(0.04)	0.02	(0.23)	0.07	1.00

Appendix 2 – Biomarkers used in Alzheimer's

Diagnostic criteria	Type of biomarker	Disease modifying component	Identify/predict	Sensitivity (have disease & tested positive)	Specificity (do not have disease and tested negative)	Comments
DSM-IV-TR / NINCDS-ADRDA	Clinical	N	AD	65-96%	23-88%	Against neuropathological gold standard
a-beta 1-42 in CSF	Lab	Y	AD	86%	90%	
t-tau in CSF	Lab	Y	AD	81%	90%	
p-tau in CSF	Lab	Y	AD	80%	92%	
Combi a-beta 1-42 and t-tau in CSF	Lab	Y	AD	85-94%	83-100%	
PET glucose	Imaging	N	AD	86%	86%	Pooled, High variability
SPECT	Imaging	N	AD	77-80%	65-93%	
SPECT	Imaging	N	AD	65-71%	79%	Meta-analysis
Satoris 18x blood biomarkers[1]	Lab	N	AD	90%	88%	
Age	Clinical	N	Prodromal AD	74%	N/A	
Age + Medial Temporal Lobe measures	Clinical + Imaging	N	Prodromal AD	81%	N/A	
Memory score	Clinical	N	Prodromal AD	88%	N/A	
Memory score + Medial Temporal Lobe measures	Clnical + imaging	N	Prodromal AD	96%	N/A	
Combi a-beta 1-42, p-tau and t-tau in CSF	Lab	Y	Prodromal AD	>90%	>85%	
PET glucose	Imaging	N	Prodromal AD	75-84%		
Delayed recall scores + PET glucose	Clinical + Imaging	N	Prodromal AD	>90%	>90%	

Diagnostic criteria	Type of biomarker	Disease modifying component	Identify/predict	Sensitivity (have disease & tested positive)	Specificity (do not have disease and tested negative)	Comments
PET glucose	Imaging	N	AD from Dementia with Lewy Bodies	86-92%	80-81%	
PET glucose	Imaging	N	AD from Fronto-Temporal Dementia	78%	71%	
PET glucose	Imaging	N	AD from Vascular Dementia	75-88%	18-53%	
Satoris 18x blood biomarkers[1]	Lab	N	MCI converting to AD and not other dementias	91%	100%	Over time period of 2-6 years
Satoris 18x blood biomarkers[1]	Lab	N	MCI remaining MCI	Need another test	Need another test	Over time period of 2-6 years

[1] Ray et al., 11/2007. Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling protein.

Appendix 3 – Phase 3 clinical trial protocols for AD disease-modifying drugs

Source Clinicaltrials.gov
Search date 5/11/2008
Serach criteria Condition: "Alzheimer" AND Study type: "Interventional" AND "Phase III" AND picked only disease-modifying treatment candidates

