Published in final edited form as:

Alzheimers Dement. 2014 March ; 10(2): 152–161. doi:10.1016/j.jalz.2013.05.1776.

Simulating effects of biomarker enrichment on Alzheimer's prevention trials: conceptual framework and example

Jeannie-Marie S. Leoutsakos, PhD^a, Alexandra L. Bartlett^b, Sarah N. Forrester, MS^a, and Constantine G. Lyketsos, MD^a

^aDepartment of Psychiatry, Division of Geriatric Psychiatry and Neuropsychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland USA

^bSTEM Magnet Program, South River High School, Edgewater, Maryland USA

Abstract

Background—We present a conceptual framework for simulations to determine the utility of biomarker enrichment to increase statistical power to detect a treatment effect in future Alzheimer prevention trials. We include a limited set of simulation results to illustrate aspects of this framework.

Methods—We simulated data based on the Alzheimer's Disease Anti-Inflammatory Prevention Trial, and a range of sample sizes, biomarker positive predictive values, and treatment effects. We also investigated the consequences of assuming homogeneity of parameter estimates as a function of dementia outcome.

Results—Use of biomarkers to increase the sample fraction that would develop AD in the absence of intervention from 0.5 to 0.8 would increase power from 0.35 to 0.69 with N=200. Ignoring sample heterogeneity resulted in overestimation of power.

Conclusion—Biomarker enrichment can increase statistical power, but estimates of the expected increase are sensitive to a variety of assumptions outlined in the framework.

Keywords

biomarker enrichment; dementia; Alzheimer's disease; simulation; random effects; statistical power; trajectory

1. Introduction

Maximizing the efficiency of randomized clinical trials (RCT) efficacy in mild cognitive impairment (MCI) and other pre-clinical stages is vitally important to the ultimate goal of prevention of Alzheimer disease (AD)^{1, 2}. Practically speaking, increased efficiency means fewer subjects and lower costs per study, allowing for a larger number of preventive treatments to be tested more rapidly.

© 2013 Elsevier Inc. All rights reserved.

*Corresponding Author: Jeannie-Marie Leoutsakos, Ph.D., MHS, Assistant Professor, Johns Hopkins University School of Medicine, Department of Psychiatry, Division of Geriatric Psychiatry and Behavioral Sciences, Bayview – Alpha Commons Building 4th Floor., Baltimore, MD 21224, Phone: 410-550-9884, Fax: 410-550-1407, jeannie-marie@jhu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

For the purposes of this discussion, we assume that individuals who enroll in an AD prevention trial fall into one of two groups: (1) individuals who will not develop AD dementia in their lifetime, regardless of treatment, and (2) individuals who will develop dementia in the absence of an intervention, but who might respond to treatment. We will refer to the former group as ‘not at risk’, and the latter as ‘at risk’. We assume individuals are members of one of these groups, despite our current inability to predict, with certainty, their outcomes in advance. Individuals who are at risk will likely already have some neuropathology at enrollment. Therefore, we conceptualize treatment response as a *delay* in onset of clinical AD in this at risk group. By MCI, we refer to individuals with measurable cognitive impairment who do not meet diagnostic criteria for dementia, and who may or may not meet criteria for “MCI due to AD”³. These people may stay the same, improve, develop AD, or develop other dementias^{4,5}. Despite their existing impairment, they are pre-clinical in the sense that they have not been diagnosed with clinical dementia.

In planning clinical trials, researchers must balance cost with the necessity of enrolling enough participants to ensure sufficient statistical power to detect a clinically significant effect. Statistical power to detect a given preventive effect between treated and untreated groups would be maximized if 100% of subjects were at risk. Individuals not at risk would likely receive minimal or no benefit from a putative AD preventive treatment, even if it were truly effective for those at risk. Currently, there are no definitive methods to predict who will be diagnosed with clinical AD. Therefore prevention studies, even those that only enroll MCI cases, have samples in which, at best, half are actually at risk^{2, 6-17}. This wastes resources, needlessly exposes research participants to potential risks, and may dilute treatment effects to an undetectable level¹.

There has been substantial interest in the use of biomarkers to identify individuals most likely to be at risk, thereby maximizing statistical power^{1, 18-20}. This enrichment strategy is theoretically sound^{21, 22}, as biomarker enrichment would work by shifting upward the proportion of individuals in the study sample who were truly at risk. Among patients with MCI, for example, the fraction who are at risk is approximately 0.50²³⁻²⁵ but among patients who have MCI and who are also biomarker positive, that fraction would shift upward to 0.65-.80^{20, 26, 27}. We argue that the appropriate simulation of the effects of biomarker enrichment is a comparison of statistical power between simulated datasets which vary as a function of the fraction of individuals who are at risk, ranging from 0.50 (to reflect an MCI inclusion criteria) to 0.80 (the upper bound for performance of the best current biomarkers). In these simulated datasets, the treatment effect is δ for those individuals who are at risk and assigned to the treatment group, and is null for all other individuals. The observed treatment effect when the model is fit using simulated subjects would therefore increase in magnitude toward δ as the at-risk fraction approaches 1.

