# Designing a large prevention trial: statistical issues

Richard J. Kryscio<sup>1,2,3,\*,†</sup>, Marta S. Mendiondo<sup>2,3</sup>, Frederick A. Schmitt<sup>3,4</sup>, and William R. Markesbery<sup>3,4,5</sup>

<sup>1</sup>Department of Statistics, University of Kentucky, U.S.A. <sup>2</sup>School of Public Health, University of Kentucky, U.S.A. <sup>3</sup>Alzheimer's Disease Research Center, University of Kentucky, U.S.A. <sup>4</sup>Department of Neurology, University of Kentucky, U.S.A. <sup>5</sup>Department of Pathology, University of Kentucky, U.S.A.

#### SUMMARY

Recent research in Alzheimer's disease (AD) is centred about the early detection and prevention of this disease. Several recent moderate size clinical trials targeted at high risk cohorts have been designed along this theme. There have been few attempts to design a large trial to prevent this disease in elderly individuals at low risk for the disease. The purpose of this paper is to suggest a framework for designing a simple, large AD prevention trial. This framework uses a discrete time hazard model for decreasing the incidence of AD when participants are randomly assigned to one or more active prevention agents or placebo. This design allows for differential incidence among participants due to age, family history, genetic disposition, and ethnicity. It takes into account the length of the follow-up period, participant mortality, drop-outs, drop-ins, and loss to follow-up. This framework is illustrated by PREADVISE, a recently initiated large add-on prevention trial investigating the use of anti-oxidants for preventing AD among men enrolled in a even larger prostate cancer prevention study, SELECT. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: prevention; clinical trials; sample size; dementia; Alzheimer's disease; aging

#### 1. INTRODUCTION

Alzheimer's disease (AD), the most common form of adult onset dementia, affects approximately 4 million individuals in the United States [1]. With the aging of our society, it has been estimated that approximately 14 million individuals will have AD by the year 2050 [2]. Additionally, there will be an estimated 2–4 million individuals with related dementias such as dementia with Lewy bodies, frontotemporal dementia, and vascular dementia. As of 1998 the estimated cost of caring for a dementia patient in the United States was \$40 000/year [3].

Copyright © 2004 John Wiley & Sons, Ltd.

<sup>\*</sup>Correspondence to: Richard J. Kryscio, Sanders-Brown Center on Aging, 800 South Limestone Street, University of Kentucky, Lexington, KY 40536-0230, U.S.A.

<sup>&</sup>lt;sup>†</sup>E-mail: kryscio@uky.edu

Contract/grant sponsor: NIA; contract/grant number: R01 AG19241

In addition, there is currently no effective therapy for treating this disease [4]. Hence, a major public health problem appears to be on the horizon.

While the search for effective treatments continues, most investigators now believe that prevention is the key to the treatment of this disease. To this end researchers are currently attacking this disease on two fronts. The first approach centres on treatment of high risk subjects. These involve elderly subjects with documented memory loss who do not meet the clinical criteria for the diagnosis of dementia. Several treatment trials for this condition known as mild cognitive impairment (MCI) are near completion; the primary goal of these trials is to delay the onset of AD. Unless a completely effective treatment for MCI is identified, this approach will not necessarily prevent an epidemic of AD because a delay in the occurrence of the disease can only alter the timing of the epidemic.

The second, more recent approach centres around preventing the disease in asymptomatic subjects. The main endpoint of these prevention trials is disease incidence. The statistical design of these trials is the subject of this paper. We begin with a brief review in Section 2 of the prevention trials currently being conducted. We then review factors affecting the incidence of this disease in Section 3. A formula for computing the probability that a naive subject (i.e. dementia free subject) will contract the disease during the follow-up period of the trial is presented in Section 4. In Section 5 we propose a sample size formula based on the well known log rank test statistic to determine the size of a prevention trial. This formula is illustrated by an application to a real trial in Section 6. Concluding remarks are made in Section 7.

## 2. AD PREVENTION TRIALS

There are five phase III AD and other dementia prevention trials that are currently enrolling patients, although others are sure to follow in the near future. All five are double blind, placebo controlled trials with the primary endpoint of AD prevention. Several of these trials have secondary endpoints of delaying memory loss. Some of the salient characteristics of these trials are summarized in Table I. Note that all of these trials have enrollment restrictions. Two of them are open only to individuals with a high risk of AD defined by either a family history of senility, memory loss, AD or other dementia (ADAPT) or simply a family history of AD (PREPARE). The Ginkgo evaluation of memory (GEM) trial enrolled part of its participants (about 8 per cent) from the Cardiovascular Health Study and the remainder from the community at large. Enrollment was completed in 1999 and the study is now in its follow-up stage. Participants in the women's Health Initiative—memory Study (WHI-MS) must first enroll in the Women's health initiative study, a set of multicentre trials on women's health issues [5]. Participants in PREADVISE must first enroll in SELECT, a multicentre trial to prevent prostate cancer. Hence, the latter three studies are ancillary to larger trials.