Drug	CT.gov ID	Study name	Sponsor	Status	Published results	Study start	Study end	Phase	Normal/MCI/AD DisMod/AD Sympt	n	Disease-mod. (Y/N)	Enrollment period	Trial period	Total duration	Primary Endpoint	Groups	n / Group
MPC-7869	NCT00322036	Global efficacy study of MPC-7869 to treat patients with AD	Myriad	Active, not recruiting	No	May-06	Dec-08	III	Mild	800	Y	N/A	N/A	135	Cognition and activities of daily living at 18 months	1 dose plus placebo	400
MPC-7869	NCT00105547	Efficacy study of MPC-7869 to treat patients with Alzheimer's	Myriad	Active, not recruiting	No	Feb-05	May-08	III	Mild	1600	Y	91	78	169	Cognition and activities of daily living at 18 months	1 dose plus placebo	800
MPC-7869	NCT00380276	Open label treatment with MPC-7869 for patients with Alzheimer's who participated in an MPC-7869 protocol	Myriad	Recruiting	No	Sep-06	Dec-08	III (open label)		1000	Y	N/A	N/A	117	Safety	1 dose	
3APS	NCT00088673	Evaluation of 3APS in patients with mild to moderate AD	Bellus health Inc. (Neurochem)	Active, not recruiting	No	Jun-04		III	mild to moderate	950	Y	N/A	78	N/A	At 18M 1. ADAS-cog and CDR-SB decline 2. brain volume change MRI	1 dose plus placebo	475
3APS	NCT00217763	European study of 3APS in mild to moderate AD patients	Bellus health Inc. (Neurochem)	Active, not recruiting	No				mild to moderate	930	Y	39	78	117	at 18 M	2doses plus placebo	310
3APS	NCT00314912	Open label extension of the phase III study with tramiprosate (3APS) in patients with mild to moderate AD	Bellus health Inc. (Neurochem)	Active, not recruiting	No	Sep-05	Dec-07	III		650	Y	N/A	52	N/A	long-term safety	1 dose	
LY450139	NCT00594568	Effect of LY450139 on the long term progression of AD	Eli Lilly	Active, not recruiting / Recruiting	No / No	May-06 / Mar-08	Mar-12	III / III	mild to moderate	1500	Y	105	104	209	cognitive and functional decline	2 doses plus placebo	500
bapineuzumab	NCT00667810	study evaluating the efficacy and safety of bapineuzumab in AD patients Who Are Apolipoprotein E ε4 Non-Carriers	Wyeth	Not yet recruiting	No	May-08	Apr-11	III		1250	Y	74	78	152	Alzheimer's Disease Assessment Scale-Cognitive Subscale; Disability Assessment for Dementia	3 doses plus placebo	313
bapineuzumab	NCT00676143	Study Evaluating the Safety and Efficacy of Bapineuzumab in AD Patients Who Are Apolipoprotein E ε4 Carriers	Wyeth	Not yet recruiting	No	May-08	Apr-11	III		800	Y	74	78	152	Alzheimer's Disease Assessment Scale-Cognitive Subscale; Disability Assessment for Dementia	3 doses plus placebo	200
bapineuzumab	NCT00575055	Bapineuzumab in patients with mild to moderate AD Who Are Apolipoprotein E4 Carriers	Elan	Recruiting	No	Dec-07	Dec-10	III	mild to moderate	800	Y	79	78	157	Cognitive and Functional	1 dose plus placebo	400
bapineuzumab	NCT00574132	Bapineuzumab in patients with mild to moderate AD Who Are Apolipoprotein E4 Non-Carriers	Elan	Recruiting	No	Dec-07	Dec-10	III	mild to moderate	1250	Y	79	78	157	Cognitive and Functional	3 doses plus placebo	313
									Avg per drug (ex safety)	2720	Avg.	77	78	152		Average n per Group	337

Drug	CT.gov ID	Age min	Age max	Probable AD	MMSE	Living in the community	Communication OK	Min. Education	ApoE pos	Positive MRI (not inconsistent with AD)	Informed consent	Caregiver available	Only women sterile or post menopause	Participated in previous study	Other AD meds/stable	Psychiatric disorders	Significant illness or unhealed surgery	Evidence of non-AD dementia	Stroke/seizure	Smoking	Alcohol abuse	Prior treatment with same drug or class of drugs	Prior or current treatment with other class of investig. drugs	Pacemakers/metal objects in body
		Inclusion criteria		Assess Scale	Assess Scale	History	History	History	Biomarker	Biomarker	Practical	Practical	Ethical	Drugs	Drugs	**Exclusion criteria**	Co-morbid	Co-morbid	Co-morbid	Co-morbid	Co-morbid	Drugs	Drugs	Practical
MPC-7869	NCT00322036	55		X		X	X	X			X	X	X				X	X				X	X	
MPC-7869	NCT00105547	55		X							X	X	X	X			X						X	
MPC-7869	NCT00380276	55		X		X	X				X	X	X		X (optional)		X	X					X	
3APS	NCT00088673	50		X	MtoM	(X)	X				X	X	X		X (need)		X	X				X	X	
3APS	NCT00217763			X	MtoM	(X)	X				X	X	X		X (need)		X	X				X	X	
3APS	NCT00314912	50				(X)					X	X	X				X						X	
LY450139	NCT00594568	50	55	X	MtoM (16-26)					X		X	X	X		GDS <=6	X		Hachinski <=4	X			X (not active immunotherapy)	
bapineuzumab	NCT00667810	50	88	X	MtoM (16-26)					X		X	X		X (optional)	X		X						X
bapineuzumab	NCT00676143	50	88	X	MtoM (16-26)					X		X	X		X (optional)	X		X						X
bapineuzumab	NCT00575055	50	89	X	MtoM (16-26)					X		X	X		X (optional)	X	X	X	X	X			X (immunotherapy + some classes non-AD drugs)	X
bapineuzumab	NCT00574132	50	89	X	MtoM (16-26)					X		X	X		X (optional)	X	X	X	X	X			X (immunotherapy + some classes non-AD drugs)	X