Simulation studies are sensitive to the assumptions under which they are conducted. Therefore, it is important that principal investigators of future prevention trials have a clear understanding of the meaning and consequences of these assumptions. In this paper, we will identify and describe each of these assumptions and present an explicit framework for future simulation studies. We will also present a small set of simulation results as an illustration of this framework. The conceptual framework is readily applicable to the planning of a wide variety of prevention trials.

2. Conceptual framework

The value of simulations is that they allow for the conduct of statistical experiments. When data from actual clinical trials are analyzed, a model is fit under a set of assumptions, and then an estimate of the treatment effect is produced. In some cases, it is difficult to

determine if those assumptions are tenable, and it is impossible to know the “true” magnitude and variability of the treatment effect and other model parameters. With simulation, we *know* the truth (because we create the data based on that ‘truth’) and are able to observe what happens when models are fit to such data. In some cases²⁸ observed data serve as the basis of simulation, and a treatment effect is added to a subset of the observed data assigned to a hypothetical treatment group. This approach may be more ‘real world’ by capturing additional unknown parameters, but carries the disadvantage that, as in the analysis of observed data, the true magnitudes and variabilities of these other model parameters are unknown. We argue that it is preferable to simulate data “from scratch”, because the magnitudes, and particularly the variabilities, of other model parameters also impact statistical power to detect a treatment effect. Therefore it is important to know, and to be able to manipulate these values in simulations to determine their effect on power.

The next decision concerns how the data will be modeled. As the goal is prevention, the natural choice might be to model time to clinical AD diagnosis as a function of treatment (e.g., a survival analysis)²⁹. However, this requires a long observation period and a large number of participants in order to have sufficient statistical power³⁰. We, and others³¹, argue that it is more efficient to model cognitive and/or functional trajectories as a function of treatment, where a decrease in rate of decline among treated subjects relative to control subjects would be taken as evidence of efficacy. Such an approach is acceptable to the FDA, when viewed as a treatment for presymptomatic or incipient AD, rather than as prevention *per se*^{1, 31}. Of the 3 AD prevention trials currently underway, all plan to study change in cognition rather than dementia incidence³². We, and others³³, argue that the model of choice is a longitudinal mixed-effects model, also known as a mixed model, random-effects model, multilevel/hierarchical linear model, covariance component model, or random-coefficient regression³⁴. Two alternatives to mixed effects models are repeated measures ANOVA and the method of generalized estimating equations. The former assumes there will be no missing data, and the latter assumes any missing data are missing completely at random³⁴. Given that there will almost certainly be missing data, and that missingness will likely not be completely at random (e.g., older individuals are more likely to die and be lost to follow-up), neither alternative seems appropriate. Further, repeated-measures ANOVA assumes constant variance over time; such an assumption is routinely violated in aging studies.

Mixed effects models consist of fixed effects (β s) and random effects (α s) and operate by fitting separate regression lines for each individual. The fixed effects are the mean intercepts and slopes for each treatment group, and the random effects are the deviations from those mean intercepts and slopes for each individual's fitted trajectory. Figure 1 shows several individual fitted trajectories along with their relationship to the mean fitted trajectories for a placebo and treatment group, respectively. The model is written as $y_{ij} = \beta_0 + \alpha_{0i} + (\beta_1 + \alpha_{1i})\text{time} + \beta_2 \text{treatment} + \beta_3 \text{time} * \text{treatment} + \epsilon_{ij}$, where y_{ij} represents the outcome for the i^{th} individual at the j^{th} timepoint, β_0 and β_1 are the mean intercept and slope over time for the placebo group, and β_2 and β_3 are the difference in mean intercept and slope over time between the two treatment groups. The effect of interest is β_3 , which is the difference in slopes (cognitive trajectories) over time between the means of the treatment groups. If β_3 is statistically significantly different from 0, it can be taken as evidence of a treatment effect. The random effects (α_{0i} and α_{1i}) represent the deviations from the mean intercept and slope, respectively, for the i^{th} individual, and are typically assumed to follow a normal distribution, though other distributions can be assumed. These random effects have means of 0 and variances of $\sigma^2_{\alpha_0}$ and $\sigma^2_{\alpha_1}$ and covariance $\sigma_{\alpha_0, \alpha_1}$. The variance of the random effect for slope is particularly important, as it will directly impact the power to detect a between-group difference in slope³⁵. The residuals ϵ_{ij} are often assumed to be uncorrelated over time within individuals with a mean of 0 and variance σ^2_{ϵ} . The assumption of within-person independence of residuals is not necessary, though. In AD, the modeling of trajectories is

complicated by the fact that subjects' courses are often curvilinear (e.g., they accelerate downward over time)³⁶. There are a number of ways that these changes in trajectory can be accommodated, including piecewise regression³⁷ and the addition of nonlinear (e.g., quadratic or “time-squared”) terms. In the context of clinical trial, we prefer the latter option, as the former necessitates choices about placement of breakpoints which could affect the inference, and which may not be made until the analysis phase is reached. In those cases, the analysis plan would be partially *post hoc* rather than *a priori*, as is typically recommended for clinical trials³⁸. If a quadratic term is added, it is possible that a treatment might have an effect both on the slope of the trajectory *and* on this quadratic term.