### 3. FACTORS AFFECTING INCIDENCE

A number of factors of varying importance affect the incidence of AD/dementia; these include age, gender, family history of disease, race/ethnicity, education, and genetic status. Because there is no registry for this disease, independent epidemiologic studies tend to report conflicting

Trial rationale	Trial acronym	Treatment arm(s)	Gender	Age at enrollment	Number sites	Length (years)	Projected N
Anti- inflammatory	ADAPT	Naproxen or Celecoxib	Both	≥70	5	5–7	2800
Alternative medicine	GEM	Ginkgo Biloba	Both	≥75	4	5	3000
Anti-oxidants	PREADVISE	Vitamin E or Selenium or Both	Male	$\geq 62 \text{ or } (\geq 60)^{\dagger}$	400	9–12	10 700
Hormone replacement	PREPARE	Estrogen or Estrogen plus Progestin	Female	≥65	21	5	500
Hormone replacement	WHI-MS	Estrogen <sup>§</sup> or Estrogen plus Progestin¶	Female	$\geq 65^{\ddagger}$	40	6	8300

Table I. Characteristics of current phase III AD/dementia prevention trials.

<sup>†</sup>If minority.

<sup>‡</sup>Must be postmenopausal.

<sup>§</sup>Must have had a hysterectomy.

¶Must have a uterus.

Table II. Age specific AD incidence/1000 person-years by age (both genders combined).

					Age			
Study	AD level $^{\dagger}$	60-64	65-69	70-74	75-79	80-84	85-89	≥90
Jorm [6]	Moderate	_	1.6	3.5	7.8	14.8	26.0	_
Rocca [8]	Moderate	0.5	1.1	3.7	7.8	18.8	32.7	36.2
Kawas [12]	Moderate	0.8	1.3	4.2	8.9	21.6	$64.8^{\ddagger}$	
Gao [7]	Mixed	0.6	1.9	5.1	11.7	23.1	38.9	60.9
Ganguli [9]	Moderate		2.1	4.6	10.0	25.8	26.8	50.9
Ganguli [9]	Mild	_	2.1	8.5	16.1	45.1	48.8	70.2
Jorm [6]	Mild	—	6.1	11.1	20.1	38.4	74.5	

<sup>†</sup>Denotes the minimum level of the disease detectable in the study. <sup>‡</sup>Rate is for 85 or older.

results on some of these risk factors due to case ascertainment methods. A review of the current knowledge base on these risk factors follows. This review emphasizes AD although almost all that is reported below for AD has also been done for dementia as well.

The single most important risk factor is age, and there are numerous studies reporting the relationship between dementia incidence and age. Incidence increases dramatically beginning in the sixth decade of life. Population-based epidemiologic studies published before 1998 are summarized below in the review of two meta-analyses [6, 7], along with more recent individual epidemiologic studies. Age-specific rates for both genders combined are presented in Table II. Note that the rates tend to be consistent across studies after accounting for the method of diagnosis (discussed below) and the standard errors in these estimates (not reported).

With respect to this table, a meta-analysis of 23 population-based incidence studies that reported the severity level of AD as mild<sup>+</sup> (clinical dementia rating (CDR  $\ge 0.5$ ) or moderate<sup>+</sup>

 $(CDR \ge 1.0)$ , and the age-sex-specific rates was conducted [6]. The authors investigated the effect of the following four factors on AD incidence: age, sex, region (Europe versus USA versus East Asia), and diagnostic criteria. The diagnostic criteria used either the Diagnostic and Statistical Manual of Mental Disorders, revised, third edition (DSM-IIIR) versus other AD criteria or National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) versus other AD criteria. Their findings revealed (a) the log of incidence versus age is approximately linear; (b) women have higher rates (especially in older age categories) of mild AD<sup>+</sup>; (c) the East Asian region has lower rates compared with the European region, although the overall region variable was not significant; and (d) DSM-IIIR criteria have lower rates for mild AD<sup>+</sup> but higher rates for moderate AD<sup>+</sup>. Incidence for studies using the NINCDS/ADRDA criteria did not differ from studies using other diagnostic criteria. The rates reported in Table II are only for studies in the U.S.A.

Gao conducted a meta-analysis of the eight studies reported by Jorm and Jolley that relied on the DSM-IIIR criteria and personal interviews for AD diagnosis [7]. A mixed linear model was used to study the effect of age and sex on the rates. They concluded that the incidence of AD increases with age but levels off in older age groups (quadratic effect of age) and women have higher incidence rates (odds ratio = 1.56 with 95 per cent confidence interval: 1.16-2.10).