Appendix 4 – Statistical model calculation (Aricept trial)

Title

Estimation of enrollment calculations in Aricept's NDA supporting Phase 3 trial

Key assumptions

Variables:	Aricept's Phase 3 trial
Patient population:	Mild to moderate AD
Outcome:	Change from baseline in ADAS-cog
Statistical model	Two sample inference test on means with unequal variances

Sample size calculations

$$n = (z_{1-\beta} + z_{1-\alpha})^2 * (\sigma_T^2 + \sigma_P^2) / (\mu_T - \mu_P)^2$$ 　　　　Sample size per group (Rosner 8.26)

α	0.05	significance level	
β	0.1	1-power	
σ_T	5.81	S.D. of ADAS-cog change for treatment group	Source: Donepezil for dementia due to Alzheimer's Disease Review
σ_P	5.86	S.D. of ADAS-cog change for placebo group	Source: Donepezil for dementia due to Alzheimer's Disease Review
μ_T	-0.9	ADAS-cog change of treatment group	Source: NDA 20-690/S-026. Ariecpt (donepzil hydrochloride tablets) and Donepezil for dementia due to Alzheimer's Disease.
μ_P	1.81	ADAS-cog change of placebo group	Source: NDA 20-690/S-026, and Donepezil for dementia due to Alzheimer's Disease Review.
Baseline score	26	ADAS-cog for both groups at baseline	Source: NDA 20-690/S-026, Page 2. Inclusion criteria: MMMSE scores between 10-26.
Decline T in %	-3.5%		
Decline P in %	7.0%		
$\mu_T - \mu_P$	-2.71	ADAS-cog points	
Treatment effect in %	10.4%		
$z_{1-\alpha}$	1.96		
$z_{1-\beta}$	1.28		
n	**97**		
Adjustment for losses to follow-up:	10%		
n	**108**		

Note: Interestingly, the Aricept Study 302 had 130 patients in each arm which explains the positive outcome of the study.

Appendix 5 – Statistical model calculation (current disease-modifying trials)

Title

Estimation of enrollment calculations in current NDA-supporting <u>disease-modifying</u> Phase 3 trial

Key assumptions

Variables:	**18-month disease-modifying treatment**
Patient population:	Mild to moderate AD
Outcome:	Change from baseline in ADAS-cog
Statistical model	Two sample inference test on means with unequal variances

Sample size calculations

$$n = (z_{1-\beta} + z_{1-\alpha})^2 * (\sigma_T^2 + \sigma_P^2) / (\mu_T - \mu_P)^2$$

Sample size per group (Rosner 8.26)

α	0.05	significance level	
β	0.1	1-power	
σ_T	9	S.D. of ADAS-cog change for treatment group	See explanations in Section 2.4
σ_P	9	S.D. of ADAS-cog change for placebo group	See explanations in Section 2.4
μ_T	4.94	ADAS-cog change of treatment group	See explanations in Section 2.4
μ_P	7.5	ADAS-cog change of placebo group	See explanations in Section 2.4
Baseline score	21.6	ADAS-cog for both groups at baseline	See explanations in Section 2.4
Decline T in %	22.9%		
Decline P in %	34.7%		
$\mu_T - \mu_P$	-2.56	ADAS-cog points	
Treatment effect in %	11.9%		
$z_{1-\alpha}$	1.96		
$z_{1-\beta}$	1.28		
n	**259**		
Adjustment for losses to follow-up:	23%		
n	**337**		

Appendix 6 – Sample size calculations for Apoe4 biomarker

$$n = (z_{1-\beta} + z_{1-\alpha})^2 * (\sigma_T^2 + \sigma_P^2) / (\mu_T - \mu_P)^2$$