Therefore, two additional terms are added to the model to characterize this curvilinear time course and treatment effect, with the new model represented as $y_{ij} = \beta_0 + \alpha_0 i + (\beta_1 + \alpha_1 i) \text{time} + \beta_2 \text{treatment} + \beta_3 * \text{time} * \text{treatment} + \beta_4 * \text{time}^2 + \beta_5 * \text{treatment} * \text{time}^2 + e_{ij}$. It is also possible to include a random effect for the quadratic term, but in practice this is seldom done, and such models often fail to converge³⁹. Since there are now two potential treatment effects, the hypothesis test is conducted by performing a likelihood ratio test for nested models, where twice the difference in log likelihood values for models with and without the two treatment effects (β_3, β_5) is known to follow a chi-squared distribution with degrees of freedom equal to the difference in parameters between the two models (in this case, $df=2$). Descriptions of simulations in which mixed effects models are used should specify the equation for the fitted model, which random effects are included, the distribution of those random effects, and the assumed correlation structure for the residuals. All of these details impact power, and without this level of detail, it would be impossible for others to replicate the simulation, or for readers to determine if the results are relevant to their own trial planning.

In a clinical trial, there are (at least) 4 groups of patients, because individuals are either at risk or not, and either biomarker positive or not. It is important to consider that each of these 4 groups may differ in terms of their meantrajectories⁴⁰ and distributions and variances of their random effects and residuals. For example, individuals who are at risk typically have greater variability in their trajectories as compared to individuals not at risk. This is the case in ADAPT (see table 1), and elsewhere⁴¹. It is also possible that individuals who are at risk, but who are biomarker negative, are atypical in some way, and may differ both in trajectory, variability of trajectory, and treatment response from individuals who are at risk but are biomarker positive. It may be that individual these biomarker negative, at risk individuals begin their decline at a later age, for example. In reporting the results of a simulation, it is necessary to specify what assumptions have been made regarding heterogeneity of these parameters, both for the purposes of simulating the data as well as in fitting the models to that simulated data.

In order to realistically simulate data from these 4 groups, it would be ideal to have longitudinal data from a study in which patients without clinical AD were enrolled, biomarker data were collected at baseline, and participants were followed to death so that it was known whether they had been at risk or not. We are not aware of such a study. The Alzheimer Disease Neuroimaging Initiative (ADNI) is a reasonable option, in that its patients are well-characterized and biomarker data is available, but the follow-up period is short. This prevents us from knowing who is at risk, and also makes it difficult to determine if their trajectories are straight or curved. An alternative is to generate data using characteristics of subjects enrolled in the Alzheimer Disease Anti-inflammatory Prevention Trial (ADAPT)⁴². We argue that using the ADAPT study may be preferable, because although biomarker data are lacking, subjects have been followed for a much longer period, allowing more certainty regarding their at risk status. The longer follow-up also allows for better characterization of the shape of the trajectory. Further, because ADAPT was an AD prevention trial, it is uniquely representative of individuals likely to enroll in future prevention trials. Use of ADNI to simulate data requires the assumption that the trajectories

of biomarker positive individuals do not differ as a function of whether they ultimately develop clinical AD. Such an assumption is untenable; clearly those who develop clinical AD will decline much more than those who do not. Use of the ADAPT data requires an assumption that the trajectories of individuals who ultimately develop clinical AD do not differ as a function of their biomarker status. Specifically, we are assuming that the mean intercept, slope, curvature, and treatment effect, as well as variability of these parameters does not differ among at-risk individuals as a function of biomarker status. Such an assumption may well be inappropriate. We argue that this assumption, while currently untestable, is preferable to the other. If subjects continue to be followed in ADNI, then in the future those (or other) data may become the preferred choice, as this assumption could then be tested.

Typically, effect sizes are parameterized as differences in slope between treated and untreated individuals, divided by baseline standard deviation of the outcome variable. This is problematic in the current setting for two reasons. First, since trajectories are likely to be curvilinear, there may be treatment effects on both the slope and quadratic aspects of the trajectory. Further, since it is likely that those at risk will have different amounts of variability in their outcome scores, even at baseline, a given difference in slope will translate into a different effect size depending on each group's baseline variability²². In the absence of a successful AD prevention trial to provide a realistic treatment effect to simulate, we argue that a reasonable approach would be to define joint differences in slope and quadratic terms which would yield delays in onset of AD of 0.5, 1, 2, and 5 years, respectively, based on commonly used outcome measure cutoffs⁴³. These delays are reasonable because their consequences are considerable and well-characterized⁴⁴.

We alluded previously to the importance of missing data in determining the appropriate statistical model. It is also important to consider missingness during the generation of simulated data. One option is to calculate, from existing data, rates of missingness at each visit for individuals at risk and not at risk, and to randomly assign missingness to the simulated data based on those probabilities. This approach ignores patterns of missingness that are likely to occur. For example, once a person has a missing visit, they are far more likely to miss subsequent visits. We argue that another reasonable approach is to sample real missing data patterns from an existing dataset, and apply those to the simulated data. It is important to note that patterns of missingness may vary as a function of at-risk status and possibly treatment assignment.

In summary, it should be clear that in simulating the effects of biomarker enrichment on statistical power, a wide range of assumptions must be made (and reported) during both the generation and modeling of data. Within the conceptual framework that we have outlined, we present the methods and results of a simulation conducted under one set of assumptions, and using parameter estimates from the ADAPT clinical trial.