Age- and sex-specific rates based on 19000 persons aged 50 and older living in Rochester, Minnesota from 1975–1984 have also been reported [8]. Diagnosis was by medical record reviews using the DMS-IIIR and NINCDS/ADRDA criteria at a moderate<sup>+</sup>severity level. Although this study was retrospective, it is useful because it is based on large sample sizes, reports both the numerators and denominators used in constructing the incidence rates, and reports rates by gender. Note that these rates are consistently lower than those reported in other studies, including the findings from two National Mortality Followback surveys conducted by the National Center for Health Statistics in 1986 and 1993.

Age- and sex-specific rates were derived from a prospective study of 1422 participants in the Monongahela Valley Independent Elders Survey (MoVIES) project [9]. The incidence was reported for the diagnosis of AD using the Clinical Dementia Rating (CDR) scales of  $0.5^+$  and  $1.0^+$ . These rates did not vary significantly with sex or education for CDR  $1.0^+$  but for CDR  $0.5^+$  men had higher rates. Persons with less than a high school education also had higher rates. Age-sex specific rates were derived from a study of 1236 participants in the Baltimore Longitudinal Study on Aging [10]. Incidence was reported for definite, probable, and possible AD (moderate<sup>+</sup>) using DSM-IIIR and NINCDS/ADRDA criteria. These rates did not vary significantly with sex or education.

Hence, the studies above indicate that the effect of gender on AD incidence is unclear with only one study reporting positive results. The studies in Table II may not be definitive for the effect of education because several of these studies did not investigate this factor. Also, the attainment of low education has been consistently reported with an increased risk of AD in case-control studies.

Other key risk factors not investigated in these studies include race and family history. In a community-based study of 1079 Medicare recipients, it was shown that race is a major risk factor for AD with African Americans and Hispanics having relative risks of 4.4 and 2.0, respectively, compared with whites. The presence of at least one APOE  $\varepsilon$ 4 allele is associated with an increased risk of AD only among whites [11]. These findings are consistent with the

meta-analysis of Farrer *et al.* [12]. The presence of a first-degree relative with AD increases the relative risk 3.5 times; with two or more first-degree relatives with AD, the risk increases to 7.5.

Finally, AD has a strong genetic component. Corder *et al.* [13] reported that the APOE  $\varepsilon 4$  allele is a major risk factor for late onset AD; others have corroborated this finding [Reviewed by Roses; 14]. Also, it has been demonstrated that the presence of the APOE  $\varepsilon 4$  allele can be associated with a decline in cognitive function in non-demented elderly individuals [15, 16] although the mechanism involved is currently unknown. Other genes associated with AD include mutations in the amyloid  $\beta$ -protein precursor and presenilin 1 and 2 genes which are associated with early onset [17]. However, since this constitutes a low proportion of cases, these will be ignored here; the main thrust of this manuscript is to design prevention trials for late-onset disease. Finally, there are other more minor risk factors associated with the incidence of AD [17].

#### 4. PROBABILITY OF AD DURING THE TRIAL

Let  $p_{\rm C}$  and  $p_{\rm E}$  denote the probabilities that a naïve subject (disease free subject at enrollment) develops or is diagnosed with AD (or dementia) during the clinical trial for members of the control group and the experimental group, respectively. The purpose of this section is to construct formulas for these two quantities based on a discrete time hazards model. A discrete time hazards model is applicable to the situation where subjects enrolled in the trial undergo periodic assessments for dementia (e.g. annual assessments).

To this end, we assume the quantity  $p_{\rm C}$  depends on the following: the length of the followup period for the subject, the incidence rate for the disease in the control arm, the drop-out rate, and the drop-in rate. We assume the quantity  $p_{\rm E}$  depends on the following: the length of the follow-up period of the subject, the incidence rate for the disease in the experimental arm, the drop-out rate and the adherence rate. We now state some assumptions that will allow us to compute these probabilities.

(1) ACCRUALS. In a typical large clinical trial, accruals take place over the first A years; for i = 1, ..., A. Let  $H_i$  = the probability the subject is accrued to the trial during year i where  $H_1 + \cdots + H_A = 1$ . In most trials, plans are made to have  $H_i = 1/A$  to assure that costs associated with administering the trial are kept reasonably equal, but this frequently fails because the  $H_i$  often form a decreasing sequence of numbers. Let D denote the duration of the trial.

(2) INCIDENCE: The incidence rate as a function of age and gender is known to be

 $i_{a,g}$  = incidence rate of AD or dementia for a subject of age a and gender g.

