Published in final edited form as:

JAMA Neurol. 2014 August ; 71(8): 947–949. doi:10.1001/jamaneurol.2014.1120.

Secondary Prevention trials in AD: the challenge of identifying a meaningful endpoint

Richard J. Kryscio, Ph.D

Department of Statistics, Department of Biostatistics, Sanders-Brown Center on Aging, University of Kentucky, Lexington, KY

Richard J. Kryscio: kryscio@email.uky.edu

The chief difference between a primary prevention trial (PPT) and secondary prevention trial (SPT) in Alzheimer's disease (AD) is related to the targeted participant group. As currently conceptualized, a PPT recruits asymptomatic individuals from the general population while an SPT recruits only asymptomatic (or preclinical) individuals who are biomarker-positive for AD¹. Recent examples of the latter include the AD kindred trial in Antioquia, Colombia, which recruits only subjects who are E280A mutation carriers with a familial history of early onset disease, and the Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease (A4) trial, which targets adults age 70 years and older with high brain amyloid levels as determined by a PET scan. The latter is the subject of the Donohue et al. manuscript², which seeks to identify a primary endpoint for that SPT.

The widely accepted gold standard endpoint of any prevention trial is AD incidence. It takes many years and a large number of subjects to demonstrate that an effective treatment delays or decreases incidence³. PPTs are usually simple but lengthy trials that are often affected by attrition, competing risks, and non-compliance issues, which makes interpretation of the findings difficult. Hence, investigators often seek alternatives by designing trials with surrogate endpoints. This is most appealing in the SPT setting due to the targeted recruitment of participants. Use of surrogate endpoints often abbreviates the trial length, attenuates the effects of competing risks, improves compliance, and in many cases, the minimum number of subjects needed to attain adequate statistical power for the endpoint of the trial. These advantages have to be balanced against demonstrating that the chosen endpoint of an SPT is a true surrogate for incidence. The many examples of failed surrogates from other fields of research including oncology, HIV, and cardiology necessitate warning investigators that this approach can lead to an SPT that is not effective in reducing incidence. Hence, this is part of the challenge investigators face when planning SPTs for AD.

The use of an SPT also addresses the increasing public pressure to identify a treatment that will slow down the impending epidemic of AD cases worldwide. Years of basic science research are paying off with translational studies that address some of the pathways commonly thought to play a critical role in AD including the amyloid, tau, inflammatory, and oxidation pathways. Evaluation of many potentially effective therapies will require informative clinical trial designs in the SPT setting⁴. This works best when the mechanism of action is well understood and when the drug operates most effectively to alter the

biological pathway or mechanism of the disease. It also works best when there is adequate information available on the target population for the SPT to inform the design of the trial.

Timing is an issue in constructing a primary endpoint for an SPT. Increasing evidence shows that the process of decline into dementia, especially in terms of neuropathology, begins decades before the clinical symptoms are appreciated. Unfortunately, most biomarkers and cognitive instruments detect very small changes over this long time span and fail to capture preclinical disease. Abrupt changes do occur but these are often difficult to detect prospectively. For example, a retrospective analysis of the cognitive assessments in the Baltimore Longitudinal Study of Aging shows that episodic memory takes a downturn seven years prior to dementia, while deficits in executive function begin to occur three years before conversion, and deficits in verbal intelligence occur in close proximity to the conversion⁵. There is considerable variability in these estimates across individuals. Choosing poor performers on any cognitive instrument for selecting participants in a prevention study is perilous due to the regression to the mean phenomenon. Hence, it is essential that a biomarker be used to select participants in an SPT and that this be paired with a primary endpoint that can capture the multifaceted changes that will occur in the short run in the individuals testing positive for that biomarker. For this reason composite endpoints are attractive in SPTs.