3. Methods

ADAPT was a randomized, placebo-controlled, multisite primary prevention trial. Study methods have been previously reported in detail³⁰. Eligible participants were aged 70+ and had at least one first-degree relative with Alzheimer-like dementia. The 2,528 participants received 200 mg of celecoxib bid, 220 mg of naproxen sodium bid, or matching placebo. Randomization was on a 1:1:1.5 basis, stratified by age and study site. At baseline and yearly, participants were evaluated with a cognitive assessment battery, including the Modified Mini-Mental State Examination (3MS)⁴⁵. The 3MS is a widely-used, 100 point global cognitive measure with higher scores indicative of better cognitive functioning. On

December 17th, 2004, treatment was suspended due to safety concerns, but yearly cognitive testing continued until September 2007³⁰.

We fit a longitudinal random effects model to the ADAPT dataset to obtain parameter values (Table 1) on which to base the simulated data. We fit separate models for individuals who were and were not diagnosed with dementia during the study follow-up, and included effects of time (in months) and time-squared (a quadratic term). We included random effects on both the intercept and slope (time effect), and allowed these two random effects to be correlated. We assumed an independent within-person residual correlation structure, and normally-distributed random effects.

We simulated data consistent with the longitudinal mixed effects model described in the conceptual framework with total sample sizes of 100, 200, 400, and 800; and a study lasting 2 years with measurements taken each 6 months. We assumed a two-arm study with individuals assigned 1:1 to treatment or placebo. To determine our treatment effects, we examined graphically a range of joint values of β_3 and β_5 and chose values which produced trajectories that were intermediate in shape (see Figure 2) between those of participants at risk and not at risk and that resulted in the desired delays in clinical disease onset (0.5, 1, 2, and 5 years). For our purposes, clinical onset was considered to have occurred when 3MS score dropped below 80 [94 Teng, E.L. 1987]. We used $\beta_3=.0325$ and $\beta_5=.0005$ to produce a 6 month delay, $\beta_3=.065$ and $\beta_5=.0012$ to produce a 1 year delay, $\beta_3=.0975$ and $\beta_5=.0017$ for a 2 year delay, $\beta_3=.13$ and $\beta_5=.00305$ for a 5 year delay. For simplicity, we assumed that individuals not at risk experienced no treatment effects.

We sampled missing data patterns from ADAPT participants as a function of their at risk status. The rate of missingness for not-at-risk participants was 6%, 12%, 17%, and 19% for the 6, 12, 18, and 24 month visits, respectively. The corresponding rates of missingness for at-risk participants were 2%, 6%, 11% and 12%. The higher retention among at-risk participants may have reflected an increased motivation to remain in the prevention trial.

We simulated datasets with a range of biomarker positive predictive values (PPV) from 0.5 to 0.8. PPV in this context is the probability that a participant is at risk, given they are biomarker positive. For example, if the PPV of a given biomarker is 0.70, then enrolling only those who are biomarker positive should result in a sample that is 70% at risk. PPV is a useful measure of biomarker ‘potency’ because it can be calculated for either a single biomarker or combination of multiple biomarkers, and relates directly to the primary means by which biomarker enrichment improves statistical power. We then simulated longitudinal outcome measures for each person as a function of their risk status and based on the parameters values for each group estimated from the ADAPT trial in Table 1, and the range of treatment effects as shown in Figure 2. Data were simulated to have an independent within-person residual correlation structure, and normally-distributed random effects.

Simulations for each scenario (combination of PPV, sample size, and joint treatment effects) were run on 1000 simulated datasets. Parameters were estimated using a longitudinal mixed effects model with normally-distributed random effects for slope and intercept, and an independent within-person residual correlation structure. To reflect common practice, the regression model was estimated under an assumption of heterogeneity of parameter estimates with regard to at-risk status. A joint treatment effect was considered statistically significant if the p-value for the likelihood ratio test was less than or equal to 0.05. Power for a given scenario was then calculated as the percentage of the 1000 datasets in which a statistically significant treatment effect was found. To demonstrate the effects of assumed variance of the random effect on slope, and assumed variance of residuals during the generation of the data, we ran two additional simulations, both with sample sizes of 100/

group. In the first, we changed the simulated slope random effect standard deviation to be equal to the pooled standard deviation for both groups (0.08), and in the second, we changed the simulated residual standard deviation to be equal to the pooled standard deviation for both groups (2.49). In both cases, as in the previous simulations, the regression models were fit under an assumption of heterogeneity of these parameters. All simulations were conducted in R 2.13.0⁴⁶, and computing code and instructions on its use are available from the first author on request.

4. Results

Table 1 shows parameter estimates from models fit to the ADAPT data. As expected, those who went on to develop dementia had a lower mean intercept and steeper downward decline, as well as greater variability in their slope and residuals. Figure 3 shows statistical power for a range of treatment effects and PPVs, using the simulated parameter estimates as shown in Table 1. As expected, statistical power (shown on the vertical axis) increases with both magnitude of treatment effect and sample size. Further, power increases with PPV, though the increase relative to PPV=0.50 is small when both the treatment effect and sample size are large. Table 2 shows power estimates as a function of sample size and PPV, when the treatment effect is a delay in onset of 2 years. Here the benefits of biomarker enrichment are clear; using a biomarker with a PPV of 0.80 and a sample size of 100/group yields an estimated power of 0.688, while a PPV of 0.50 (e.g., what you might expect by enrolling MCI patients without enrichment) and a sample size of 200/group yields an estimated power of 0.715. In this example, using a biomarker with a PPV of 0.80 would reduce the sample size by half, while maintaining comparable statistical power.