	Trial design 18-month dis.-mod. treatment	Full 32 AD patient sample	Sub population (both Apoe4 genes)	Trial design Adjusted Both Apoe4 genes	Sub population (not both Apoe4 genes)	Trial design Adjusted Not both Apoe4 genes
Leverage factor scenario: SD & TE levers (μ_T remains unchanged)						
α	0.05			0.05		0.05
β	0.1			0.1		0.1
σ_T	9			10.5		7.9
σ_P	9	7.2	8.7	10.5	6.1	7.9
μ_T	4.94			4.94		4.94
μ_P	7.5	5.4	8.5	10.6	4	6.1
Baseline score	21.6			21.6		21.6
Decline T in %	22.9%			22.9%		22.9%
Decline P in %	34.7%			49.1%		28.2%
$\mu_T - \mu_P$	-2.56			-5.66		-1.16
Treatment effect in %	11.9%			26.2%		5.4%
$z_{1-\alpha}$	1.96			1.96		1.96
$z_{1-\beta}$	1.28			1.28		1.28
n	259			72		974
Adjustment for losses to follow-up	23%			23%		23%
n	**337**			**94**		**1,265**
Leverage factor scenario: only SD lever ($\mu_T - \mu P$ remains unchanged)						
α	0.05			0.05		0.05
β	0.1			0.1		0.1
σ_T	9			10.5		7.9
σ_P	9	7.2	8.7	10.5	6.1	7.9
μ_T	4.94			8.04		3.54
μ_P	7.5	5.4	8.5	10.6	4	6.1
Baseline score	21.6			21.6		21.6
Decline T in %	22.9%			37.2%		16.4%
Decline P in %	34.7%			49.1%		28.2%
$\mu_T - \mu_P$	-2.56			-2.56		-2.56
Treatment effect in %	11.9%			11.9%		11.9%
$z_{1-\alpha}$	1.96			1.96		1.96
$z_{1-\beta}$	1.28			1.28		1.28
n	259			353		200
Adjustment for losses to follow-up	23%			23%		23%
n	**337**			**459**		**260**

Appendix 7 - Sample size calculations for ABeta 1-42 biomarker

$$n = (z_{1-\beta} + z_{1-\alpha})^2 * (\sigma_T^2 + \sigma_P^2) / (\mu_T - \mu_P)^2$$

	18-month dis.-mod. treatment	Full 32 AD patient sample	Sub population (A-beta Quartile 1)	Trial design Adjusted Abeta Quartile 1	Sub population (A-beta Quartile 2)	Trial design Adjusted Abeta Quartile 2	Sub population (A-beta Quartile 3)	Trial design Adjusted Abeta Quartile 3	Sub population (A-beta Quartile 4)	Trial design Adjusted Abeta Quartile 4	Sub population (A-beta Half 1)	Trial design Adjusted Abeta Half 1	Sub population (A-beta Half 2)	Trial design Adjusted Abeta Half 2
Leverage factor scenario: SD & TE levers (μ_T remains unchanged)														
α	0.05			0.05		0.05		0.05		0.05		0.05		0.05
β	0.1			0.1		0.1		0.1		0.1		0.1		0.1
σ_T	9			11.7		5.7		9		6.1		9.1		8.3
σ_P	9	7.2	9.9	11.7	3.9	5.7	7.2	9	4.3	6.1	7.3	9.1	6.5	8.3
μ_T	4.94			4.94		4.94		4.94		4.94		4.94		4.94
μ_P	7.5	5.4	8.6	10.7	6.9	9	6.1	8.2	0.1	2.2	7.8	9.9	3.1	5.2
Baseline score	21.6			21.6		21.6		21.6		21.6		21.6		21.6
Decline T in %	22.9%			22.9%		22.9%		22.9%		22.9%		22.9%		22.9%
Decline P in %	34.7%			49.5%		41.7%		38.0%		10.2%		45.8%		24.1%
$\mu_T - \mu_P$	-2.56			-5.76		-4.06		-3.26		2.74		-4.96		-0.26
Treatment effect in %	11.9%			26.7%		18.8%		15.1%		-12.7%		23.0%		1.2%
$z_{1-\alpha}$	1.96			1.96		1.96		1.96		1.96		1.96		1.96
$z_{1-\beta}$	1.28			1.28		1.28		1.28		1.28		1.28		1.28
n	259			87		41		160		N/A		71		N/A
Adjustment for losses to follow-up:	23%			23%		23%		23%		23%		23%		23%
n	**337**			**113**		**54**		**208**		**N/A**		**92**		**N/A**
Leverage factor scenario: only SD lever (μ_T - µP remains unchanged)														
α	0.05			0.05		0.05		0.05		0.05		0.05		0.05
β	0.1			0.1		0.1		0.1		0.1		0.1		0.1
σ_T	9			11.7		5.7		9		6.1		9.1		8.3
σ_P	9	7.2	9.9	11.7	3.9	5.7	7.2	9	4.3	6.1	7.3	9.1	6.5	8.3
μ_T	4.94			7.44		7.14		5.64		-0.36		7.34		2.64
μ_P	7.5	5.4	8.6	10.7	6.9	9	6.1	8.2	0.1	2.2	7.8	9.9	3.1	5.2
Baseline score	21.6			21.6		21.6		21.6		21.6		21.6		21.6
Decline T in %	22.9%			34.4%		33.1%		26.1%		-1.7%		34.0%		12.2%
Decline P in %	34.7%			49.5%		41.7%		38.0%		10.2%		45.8%		24.1%
$\mu_T - \mu_P$	-2.56			-3.26		-1.86		-2.56		-2.56		-2.56		-2.56
Treatment effect in %	11.9%			15.1%		8.6%		11.9%		11.9%		11.9%		11.9%
$z_{1-\alpha}$	1.96			1.96		1.96		1.96		1.96		1.96		1.96
$z_{1-\beta}$	1.28			1.28		1.28		1.28		1.28		1.28		1.28
n	259			270		197		259		119		265		221
Adjustment for losses to follow-up:	23%			23%		23%		23%		23%		23%		23%
n	**337**			**351**		**256**		**337**		**155**		**345**		**287**