However, the dependence of the incidence rate on other key risk factors—family history, race, APOE genetic status, and low educational attainment—is generally not known as a function of age and gender. As pointed out in the previous section, case-control studies and/or proportional hazards regression models based on large population studies often yield estimates of the relative risk of disease for these risk factors. Thus, under a proportional hazards assumption with respect to age and gender, let  $R_{f,m,\epsilon,b}$  represent the relative risk for a subject with family history f, minority status m, genetic status  $\epsilon$ , and educational attainment b. For simplicity we assume that  $f, m, \epsilon$ , and b are each 1 or 2, implying these risk factors are dichotomous

variables such as the presence/absence of a positive family history, minority status, at least one 4 APOE allele, and low educational attainment, respectively. Generalization to the case where one or more of these factors has more than two levels is straightforward. We further assume that  $R_{1,1,1,1} = 1$ , corresponding to the control condition.

(3) DROP-OUTS: Subjects are often lost to follow-up in large, long clinical trials, even though special procedures are used to minimize these events. We assume that L per cent of subjects are lost to follow up/year, which is assumed to be independent of all factors. Further, since these prevention trials necessarily involve elderly subjects (see Table I), uncontrollable and sometimes substantial losses occur due to subject deaths. Let  $d_{a,g,m}$  denote the known death rate for a subject of age a, gender g, and minority status m. We assume that deaths do not depend on any other risk factors.

(4) DROP-INS: At the beginning of each year of the study, we assume that s per cent of controls switch from the control arm to the experimental arm and remain so for the duration of the study.

(5) ADHERENCE: At the beginning of each year of the study, assume  $s^*$  per cent of subjects switch (permanently) from the experimental arm to the control arm.

(6) REDUCTION IN INCIDENCE: We assume that the experimental treatment reduces AD incidence by the relative risk R < 1 which is independent of all risk factors. We further assume that this reduced risk applies only while the participant is adhering to the experimental treatment; once the treatment is stopped, risk returns to that of a matched control (match on risk factors) with no immunity built up for taking the experimental treatment for several years before becoming a non-adherer.

We now compute  $p_{\rm C}$  by focusing on  $T_{\rm C}$ , the waiting time in years from enrollment to the occurrence of dementia in a randomly selected control subject. According to assumption (1) we have

$$p_{\rm C} = \sum_{t=1}^{A} H_t \sum_{k=1}^{D-t+1} P(T_{\rm C} = k)$$

Note that the probability that  $T_{\rm C} = k$  depends on the risk status of the individual and hence we obtain

$$p_{\rm C} = \sum_{t=1}^{A} H_t \sum_{f=1}^{2} \sum_{m=1}^{2} \sum_{\epsilon=1}^{2} \sum_{b=1}^{2} \sum_{g=1}^{2} \sum_{a=Y_m}^{O_m} W_{f,m,\epsilon,b,a,g} \sum_{k=1}^{D-t+1} P\{T_{\rm C}(f,m,\epsilon,b,a,g) = k\}$$
(7)

Here  $W_{f,m,\varepsilon,b,a,g}$  denotes the proportion of accruals that fall into the strata defined by the risk factors in (2) above. Also,  $Y_m$  and  $O_m$  denote the lower and upper bounds on the age of subjects at the time of recruitment to the trial as defined by the study protocol. These depend on the ethnicity of the participant. If we temporarily ignore drop-outs, then defining the sum of the death rates and lost to follow-up rates to be  $e_{a,g,m} = d_{a,g,m} + L$  yields

$$P\{T_{C}(f, m, \varepsilon, b, a, g) = k\}$$

$$= \begin{cases} (1 - e_{a,g,m})i_{a,g}R_{f,m,\varepsilon,b} & \text{if } k = 1\\ \prod_{j=1}^{k-1} (1 - e_{a+j-1,g,m} - R_{f,m,\varepsilon,b}i_{a+j-1,g})(1 - e_{a+k-1,g,m})R_{f,m,\varepsilon,b}i_{a+k-1,g} & \text{if } k > 1 \end{cases}$$

Copyright © 2004 John Wiley & Sons, Ltd.

Statist. Med. 2004; 23:285-296

Accounting for s, the drop-in rate, and using the fact that a crossover to the experimental treatment can occur at any year prior to disease yields for k = 1

$$P\{T_{C}(f,m,\varepsilon,b,a,g) = k\} = (1 - e_{a,g,m})R_{f,m,\varepsilon,b}[(1 - s) + sR]i_{a,g}$$
(8)

If k > 1, a geometric waiting time argument yields

$$P\{T_{C}(f, m, \varepsilon, b, a, g) = k\}$$

$$= \sum_{l=0}^{k} (1 - s)^{l} s^{\rho_{lk}} \left( \prod_{j=1}^{k-1} (1 - e_{a+j-1,g,m} - R_{f,m,\varepsilon,b} R^{\rho_{lj}} i_{a+j-1,g}) \right)$$

$$\times (1 - e_{a+k-1,g,m}) R_{f,m,\varepsilon,b} R^{\rho_{lk}} i_{a+k-1,g}$$
(9)

Here,  $\rho_{uv} = 1$  if u < v, and 0 otherwise.