There are many methods for constructing composite endpoints; the three most popular being principal components analysis (PCA), item response theory (IRT), and mean to standard deviation ratio (MSDR)^{6,7}. PCA is the least preferred method since it relies on an analysis of only the variances and correlations among the measures involved in the composite but ignores the change in mean response, which is often of interest in the clinical trials setting. IRT, which is grounded in modern psychometric theory, works best when the instruments being combined lie along one latent construct (e.g., memory or executive function). Although limited in this fashion, it is much easier to establish internal and external validity when correlating changes in the derived IRT solution with change in biomarkers. However, composites that span several cognitive domains are likely to have higher power than those that rely on one domain.

Donohue et al. chose the MSDR route since this allows combining instruments across cognitive domains. The change in their composite from baseline to final visit is the primary endpoint, taking into account the standard deviation associated with this change. The proposed composite is based on averaging normalized versions of two measures of memory (Free and Cued Selective Reminding Test or FCSRT and delayed recall in the Wechsler Logical Memory task), a measure of processing speed (Digit Substitution test), and a measure of global function (Mini-Mental State Exam or MMSE). It is denoted by the acronym Alzheimer's Disease Cooperative Studies Preclinical Alzheimer Cognitive Composite (ADCS-PACC). The authors choose to weigh each instrument equally in the composite, while others suggest using some criteria such as maximizing the probability that an individual will experience a decline over time in the derived composite⁷. The authors have chosen a pragmatic solution with equal weighing since their derivation is based on a retrospective analysis, and it is unclear that assigning more weight to one instrument over another will be optimal in the upcoming A4 trial. They argue that since they are relying on a

retrospective data analysis, there is insufficient evidence to weigh more heavily one component over another at this point in the A4 trial.

The composite endpoint needs to be reliable and have validity; it is not enough to argue that the instruments making up the composite each have these properties since the weighting of the instruments in the composite will change these properties. Validation is best done in the prospective setting. In addition, to be effective the composite must be responsive over time within the target population. Since decline in a composite measure in the subgroup of biomarker-positive elderly is of interest, the composite must show that it is sensitive to change over time in that subgroup. The proposed ADCS-PACC composite demonstrates a mixed result for decline over time using a retrospective data analysis on three data sets. The analysis based on the Alzheimer Disease Neuroimaging Initiative dataset fails to show a difference between the amyloid beta positive ($A\beta+$) group and the $A\beta-$ group, but since this dataset does not contain information on the FCSRT, a substitute had to be chosen. The analysis based on the Australian Imaging, Biomarker, and Lifestyle Study on Aging also lacked data on the FCRST. But, using a substitute, it did show a significant difference between the $A\beta+$ and $A\beta-$ subgroups at 24 months in the proposed composite. The ADCS Prevention Instrument (ADCS-PI) study needed substitutes for the MMSE and the Logical Memory test but showed differences between subgroups defined by APOE $\epsilon 4$ status. A corresponding analysis based on stratifying participants by progression status showed, as would be expected, an even greater difference. These analyses illustrate that despite using data from well conducted longitudinal studies, it is difficult to establish validity for a proposed composite when no SPTs have been conducted to date.

A second consideration is the role of missing data. The analysis tool chosen by the authors is the mixed model of repeated measures (MMRM) analysis, a well-accepted tool that uses all data available in the analysis. That is, if a participant misses one or more follow-up visits in any of the longitudinal studies mentioned above, that person's observed responses are included in the MMRM. This is valid under a data missing at random assumption, which implies that a missing observation or visit on a patient is not due to the missing value for that visit. While not usually violated in an observational study, this can be easily violated in a randomized trial. There is no simple way to adjust for data missing not at random without some auxiliary information that is rarely available in practice. Data missing not at random affect the power of the study and may bias its findings.