Appendix 8 - Sample size calculations for the Ratio Tau/ABeta 1-42 biomarker

$$n = (z_{1-\beta} + z_{1-\alpha})^2 \cdot (\sigma_T^2 + \sigma_P^2) / (\mu_T - \mu_P)^2$$

	Trial design 18-month dis.-mod. treatment	Full 32 AD patient sample	Sub population (Ratio Tau/A-beta Quartile 1)	Trial design Adjusted Ratio Tau/A-beta Quartile 1	Sub population (Ratio Tau/A-beta Quartile 2)	Trial design Adjusted Ratio Tau/A-beta Quartile 2	Sub population (Ratio Tau/A-beta Quartile 3)	Trial design Adjusted Ratio Tau/A-beta Quartile 3	Sub population (Ratio Tau/A-beta Quartile 4)	Trial design Adjusted Ratio Tau/A-beta Quartile 4	Sub population (Ratio Tau/A-beta Half 1)	Trial design Adjusted Ratio Tau/A-beta Half 1	Sub population (Ratio Tau/A-beta Half 2)	Trial design Adjusted Ratio Tau/A-beta Half 2
Leverage factor scenario: SD & TE levers (μ_T remains unchanged)														
α	0.05			0.05		0.05		0.05		0.05		0.05		0.05
β	0.1			0.1		0.1		0.1		0.1		0.1		0.1
σ_T	9			6.3		9		12.1		5.3		7.9		9.4
σ_P	9	7.2	4.5	6.3	7.2	9	10.3	12.1	3.5	5.3	6.1	7.9	7.6	9.4
μ_T	4.94			4.94		4.94		4.94		4.94		4.94		4.94
μ_P	7.5	5.4	1.1	3.2	5.1	7.2	9.3	11.4	6.2	8.3	3.1	5.2	7.8	9.9
Baseline score	21.6			21.6		21.6		21.6		21.6		21.6		21.6
Decline T in %	22.9%			22.9%		22.9%		22.9%		22.9%		22.9%		22.9%
Decline P in %	34.7%			14.8%		33.3%		52.8%		38.4%		24.1%		45.8%
$\mu_T - \mu_P$	-2.56			1.74		-2.26		-6.46		-3.36		-0.26		-4.96
Treatment effect in %	11.9%			-8.1%		10.5%		29.9%		15.6%		1.2%		23.0%
$z_{1-\alpha}$	1.96			1.96		1.96		1.96		1.96		1.96		1.96
$z_{1-\beta}$	1.28			1.28		1.28		1.28		1.28		1.28		1.28
n	259			N/A		333		74		52		N/A		75
Adjustment for losses to follow-up:	23%			23%		23%		23%		23%		23%		23%
n	337			N/A		432		96		68		N/A		98
Leverage factor scenario: only SD lever ($\mu_T - \mu P$ remains unchanged)														
α	0.05			0.05		0.05		0.05		0.05		0.05		0.05
β	0.1			0.1		0.1		0.1		0.1		0.1		0.1
σ_T	9			6.3		9		12.1		5.3		7.9		9.4
σ_P	9	7.2	4.5	6.3	7.2	9	10.3	12.1	3.5	5.3	6.1	7.9	7.6	9.4
μ_T	4.94			1.14		4.54		8.84		5.74		2.64		7.34
μ_P	7.5	5.4	1.1	3.2	5.1	7.2	9.3	11.4	6.2	8.3	3.1	5.2	7.8	9.9
Baseline score	21.6			21.6		21.6		21.6		21.6		21.6		21.6
Decline T in %	22.9%			5.3%		21.0%		40.9%		26.6%		12.2%		34.0%
Decline P in %	34.7%			14.8%		33.3%		52.8%		38.4%		24.1%		45.8%
$\mu_T - \mu_P$	-2.56			-2.06		-2.66		-2.56		-2.56		-2.56		-2.56
Treatment effect in %	11.9%			9.5%		12.3%		11.9%		11.9%		11.9%		11.9%
$z_{1-\alpha}$	1.96			1.96		1.96		1.96		1.96		1.96		1.96
$z_{1-\beta}$	1.28			1.28		1.28		1.28		1.28		1.28		1.28
n	259			196		240		469		90		200		283
Adjustment for losses to follow-up:	23%			23%		23%		23%		23%		23%		23%
n	337			255		312		609		117		260		368