Using similar arguments, the quantity  $p_{\rm E}$  can be computed as follows:

$$p_{\rm E} = \sum_{t=1}^{A} H_t \sum_{f=1}^{2} \sum_{m=1}^{2} \sum_{\epsilon=1}^{2} \sum_{b=1}^{2} \sum_{g=1}^{2} \sum_{a=Y_m}^{O_m} W_{f,m,\epsilon,b,a,g} \sum_{k=1}^{D-t+1} P\{T_{\rm E}(f,m,\epsilon,b,a,g) = k\}$$
(10)

Here,  $T_{\rm E}$  is the waiting time for AD in a treated subject. If k = 1,

$$P\{T_{E}(f,m,\varepsilon,b,a,g) = k\} = (1 - e_{a,g,m})R_{f,m,\varepsilon,b}[(1 - s^{*})R + s^{*}]i_{a,g}$$
(11)

while, if k > 1,

$$P\{T_{\rm E}(f,m,\varepsilon,b,a,g) = k\}$$

$$= \sum_{l=0}^{k} (1-s^*)^l s^{*\rho_{lk}} \left( \prod_{j=1}^{k-1} (1-e_{a+j-1,g,m}-R_{f,m,\varepsilon,b}R^{1-\rho_{lj}}i_{a+j-1,g}) \right)$$

$$\times (1-e_{a+k-1,g,m})R_{f,m,\varepsilon,b}R^{1-\rho_{lk}}i_{a+k-1,g}$$
(12)

### 5. SAMPLE SIZE FORMULA

We can use the formulas derived in the previous section to construct a statistical test of the following hypotheses:

$$H_0: p_{\rm C} = p_{\rm E} \quad \text{versus} \quad H_1: p_{\rm C} > p_{\rm E} \tag{13}$$

Since  $p_i$  for i = C, E represents the cumulative probability distribution function of a waiting time variable T until an event occurs in group i, it follows that, assuming a proportional hazards model, the log rank test can be applied to test  $H_0$  versus  $H_1$ . The minimum sample size needed to assure power  $100(1 - \beta)$  per cent, assuming a significance level  $\alpha$  [18], is

$$n \ge \frac{(z_{1-\alpha} + z_{1-\beta})^2}{E^2(p_{\rm C} + p_{\rm E})}$$
(14)

Copyright © 2004 John Wiley & Sons, Ltd.

Statist. Med. 2004; 23:285-296

where  $z_a$  represents the *q*th percentile on the standard normal curve and

$$E = (1 - \Delta)/(1 + \Delta) \tag{15}$$

Here the hazard ratio  $\Delta$ , sometimes called the effect size, is defined by

$$\Delta = \log(1 - p_{\rm E}) / \log(1 - p_{\rm C}).$$
(16)

This formula is based on an asymptotic normal approximation to the sampling distribution of the hazard ratio under the null hypothesis first suggested by Schoenfeld [19]. Since the  $p_i$  are often small quantities, approximations for  $\Delta$  and n are

$$\Delta \approx p_{\rm E}/p_{\rm C}$$

and

$$n \approx (z_{1-\alpha} + z_{1-\beta})^2 (p_{\rm C} + p_{\rm E})/(p_{\rm C} - p_{\rm E})^2$$

### 6. APPLICATION

As mentioned in Section 2, the PREADVISE study has as its primary endpoint the prevention of AD. It is an ancillary study to a much larger trial, SELECT, which has as its primary endpoint the prevention of prostate cancer. SELECT plans to enroll 32 000 men 55 years or older (50 years or older if African-American) at up to 400 sites; men are randomized into the cells of a  $2 \times 2$  factorial design. Factor one is vitamin E or placebo, while factor two is selenium or placebo. As of this writing, SELECT has enrolled over 10 000 men [20]. A man must enroll in SELECT to be eligible for PREADVISE.

In designing PREADVISE, certain assumptions were made by the SELECT trial. With respect to (1), accruals were planned to occur uniformly over a 5-year period, but based on first year experience, it appears that accruals to SELECT will occur more quickly, perhaps over 4 years with 30 per cent of the target enrollment attained in each of the years 1 and 2, and 20 per cent in each of the remaining 2 years. The SELECT trial will last D = 11 years, unless some interim stopping rule is applied to the trial.