Figure 4 shows simulated power as a function of whether the generation of simulated data takes into account differences in variability of slopes and residuals between those at risk and not at risk. In both cases, sample size is 200 per group. It is clear that using pooled variability estimates (grey lines) for the two groups resulted in an underestimation of the variability of both the slope and residuals in the simulated data sets, leading to an over-estimation of power to detect a statistically significant treatment effect. This example is important, because it demonstrates that biomarker enrichment can also negatively impact statistical power by increasing heterogeneity (in terms of outcome scores) among participants. It is therefore inappropriate to use pooled estimates of variances to simulate data, because they are essentially weighted averages of the variances of the two groups, and when the relative sizes of the two groups shifts, so too must the pooled estimate of variance.

5. Discussion

The purpose of this paper was to provide a conceptual framework for simulations intending to predict the effects of biomarker enrichment on statistical power to detect a treatment effect in an AD prevention trial. Under the limited set of conditions that we explored, power increases as a function of both treatment effect and biomarker PPV. In additional simulations we demonstrated that power estimates were sensitive to assumptions made about specific model parameters (slope and residual variability), and that pooled variance estimates should not be used to simulate data, as they can result in an over-estimation of statistical power.

Our finding that power increased with biomarker PPV was consistent with most other estimations of the effects of enrichment¹⁸⁻²¹. One notable exception is a simulation study which sampled actual MCI cases from ADNI, and compared power when samples consisted of those who were biomarker positive only, versus when no selection was made²⁸. In both types of samples, the authors simulated a null treatment effect for those in a hypothetical control group, and treatment effects were drawn from the same distribution for all treated

participants^{28;47}. Because they did not simulated a shift in the sample fraction that were at risk and could respond to treatment, the only mechanism by which selection could have increased power was through changes in sample variance which may have occurred due to their selection.

There are a number of caveats that should be noted. These simulations were conducted under a very specific set of assumptions, intended to be illustrative only. Researchers planning a prevention trial should conduct a much more comprehensive set of simulations under the range of assumptions appropriate to their application. As noted, our simulations assumed that the distributions of the random effects and residuals within the at risk and not at risk groups would be the same for individuals who were and were not biomarker positive. Continued follow-up of the ADNI or other cohorts may prove this assumption incorrect. Further, we assumed that these distributions would be the same within each group regardless of treatment status; this too may prove untrue. It is even possible that these parameter values, as well as treatment response, may vary as a function of the *interaction* between biomarker status and treatment assignment. For simplicity, our models and simulations assumed constant error variance over time and no practice effects; this may not have been appropriate^{48; 49}. In cases where this is of concern, researchers can simulate data under a wide range of assumptions regarding the within-person error covariance structure. Researchers may also wish to investigate sensitivity to small changes in simulated parameter estimates to ensure that estimates are robust to such variations.

It is possible that treatment effect among those at risk varies as a function biomarker status. Biomarker enrichment would then negatively impact the external validity or generalizability of the study by excluding individuals who are at risk, but are biomarker negative. A specific example relates to the timing of biomarker measurement⁵⁰. Since AD pathogenesis is a dynamic process, it may be that at risk individuals who are biomarker positive are further along in the process, and therefore more or less responsive to a given treatment than at risk individuals who are biomarker negative. It is also possible that at risk individuals would only be biomarker positive during a certain window. For example, certain plasma ceramides are decreased in MCI, but not in normal controls *or* AD⁵¹. In these cases, biomarker PPVs that are specific to the stage of pre-clinical AD in which the putative preventive agent is most likely to be effective should be considered. While this level of specificity may not be currently possible, it will be a byproduct of the growing study of AD biomarkers.

The ADAPT sample itself is enriched already, in that individuals were required to have a first-degree relative with AD-like dementia. This would have increased the percent of individuals who are at risk in the sample. We would argue that this is only problematic, though, if the trajectories or treatment effects of at-risk individuals with a family history are different than those without. In the ADAPT trial, at risk individuals had greater retention, which was unexpected. This may result in parameter estimates which are not generalizable to future studies. As such, researchers might consider conducting simulations with more typical missing data patterns.

In our simulations, the data were simulated based on the same model that was then fit to the data. In actual practice, analysts would not know the “true” model, and fitting a different model would likely negatively impact statistical power. Researchers should consider conducting simulations in which the fitted model is mis-specified. For example, they might simulate data that follow a curvilinear trajectory, but then eliminate the quadratic terms from the fitted model to see how sensitive the power estimates would be to this type of model mis-specification. We assumed that individuals not at risk would experience no treatment affect at all. The effects of biomarker enrichment will be diluted if not at risk individuals also respond to treatment. For example, in previous simulations²¹, biomarker enrichment

was not useful when not at risk patients experienced a 50% treatment effect (relative to the treatment effect in at-risk individuals). Our computing code can accommodate non-zero treatment effects in the not-at-risk sample, expressed as a fraction of the at-risk treatment effect.