Appendix 9 - Sample size calculations for the Average Hippocampal Volume biomarker

$$n = (z_{1-\beta} + z_{1-\alpha})^2 * (\sigma_T^2 + \sigma_P^2) / (\mu_T - \mu_P)^2$$

	Trial design 18-month dis.-mod. treatment	Full 32 AD patient sample	Sub population (Avg Hippoc Vol Quartile 1)	Trial design Adjusted Avg Hippoc Vol Quartile 1	Sub population (Avg Hippoc Vol Quartile 2)	Trial design Adjusted Avg Hippoc Vol Quartile 2	Sub population (Avg Hippoc Vol Quartile 3)	Trial design Adjusted Avg Hippoc Vol Quartile 3	Sub population (Avg Hippoc Vol Quartile 4)	Trial design Adjusted Avg Hippoc Vol Quartile 4	Sub population (Avg Hippoc Vol Half 1)	Trial design Adjusted Avg Hippoc Vol Half 1	Sub population (Avg Hippoc Vol Half 2)	Trial design Adjusted Avg Hippoc Vol Half 2
Leverage factor scenario: SD & TE levers (μ_T remains unchanged)														
α	0.05			0.05		0.05		0.05		0.05		0.05		0.05
β	0.1			0.1		0.1		0.1		0.1		0.1		0.1
σ_T	9			11.6		7.7		4.7		9.8		10.1		7.7
σ_P	9	7.2	9.8	11.6	5.9	7.7	2.9	4.7	8	9.8	8.3	10.1	5.9	7.7
μ_T	4.94			4.94		4.94		4.94		4.94		4.94		4.94
μ_P	7.5	5.4	9.4	11.5	3.8	5.9	3.1	5.2	5.4	7.5	6.6	8.7	4.3	6.4
Baseline score	21.6			21.6		21.6		21.6		21.6		21.6		21.6
Decline T in %	22.9%			22.9%		22.9%		22.9%		22.9%		22.9%		22.9%
Decline P in %	34.7%			53.2%		27.3%		24.1%		34.7%		40.3%		29.6%
$\mu_T - \mu_P$	-2.56			-6.56		-0.96		-0.26		-2.56		-3.76		-1.46
Treatment effect in %	11.9%			30.4%		4.4%		1.2%		11.9%		17.4%		6.8%
$z_{1-\alpha}$	1.96			1.96		1.96		1.96		1.96		1.96		1.96
$z_{1-\beta}$	1.28			1.28		1.28		1.28		1.28		1.28		1.28
n	259			66		1,351		6,861		308		151		584
Adjustment for losses to follow-up:	23%			23%		23%		23%		23%		0%		0%
n	**337**			**85**		**1,754**		**8,910**		**400**		**151**		**N/A**
Leverage factor scenario: only SD lever ($\mu_T - \mu_P$ remains unchanged)														
α	0.05			0.05		0.05		0.05		0.05		0.05		0.05
β	0.1			0.1		0.1		0.1		0.1		0.1		0.1
σ_T	9			11.6		7.7		4.7		9.8		10.1		7.7
σ_P	9	7.2	9.8	11.6	5.9	7.7	2.9	4.7	8	9.8	8.3	10.1	5.9	7.7
μ_T	4.94			8.94		3.34		2.64		4.94		6.14		3.84
μ_P	7.5	5.4	9.4	11.5	3.8	5.9	3.1	5.2	5.4	7.5	6.6	8.7	4.3	6.4
Baseline score	21.6			21.6		21.6		21.6		21.6		21.6		21.6
Decline T in %	22.9%			41.4%		15.5%		12.2%		22.9%		28.4%		17.8%
Decline P in %	34.7%			53.2%		27.3%		24.1%		34.7%		40.3%		29.6%
$\mu_T - \mu_P$	-2.56			-2.56		-2.56		-2.56		-2.56		-2.56		-2.56
Treatment effect in %	11.9%			11.9%		11.9%		11.9%		11.9%		11.9%		11.9%
$z_{1-\alpha}$	1.96			1.96		1.96		1.96		1.96		1.96		1.96
$z_{1-\beta}$	1.28			1.28		1.28		1.28		1.28		1.28		1.28
n	259			431		190		71		308		327		190
Adjustment for losses to follow-up:	23%			23%		23%		23%		23%		23%		23%
n	**337**			**560**		**247**		**92**		**400**		**424**		**247**