With respect to (3) and (4), L=0.5 per cent is the assumed lost to follow-up rate per year, s=1.0 per cent is the assumed drop-in rate, and  $s^*=5.0$  per cent is the assumed non-adherence rate taken to be uniform over all active treatment arms. This is a rather large number, but the philosophy behind this assumption is to avoid losing subjects in the study due to insistence on being compliant with study medications. This philosophy is based on the experience that the Southwest Oncology Group, the co-ordinating centre for SELECT, gained in a recently completed long-term prostate cancer prevention trial in which men were randomized to placebo or the drug finesteride (PCPT) [21]. Many of the sites involved in SELECT participated in PCPT.

To complete the sample size calculations, the death rates  $d_{a,g,m}$  in (3) were taken from the National Vital Statistics Life Tables [22]. The incidence rates  $i_{a,g}$  were taken from Rocca *et al.* [8], who listed incidence based on a large sample of almost exclusively Caucasian males. These rates are listed in Table III below and assume the diagnosis of AD during the trial will be made at a moderate level of the disease. Notice that these rates are conservative because the incidence rate for men 90 years old or older is the same as for men aged 85-89.

Age interval	60-64	65-69	70-74	75–79	80-84	≥85
Incidence/100000	67.5	181.5	392.1	703.7	1610.6	2756.7

Table III. AD incidence rates used in power calculations.

To adjust these rates for minorities (African Americans and Hispanics), it is assumed that 10 per cent of men recruited to PREADVISE will have minority status, even though the stated recruitment goal for the trial is higher (20 per cent). The relative risk for this cohort of minorities will be taken to be 2.0 although even higher risk ratios have been reported in the literature. No adjustments are made for a positive family history, low educational attainment, or APOE status since it is assumed that these are represented in the community based incidence figures quoted above.

Next, two decisions were made in the PREADVISE study design:

- (1) Recruitment from SELECT will be limited to men aged 62 or older (60 or older if an African American or of Hispanic origin). This was based on two facts: the median age at enrollment into SELECT is 62 (verified by the first six months of enrollment) and incidence for men in their 50s is so low that even with 5–10 years of follow-up, we would not find enough conversions to AD to justify following that half of the men enrolled in SELECT.
- (2) Enrollment of men aged 62 or older would decline linearly with age over the effective age range of recruitments into SELECT: 62–88. This was also verified empirically using both PCPT data and early enrollments into SELECT. This defines the proportion W in (7)-(12) above.

Finally, rather than choosing an effect size and computing the minimum sample size required to attain a certain power, it was decided to estimate the accrual to PREADVISE as 10800 men (or 2700 men per arm) and then to determine the detectable effect size for 90 per cent power and significance level 0.05. The estimate 10800 came from a conservative estimate: median age in SELECT is 62 and total accrual to SELECT is 32000, leaving at least 16000 men eligible for PREADVISE. We assumed one out of every three eligible men will not enter the trial either because of personal reasons or because he is enrolled at a SELECT site that will not participate in PREADVISE.

To this end, a SAS macro was created by one of the authors (MSM) to calculate the quantities  $p_{\rm C}$  and  $p_{\rm E}$  given by (8)–(12). This program is available upon request. Working with a reduction in incidence R=0.5 and 0.55, the minimum sample size needed to attain 90 per cent power was computed using (14). These calculations are summarized in Table IV below. Notice that with 2700 men available per arm, the best treatment arm must reduce incidence by the factor R=0.52; this corresponds to a hazard ratio of approximately 0.656, assuming uniform accrual. The best treatment arm in PREADVISE is defined to be the combination therapy: selenium plus vitamin E. On the other hand, if the study is powered to detect a difference in incidence between any treatment arm and placebo, then applying a Bonferroni correction factor to the alpha level reduces this power to 79 per cent (calculation not shown).

	R	= 0.50	R = 0.55				
Accruals	Uniform 5 years	Non-uniform 4 years	Uniform 5 years	Non-uniform 4 years			
<i>p</i> <sub>C</sub>	0.045	0.049	0.045	0.049			
$p_{\rm E}$	0.029	0.032	0.031	0.034 0.689 3031			
$\Delta$	0.639	0.647	0.684				
n	2387	2306	3198				
No. AD cases:							
placebo	107.0	113.0	143.9	148.5			
best treatment	69.2	73.8	99.1	103.1			

Table	IV.	Characteristics	of	the	PREADVISE	study	design	for	90	per	cent	power	and	alpha	0.05
					reductio	n in ir	ncidence								

### 7. DISCUSSION

A randomized controlled trial provides the strongest evidence for the preventive effect of a treatment. This usually requires a large trial to detect the reduction in AD incidence. In this paper we present a formula based on a discrete time hazards model for computing the probability that a subject who is disease free will be diagnosed with AD during a prospective prevention trial. This formula adjusts for differential recruiting by several risk factors known to affect the incidence of the disease including age, gender, race, genetic status, family history of disease, and low educational attainment. It could be extended to apply to other risk factors once they are identified. This formula also adjusts for the effect of drop-outs, drop-ins, and treatment non-adherence.