We did not include additional covariates in our simulated data. In randomized trials, inferences from adjusted analyses do not typically differ with those from unadjusted analyses, though adjustment can increase statistical power through increased efficiency⁵². Simulations without additional covariates are therefore likely to be slightly conservative.

Going forward, we make a number of recommendations. First, simulation studies should be reported at a level of detail that enables their replication, and computing code should be available upon request. Second, the benefits of increased statistical power must be evaluated within the context of both the costs (human and monetary) of measuring the relevant biomarker, and the number of individuals who would have to be screened in order to enroll the required number of biomarker positive subjects. Careful consideration must be given to assumptions about trajectories and treatment effects, including whether trajectories and treatment effects among biomarker positive at risk individuals can be assumed to generalize to all at risk individuals. Further data collection in ADNI and elsewhere will provide answers to some of these questions. Finally, and most importantly, we strongly encourage principal investigators of prevention trials to ensure that they have a solid understanding of the assumptions and limitations of the methods of sample size estimation that are used for their trials, because final decisions regarding trial design are made by these principal investigators. If they are considering biomarker enrichment, it is important that they understand the mechanism through which it can alter power, as well as the effects it may have on the generalizability of their findings.

Acknowledgments

The work was supported by NIH grantU01-AG15477. The authors wish to thank the ADAPT Research Group for making the data available.

Literature Cited

1. Reiman EM, Langbaum JB, Tariot PN. Alzheimer's prevention initiative: a proposal to evaluate presymptomatic treatments as quickly as possible. *Biomark Med.* 2010; 4:3–14. [PubMed: 20383319]
2. Amieva H, Letenneur L, Dartigues JF, et al. Annual rate and predictors of conversion to dementia in subjects presenting mild cognitive impairment criteria defined according to a population-based study. *Dement Geriatr Cogn Disord.* 2004; 18:87–93. [PubMed: 15087583]
3. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011; 7:270–279. [PubMed: 21514249]
4. Brodaty H, Heffernan M, Kochan NA, et al. Mild cognitive impairment in a community sample: The Sydney Memory and Ageing Study. *Alzheimers Dement.* 2012
5. Aretouli E, Tsilidis KK, Brandt J. Four-year outcome of mild cognitive impairment: The contribution of executive dysfunction. *Neuropsychology.* 2013; 27:95–106. [PubMed: 23106114]
6. Buerger K, Uspenskaya O, Hartmann O, et al. Prediction of Alzheimer's disease using midregional proadrenomedullin and midregional proatrial natriuretic peptide: a retrospective analysis of 134 patients with mild cognitive impairment. *J Clin Psychiatry.* 2011; 72:556–563. [PubMed: 21208578]

7. Devier DJ, Villemarette-Pittman N, Brown P, et al. Predictive utility of type and duration of symptoms at initial presentation in patients with mild cognitive impairment. *Dement Geriatr Cogn Disord*. 2010; 30:238–244. [PubMed: 20847554]
8. Landau SM, Harvey D, Madison CM, et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*. 2010; 75:230–238. [PubMed: 20592257]
9. Luis CA, Barker WW, Loewenstein DA, et al. Conversion to dementia among two groups with cognitive impairment. A preliminary report. *Dement Geriatr Cogn Disord*. 2004; 18:307–313. [PubMed: 15305108]
10. Maioli F, Coveri M, Pagni P, et al. Conversion of mild cognitive impairment to dementia in elderly subjects: a preliminary study in a memory and cognitive disorder unit. *Arch Gerontol Geriatr*. 2007; 44(1):233–241. [PubMed: 17317458]
11. Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand*. 2009; 119:252–265. [PubMed: 19236314]
12. Nordlund A, Rolstad S, Gothlin M, et al. Cognitive profiles of incipient dementia in the Goteborg MCI study. *Dement Geriatr Cogn Disord*. 2010; 30:403–410. [PubMed: 20975271]
13. Perri R, Serra L, Carlesimo GA, et al. Preclinical dementia: an Italian multicentre study on amnesic mild cognitive impairment. *Dement Geriatr Cogn Disord*. 2007; 23:289–300. [PubMed: 17356272]
14. Schmidtke K, Hermeneit S. High rate of conversion to Alzheimer's disease in a cohort of amnesic MCI patients. *Int Psychogeriatr*. 2008; 20:96–108. [PubMed: 17506911]
15. Taragano FE, Allegri RF, Krupitzki H, et al. Mild behavioral impairment and risk of dementia: a prospective cohort study of 358 patients. *J Clin Psychiatry*. 2009; 70:584–592. [PubMed: 19323967]
16. Wang PN, Hong CJ, Lin KN, et al. APOE epsilon4 increases the risk of progression from amnesic mild cognitive impairment to Alzheimer's disease among ethnic Chinese in Taiwan. *J Neurol Neurosurg Psychiatry*. 2011; 82:165–169. [PubMed: 20660919]
17. Mitchell AJ, Shiri-Feshki M. Temporal trends in the long term risk of progression of mild cognitive impairment: a pooled analysis. *J Neurol Neurosurg Psychiatry*. 2008; 79:1386–1391. [PubMed: 19010949]
18. Kohannim O, Hua X, Hibar DP, et al. Boosting power for clinical trials using classifiers based on multiple biomarkers. *Neurobiol Aging*. 2010; 31:1429–1442. [PubMed: 20541286]
19. Lorenzi M, Donohue M, Paternico D, et al. Enrichment through biomarkers in clinical trials of Alzheimer's drugs in patients with mild cognitive impairment. *Neurobiol Aging*. 2010; 31:1443–51. 1451.e1. [PubMed: 20541287]
20. van Rossum IA, Vos S, Handels R, et al. Biomarkers as predictors for conversion from mild cognitive impairment to Alzheimer-type dementia: implications for trial design. *J Alzheimers Dis*. 2010; 20:881–891. [PubMed: 20413876]
21. Maitournam A, Simon R. On the efficiency of targeted clinical trials. *Stat Med*. 2005; 24:329–339. [PubMed: 15551403]
22. Hendrix SB. Requiring an amyloid-beta1-42 biomarker may improve the efficiency of a study, and simulations may help in planning studies. *Alzheimers Res Ther*. 2011; 3:10. [PubMed: 21457497]
23. Larrieu S, Letenneur L, Orgogozo JM, et al. Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. *Neurology*. 2002; 59:1594–1599. [PubMed: 12451203]
24. Matthews FE, Stephan BC, McKeith IG, et al. Two-year progression from mild cognitive impairment to dementia: to what extent do different definitions agree? *J Am Geriatr Soc*. 2008; 56:1424–1433. [PubMed: 18662209]
25. Ishikawa T, Ikeda M. Mild Cognitive Impairment in a population-based epidemiological study. *Psychogeriatrics*. 2007; 7:104.
26. Blom ES, Giedraitis V, Zetterberg H, et al. Rapid progression from mild cognitive impairment to Alzheimer's disease in subjects with elevated levels of tau in cerebrospinal fluid and the APOE epsilon4/epsilon4 genotype. *Dement Geriatr Cogn Disord*. 2009; 27:458–464. [PubMed: 19420940]