Appendix 10 – Biomarker data of the mild AD patient sample from the ADNI database

RID	MMSE (BL)	Ab1-42 (BL)	Tau (BL)	P-Tau 181P (BL)	Tau/Ab1-42 (BL)	P-Tau181P/Ab1-42 (BL)	Both APOE4 genes (BL)	At least one APOE4 genes (BL)	AVG HIPPO VOL (BL)	AVG HIPPO VOL change (BL->M12)	ADAS 11 change (BL->M12)
366	23.00	140.5	117.4	31.04	0.84	0.22	0.00	0.00	1,901.2	-192.6	17.33
372	23.00	137.3	76.5	34.58	0.56	0.25	0.00	0.00	1,082.4	-110.7	9.67
404	20.00	283.8	166.0	31.55	0.58	0.11	0.00	0.00	1,159.7	11.2	5.66
426	24	154.0	62.1	17.23	0.40	0.11	0.00	1.00	2,170.2	-153.9	3.33
431	24	122.8	115.5	54.54	0.94	0.44	0.00	1.00	1,593.2	-215.7	9.00
470	21	119.6	164.4	40.98	1.37	0.34	0.00	1.00	1,632.5	-69.3	1.00
474	26	133.7	136.2	54.82	1.02	0.41	1.00	1.00	1,699.4	-105.4	9.00
517	20	93.3	93.1	34.51	1.00	0.37	0.00	0.00	1,833.6	-116.1	16.33
535	23	165.4	102.6	28.36	0.62	0.17	0.00	1.00	1,298.8	-85.6	-5.00
543	25	179.2	217.1	76.81	1.21	0.43	0.00	1.00	1,647.5	-297.9	4.67
547	26	211.8	58.0	28.95	0.27	0.14	0.00	0.00	1,706.4	221.4	0.33
565	24	81.6	82.2	28.03	1.01	0.34	1.00	1.00	1,216.0	-72.0	29.66
577	24	150.6	204.2	41.42	1.36	0.28	1.00	1.00	1,221.8	-132.0	12.00
606	23	107.0	142.0	72.16	1.33	0.67	1.00	1.00	1,276.0	7.0	4.00
619	22	129.4	59.7	20.87	0.46	0.16	1.00	1.00	1,670.0	-43.3	2.34
627	23	104.6	270.9	81.81	2.59	0.78	1.00	1.00	1,506.8	-11.0	5.34
724	21	151.6	153.8	44.78	1.01	0.30	0.00	0.00	2,270.8	-319.7	-3.34
733	23	129.9	157.6	50.02	1.21	0.38	1.00	1.00	1,863.2	352.3	9.34
753	24	129.7	109.6	61.90	0.84	0.48	1.00	1.00	1,523.7	-144.1	2.00
754	23	132.3	118.8	35.13	0.90	0.27	0.00	1.00	1,744.3	-152.2	2.00
784	23	163.5	36.5	17.43	0.22	0.11	0.00	0.00	1,553.3	-72.5	-3.66
814	21	138.3	157.5	48.63	1.14	0.35	0.00	1.00	1,279.2	-12.3	8.67
836	26	179.2	91.6	29.50	0.51	0.16	0.00	1.00	1,400.9	-41.2	-4.66
850	22	134.0	90.0	32.13	0.67	0.24	0.00	1.00	1,482.9	-97.2	6.00
852	24	140.3	134.6	32.41	0.96	0.23	0.00	1.00	1,141.4	-42.9	10.67
891	22	160.1	56.6	26.04	0.35	0.16	0.00	1.00	1,835.5	-192.7	1.67
1041	22	146.0	88.0	39.97	0.60	0.27	1.00	1.00	2,202.7	-240.6	-1.33
1082	22	131.3	109.0	30.36	0.83	0.23	1.00	1.00	1,533.6	14.6	12.67
1109	21	162.8	104.4	32.93	0.64	0.20	0.00	1.00	1,549.8	-72.3	3.66
1170	23	132.8	178.8	90.95	1.35	0.69	0.00	1.00	1,614.6	-126.6	4.67
1171	24	185.0	41.0	21.26	0.22	0.11	0.00	1.00	2,370.3	-41.6	-0.33
1221	21	124.9	131.5	88.56	1.05	0.71	0.00	1.00	1,773.6	-262.0	1.00
n	32	32	32	32	32	32	32	32	32.0	32.0	32
AVG	22.9	145.5	119.6	42.5	0.88	0.32	0.3	0.8	1,617.3	-88.0	5.4
MED	23	137.8	112.5	34.5	0.87	0.27	0	1	1,603.9	-91.4	4.3
S.D.	1.61	36.8	53.1	20.4	0.46	0.18	0.47	0.42	331.7	132.7	7.2
MIN	20	81.6	36.5	17.2	0.22	0.11	0	0	1,082.4	-319.7	-5.0
MAX	26	283.8	270.9	90.9	2.59	0.78	1	1	2,370.3	352.3	29.7