The formula is illustrated by an application to what is anticipated to be the largest current prevention trial, PREADVISE (see Table I). This trial makes good use of resources because it is an add-on trial to a large prostate cancer prevention trial for men, SELECT. However, this means that the sample size for this trial is predetermined by the limits of SELECT. The application of the formula in this paper indicates that a large reduction in incidence will be needed for the trial to be successful. Meta analyses of case-control studies shows that non-steroidal anti-inflammatory medications could offer a 50 per cent reduction in the odds ratio for the disease, that estrogen replacement therapy could offer a 30 per cent reduction in the same odds ratio, while the same evidence for antioxidants is mixed (cf. Tables II–IV of [23]). These provide evidence that some agents actively under consideration in prospective trials could potentially have a large impact on AD incidence.

Prevention trials do not always work out in practice. For example, the WHI-MS trial was designed to reduce the incidence of any form of dementia by use of estrogen with progestin by 40 per cent. This was based on following 8300 women over 64 years of age for at least 6 years with an anticipated 165 new cases of dementia during that trial. Part of the parent trial WHI was recently discontinued due to increased health risks for women receiving the combined hormone therapy [24]. This overall risk apparently extended to dementia because when the data were examined in the ancillary study, women taking combined hormone therapy had a two fold increase for dementia [25]. Specifically, after accruing only 4532 women to

the WHI-MS trial, and after only a mean follow-up of 4.05 years and 61 incident cases, the hazard ratio against the combined therapy compared to placebo was 2.05 (95 per cent C.I. 1.21-3.48). This result did not however extend to the incidence of mild cognitive impairment given the hazard ratio for MCI of 1.07 (95 per cent C.I. 0.74-1.55). Results for the estrogen only arm are not yet available.

The WHI-MS finding contradicts the results of many other studies including the Cache County Study, a prospective study of incident dementia [26]. The latter study showed that women who used hormone replacement therapy (HRT) were at a reduced risk for AD and that there was a dose effect (cf. Figure 2 of [26]). Some possible explanations for this apparent discrepancy follows. Women enrolled in the WHI-MS were younger than those enrolled in the Cache study (46 per cent versus approximately 20 per cent under 70). Women enrolled in the WHI-MS study were estrogen naïve at baseline and had only an average of 4.05 years of follow-up on estrogen. The Cache County Study showed that incidence curves diverge between HRT users and non-users as age increases (age 80 plus) and that the greatest difference between users and non-users is observed for those that who used HRT for at least 10 years. Hence, the WHI-MS study could be examining the wrong part of the age incidence curve or it could be initiating HRT when it is too late to prevent the disease. It is possible that initiating the therapy at an older age could cause more harm than good. Finally, the Cache study examined incidence of AD while the significant WHI-MS finding applied to all forms of dementia since there was no statistically significant AD risk for HRT users in the WHI-MS study. However, since the WHI-MS is a randomized prospective study its results should be taken seriously and it does demonstrate the need to examine prevention trials on an interim basis.

The detectable effect size in this paper expressed either as a reduction in incidence or as a hazard ratio indicates that future trials will likely have to involve many more subjects because smaller effect sizes are more likely to occur in practice. Another reason the sample size may have to increase is that the proposed formula ignores the effect of misdiagnosis, especially the inability of the diagnostic instrument to identify new cases of disease. As pointed out in Table I, it is also critical to know at which level the diagnostic instrument is operating: mild or moderate cases of disease, since that profoundly affects incidence rates. Since prevention is the key to avoiding the projected large increase in AD and/or dementia cases, more work needs to be done to sort out these issues. The experience gained in the five current AD prevention trials will also add to our base of knowledge on how to design optimum trials in the future.

#### ACKNOWLEDGEMENT

Research supported by the NIA grants: R01 AG19241 AG05144 and AG20146.

#### REFERENCES

- 1. Brookmeyer R, Gray S, Kawas C. Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *American Journal of Public Health* 1998; **88**:1337–1342.
- 2. Katzman R, Fox PJ. The worldwide impact of dementia: projections of prevalence and costs. In *Epidemiology* of Alzheimer's Disease: From Gene to Prevention, Mayeaux R, Christen Y (eds). Springer: Berlin, 1999.
- Petersen RC, Stevens JC, Ganguli M, Tangalos EG. Cummings JC, DeKosky ST. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). *Neurology* 2001; 56:1133–1142.

Copyright © 2004 John Wiley & Sons, Ltd.