27. Mattsson N, Zetterberg H, Hansson O, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA*. 2009; 302:385–393. [PubMed: 19622817]
28. Schneider LS, Kennedy RE, Cutter GR, et al. Requiring an amyloid-beta1-42 biomarker for prodromal Alzheimer's disease or mild cognitive impairment does not lead to more efficient clinical trials. *Alzheimers Dement*. 2010; 6:367–377. [PubMed: 20813339]
29. Vellas B, Aisen PS, Sampaio C, et al. Prevention trials in Alzheimer's disease: an EU-US task force report. *Prog Neurobiol*. 2011; 95:594–600. [PubMed: 21925234]
30. Martin BK, Szekely C, et al. ADAPT Research Group. Cognitive function over time in the Alzheimer's Disease Anti-inflammatory Prevention Trial (ADAPT): results of a randomized, controlled trial of naproxen and celecoxib. *Arch Neurol*. 2008; 65:896–905. [PubMed: 18474729]
31. Aisen PS, Andrieu S, Sampaio C, et al. Report of the task force on designing clinical trials in early (predementia) AD. *Neurology*. 2011; 76:280–286. [PubMed: 21178097]
32. Mullard A. Sting of Alzheimer's failures offset by upcoming prevention trials. *Nat Rev Drug Discov*. 2012; 11:657–660. [PubMed: 22935790]
33. Coley N, Gardette V, Cantet C, et al. How should we deal with missing data in clinical trials involving Alzheimer's disease patients? *Curr. Alzheimer Res*. 2011; 8:421–433.
34. Fitzmaurice, GM.; Laird, NM.; Ware, JH. *Applied Longitudinal Analysis*. John Wiley & Sons; Hoboken, NJ: 2011.
35. Schneider LS, Kennedy RE, Cutter GR. Estimating power with effect size versus slope differences: Both means and variance matter. *Alzheimer's & dementia: the journal of the Alzheimer's Association*. 2011; 7(2):247.
36. Leoutsakos JM, Muthen BO, Breitner JC, et al. Effects of non-steroidal anti-inflammatory drug treatments on cognitive decline vary by phase of pre-clinical Alzheimer disease: findings from the randomized controlled Alzheimer's Disease Anti-inflammatory Prevention Trial. *Int J Geriatr Psychiatry*. 2012; 27:364–374. [PubMed: 21560159]
37. Small BJ, Dixon RA, McArdle JJ. Tracking cognition-health changes from 55 to 95 years of age. *J Gerontol B Psychol Sci Soc Sci*. 2011; 66(1):153–61. [PubMed: 21196437]
38. Piantadosi, S. *Clinical Trial: A Methodological Perspective*. John Wiley and Sons; Hoboken, New Jersey: 2005.
39. Naan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med*. 1997; 16:2349–2380. [PubMed: 9351170]
40. Kennedy RE, Schneider LS, Cutter GR. Biomarker Positive and Negative Subjects in the ADNI Cohort: Clinical Characterization. *Curr Alzheimer Res*. 2012
41. Chen P, Ratcliff G, Belle SH, et al. Patterns of cognitive decline in presymptomatic Alzheimer disease: a prospective community study. *Arch Gen Psychiatry*. 2001; 58:853–858. [PubMed: 11545668]
42. Meinert CL, McCaffrey LD, et al. ADAPT Research Group. Alzheimer's Disease Anti-inflammatory Prevention Trial: design, methods, and baseline results. *Alzheimers Dement*. 2009; 5:93–104.
43. Hayden KM, Khachaturian AS, Tschanz JT, et al. Characteristics of a two-stage screen for incident dementia. *J Clin Epidemiol*. 2003; 56:1038–1045. [PubMed: 14614994]
44. Brookmeyer R, Gray S, Kawas C. Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *Am J Public Health*. 1998; 88:1337–1342. [PubMed: 9736873]
45. Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. *J Clin Psychiatry*. 1987; 48:314–8. [PubMed: 3611032]
46. R Development Core T. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2011.
47. Schneider LS, Kennedy RE, Cutter GR. Estimating power with effect size versus slope differences: Both means and variance matter. *Alzheimer's & dementia: the journal of the Alzheimer's Association*. 2011; 7(2):247.
48. Abner EL, Dennis BC, Mathews MJ, et al. Practice effects in a longitudinal, multi-center Alzheimer's disease prevention clinical trial. *Trials*. 2012; 13:217–6215. 13–217. [PubMed: 23171483]