References

[1] NI Research. NeuroInvestment, Mid Year Review and Preview. July-August 2008. No. 158-159.

[2] Dubois, B. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ARDA criteria. *Lancet Neurol.* **6**, 734-46 (2007).

[3] Harrison, J. A Neuropsychological Test Battery for Use in Alzheimer Disease Clinical Trials. *Archives of Neurology*, **64**, 9 (2007).

[4] Nasreddine, Z. The Montreal Cognitive Assessment, MoCA: A Brief ascreening Tool For Mild Cognitive Impairment. *JAGS* **53**, 695-699 (2005).

[5] Press release Myriad Genetics June 11, 2007. Myriad Genetics Presents Mathematical Comparison of Disease Modification trial Designs at Alzgheimer's Conference. Viewed online on http://www.myriad.com/news/release/1013529 on 2008-06-02.

[6] Alzforum article posted on 2008-03-31. Alzheimer Activism: How To Modernize Clinical Trials?. Viewed online on www.alzforum.org/new/detailprint.asp?id=1787 on 2008-04-02.

[7] Dubois, B. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ARDA criteria. *Lancet Neurol.* **6**, 736 (2007).

[8] Carlson, , N.E. et al. Trajectories of brain loss in aging and the development of cognitive impairment. *Neurology.* 70: 828-833 (2008).

[9] Jack, C.R. et al. MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. *Neurology*, 60, 253-260 (2003)

[10] Birks, J., Harvey, R.J. Donepezil for dementia due to Alzheimer's disease (Review). The Cochrane Collaboration. JohnWiley & Sons, Ltd (2008).

[11] NDA 20-690/S-026. Ariecpt (donepzil hydrochloride tablets, Aricept ODT (Donepezil Hydrochloride Orally Disintegrating Tablets). Page 5-6.

[12] Grundman, M. Mild Cognitive Impairment Can Be Distinguished from Alzheimer' Disease and Normal Aging for Clinical Trials. *Archives of Neurology*, **61** , 59-66 (2004).

[13] Doody, R.S. Eff ect of dimebon on cognition, activities of daily living, behaviour, and global function in patients with mild-to-moderate Alzheimer's disease: a randomised, double-blind, placebo-controlled study. *Lancet*; 372: 207–15 (2008)

[14] Doody, R.S. Effect of dimebon on cognition, activities of daily living, behaviour, and global function in patients with mild-to-moderate Alzheimer's disease: a randomised, double-blind, placebo-controlled study. *Lancet*; 372: 207–15 (2008)

[15] Wyeth/Elan investor presentation on July 29th 7pm EST presenting Phase 2 reslts of bapinuzemab (disease-modifying AD treatment).

[16] Wyeth/Elan investor presentation on July 29th 7pm EST presenting Phase 2 reslts of bapinuzemab (disease-modifying AD treatment).

[17] Birks, J., Harvey, R.J. Donepezil for dementia due to Alzheimer's disease (Review). The Cochrane Collaboration. JohnWiley & Sons, Ltd (2008).

[18] Doody, R.S. Effect of dimebon on cognition, activities of daily living, behaviour, and global function in patients with mild-to-moderate Alzheimer's disease: a randomised, double-blind, placebo-controlled study. *Lancet*; 372: 207–15 (2008).