#### R. J. KRYSCIO ET AL.

- Kaye JA. Methods for discerning disease-modifying effects in Alzheimer's disease treatment trials. Archives of Neurology 2000; 57:312–314.
- Shumaker SA, Reboussin BA, Espeland MA, Rapp SR, McBee WL, Dailey M, Bowen D, Terrell T, Jones III BN. The Women's Health Initiative Memory Study (WHIMS): a trial of the effect of estrogen therapy in preventing and slowing the progression of dementia. *Controlled Clin Trials* 1998; 19:604–621.
- 6. Jorm AF, Jolley D. The incidence of dementia: a meta-analysis. Neurology 1998; 51:728-733.
- 7. Gao S, Hendrie HC, Hall KS, Hui S. The relationship between age, sex, and the incidence of dementia and Alzheimer's disease: a meta-analysis. *Archives of General Psychiatry* 1998; **55**:809-815.
- Rocca WA, Cha RH, Waring SC, Kokmen E. Incidence of dementia and Alzheimer's disease: a reanalysis of data from Rochester, Minnesota, 1975–1984. *American Journal of Epidemiology* 1998; 148:51–62.
- 9. Ganguli M, Dodge HH, Chen P, Belle S, DeKosky ST. Ten-year incidence of dementia in a rural elderly US community population: the MoVIES Project. *Neurology* 2000; **54**:1109–1116.
- Kawas C, Gray S, Brookmeyer R, Fozard J, Zonderman A. Age-specific incidence rates of Alzheimer's disease: the Baltimore Longitudinal Study of Aging. *Neurology* 2000; 54:2072–2077.
- Tang MX, Stern Y, Marder K, Bell K, Gurland B, Lantigua R, Andrews H, Feng L, Tycko B, Mayeux R. The APOE epsilon-4 allele and the risk of Alzheimer's disease among African Americans, Whites, and Hispanics. *Journal of the American Medical Association* 1998; 279:751–755.
- 12. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, Van-Duijn. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotypes and Alzheimer's disease. A meta analysis, APOE and Alzheimer's disease meta analysis consortium. *Journal of the American Medical Association* 1997; 278:1349–1356.
- Corder E, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993; 261:921–923.
- Roses AD. Apolipoprotein E alleles as risk factors in Alzheimer's disease. Annual Review of Medicine 1996; 47:387–400.
- 15. Bondi MW, Salmon DP, Monsch AU, Galasko D, Butters N, Klauber MR, Thal LJ, Saitoh T. Episodic memory changes are associated with the APOE-epsilon 4 allele in non-demented older adults. *Neurology* 1995; **45**: 2203–2206.
- Riley KP, Snowdon DA, Saunders AM, Roses AD, Mortimer JA, Nanayakkara N. Cognitive function and Apolipoprotein E in very old adults: findings from the Nun Study. *Journal of Gerontology Psychological Sciences and Social Sciences* 2000; 55B:S69–S75.
- 17. Schofield P, Mayeux R. Alzheimer's disease: clinical features, diagnosis and epidemiology. In: *Neuropathology* of *Dementing Disorders*, Markesbery WR (ed.). Arnold Publishers: London, 1998.
- Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. Statistics in Medicine 1982; 1:121–129.
- Schoenfeld D. The asymptotic properties of comparative tests for comparing survival distributions. *Biometrics* 1981; 68:316–319.
- Klein EA, Thompson IM, Lippman SM, Goodman PJ, Albanes D, Taylor PR, Coltman C. SELECT: the next prostate cancer prevention trial. *The Journal of Urology* 2001; 166:1311–1315.
- Thompson IM, Coltman CA, Crowley J. Chemopreventon of prostate cancer: the Prostate Cancer Prevention Trial. Prostate 1997; 33:217–221.
- 22. National Center for Health Statistics, (1999). Life tables for males in the United States, vol. 47. Hyattsville, MD.
- Jorm AF. Prevention of dementia. In *Evidence-based Dementia Practice*, Qizilbash N (ed.). Blackwell Science: Malden, MA, 2002.
- 24. Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association* 2002; **288**:321–333.
- 25. Shumaker SA, Legault C, Rapp SR, Thal L, Wallace RB, Ockene JK, Hendrix SL, Jones III BN, Assaf AR, Jackson RD, Kotchen JM, Wassertheil-Smoller S, Wactawski-Wende J. Estrogen plus progestrin and the incidence of dementia and mild cognitive impairment in postmenopausal women. The women's health initiative memory study: a randomized controlled trial. *Journal of the American Medical Association* 2003, 289(20):2651–2662.
- Zandi PP, Carlson MC, Plassman BL, Welsh-Bohmer KA, Mayer LS, Steffens DC, Breitner JC. Hormone replacement therapy and incidence of Alzheimer Disease in older women: the Cache County Study. *Journal of* the American Medical Association 2002; 288(17):2123–2129.