49. Terrera GM, Brayne C, Matthews F, et al. One size fits all? Why we need more sophisticated analytical methods in the explanation of trajectories of cognition in older age and their potential risk factors. *Int Psychogeriatr*. 2010; 22:291–299. [PubMed: 19906326]
50. Wu L, Rosa-Neto P, Gauthier S. Use of biomarkers in clinical trials of Alzheimer disease: from concept to application. *Mol Diagn Ther*. 2011; 15:313–325. [PubMed: 22188635]
51. Mielke MM, Zandi PP, Shao H, et al. The 32-year relationship between cholesterol and dementia from midlife to late life. *Neurology*. 2010; 75:1888–1895. [PubMed: 21068429]
52. Pocock SJ, Assmann SE, Enos LE, et al. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002; 21:2917–2930. [PubMed: 12325108]

Systematic review

It has been suggested that samples for clinical trials for Alzheimer's disease (AD) prevention should be enriched with individuals who have biomarkers associated with increased risk of AD. However, a recent simulation study suggested this would not improve power to detect a statistically significant preventive effect. We argue that biomarker enrichment can increase power, under certain circumstances.

Interpretation

We present a conceptual framework for the design of simulation studies. Within that framework, we demonstrate several specific cases where biomarker enrichment would increase power.

Future directions

Simulation studies are only meaningful when the assumptions on which they are based are both clear and reasonable. Going forward, we argue for greater understanding of these issues among future prevention trial principal investigators.

Figure 1. Mixed Effects Model Parameters

Group mean (black) and individual (grey) trajectories are shown for each treatment group, corresponding to the regression equation $y_{ij} = \beta_0 + \alpha_{0i} + (\beta_1 + \alpha_{1i})\text{time} + \beta_2\text{treatment} + \beta_3\text{time} \times \text{treatment} + e_{ij}$, where β_0 and β_1 are the mean intercept and slope over time for the placebo group, and β_2 and β_3 are the difference in mean intercept and slope over time between the two treatment groups. The random effects have means of 0 and variances of $\sigma^2_{\alpha_0}$ and $\sigma^2_{\alpha_1}$ and covariance $\sigma_{\alpha_0, \alpha_1}$. Likewise, the residuals e_{ij} are typically assumed to be uncorrelated over time within individuals with a mean of 0 and variance σ^2_{ϵ} .

Figure 2. Fitted Trajectories for At Risk, Not At Risk, and AD Onset Delays

Figure 3. Simulated Power as a Function of Sample Size, Treatment Effect, and PPV.

Figure 4. Simulated Power As a Function of Assumptions about Slope and Residual Variances

Table 1
Parameter Estimates Derived from the ADAPT Clinical Trial

	Combined (N=2528)	No Dementia (n=2446)	Dementia (n=82)
Intercept [*]	94.05 (0.07)	94.17 (0.07)	90.21 (0.57)
Slope [†]	-0.03 (0.01)	-0.03 (0.01)	-0.14 (0.06)
Quadratic [‡]	<0.001 (<0.001)	<0.001 (<0.001)	-0.004 (0.001)
SD [§] (Intercept)	2.61	2.59	1.92
SD(Slope)	0.08	0.04	0.20
Correlation(Intercept, Slope)	0.27	0.17	-0.01
SD(Residual [¶])	2.49	2.33	5.06

* Model-based baseline 3MS score

† Model-based change in 3MS score per month

‡ Model-based change in 3MS score per month squared

§ For specification of a multivariate normal random effects distribution for the intercept and slope with means of 0 and variance-covariance matrix based on the standard deviations and correlation of the intercept and slope.

¶ For specification of normally distributed residuals.

Table 2
Simulated Power as a Function of PPV and Sample Size with a Delay of 2 Years

		Sample Size			
		50/group	100/group	200/group	400/group
Positive Predictive Value	0.50	0.199	0.352	0.715	0.962
	0.60	0.253	0.499	0.829	0.988
	0.70	0.323	0.623	0.918	0.997
	0.80	0.404	0.688	0.956	0.999