A conservative assumption made by Donohue et al. is that changes would occur, mostly at the last visit so that the primary endpoint is based on changes at that visit compared to baseline. Power is affected by the number and spacing of visits included in the primary endpoint⁸. Hence, if as often is the case when the Alzheimer Disease Assessment Scale-Cognitive Subscale is used as the primary endpoint in a dementia treatment trial, decline is linear over the short run. This allows investigators to use the slope of the decline to power the trial. The figures for the AIBL and ADCS-PI by APOE $\epsilon 4$ studies support a linear decline in the $A\beta+$ groups. The problem is that the changes in the $A\beta-$ groups are affected by practice effects since those groups are likely not declining over time. Thus, obtaining an accurate estimate of the effect size for slope as the primary endpoint is difficult.

The accuracy of a composite endpoint may be affected by the sensitivity and specificity of the diagnostic instrument used to screen participants into the high risk group. In the A4 trial, being amyloid positive (determined by use of a cut point) could create a pool of false positives who may not decline at the same rate as true positives. This increases the variance associated with the composite endpoint and could affect the power of the study. Similarly, false negatives increase the cost of the trial since additional subjects need to be screened to enter the study.

In summary, the authors are to be congratulated for conducting a careful and detailed retrospective data analysis. With so few studies to guide them through virtually uncharted territory, several hurdles needed to be addressed. Given the lack of a discovery of a neuroprotective agent for treating patients with AD, investigators are correct in turning their attention to prevention studies. Given that PPTs take too long to conduct, the use of SPTs with surrogate endpoints is attractive. Hence, the era of SPTs in AD is now being launched. Biomarkers are useful for selecting participants in these studies, but no biomarker has yet been validated to become a primary endpoint in these studies^{9,10}. A drug cannot be approved on that basis. This sets the stage for using composite endpoints based on combining cognitive decline across multiple domains as the primary endpoint of an SPT. The approach by Donohue et al. is a major first step in meeting the challenge of designing effective and informative SPTs. Given all the advances in the basic sciences of AD research and the competing theories on what pathways are responsible for the disease, the time is right to adopt this alternate approach to identifying potential effective treatments for preventing a devastating illness.

References

1. Pillai JA, Cummings JL. Clinical trials in predementia stages of Alzheimer disease. *Med Clin North Am.* 2013; 97(3):439–457. [PubMed: 23642580]
2. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurology.* to appear.
3. Kryscio RJ, Mendiola MS, Schmitt FA, Markesbery WR. Designing a large prevention trial: statistical issues. *Stat Med.* 2004; 23:285–296. [PubMed: 14716729]
4. Carrillo MC, Brashers R, Logovinsky V, et al. Can we prevent Alzheimer's disease secondary "prevention" trials in Alzheimer's disease. *Alzheimers Dement.* 2013; 9(2):123–131. [PubMed: 23411394]
5. Grober E, Hall CB, Lipton RB, Zonderman AB, Resnick SM, Kawas C. Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. *J Int Neuropsychol Soc.* 2008; 14(2):266–278. [PubMed: 18282324]
6. Crane PK, Calne A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav.* 2012; 6:502–516. [PubMed: 22782295]
7. Xiong C, van Belle G, Chen K, et al. Combining multiple markers to improve the longitudinal rate of progression-Application to clinical trials on the early stage of Alzheimer's disease. *Stat Biopharm Res.* 2013; 5(1)10.1080/19466315.2012.756662
8. Ard MC, Edland SD. Power calculations for clinical trials in Alzheimer's disease. *J Alzheimers Dis.* 2011; 26(Suppl 3):369–377. [PubMed: 21971476]
9. Schneider LS, Kennedy RE, Cutter GR. the Alzheimer's Disease Neuroimaging Initiative. Requiring an amyloid- β_{1-42} biomarker for prodromal Alzheimer's disease or mild cognitive impairment does not lead to more efficient clinical trials. *Alzheimers Dement.* 2010; 6(5):367–377. [PubMed: 20813339]

10. Beckett LA, Harvey DJ, Gamst A, et al. The Alzheimer's Disease Neuroimaging Initiative: annual change in biomarkers and clinical outcomes. *Alzheimers Dement*. 201-; 6(3):257–264. [PubMed: 20451874]