INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

THE STUDY OF PROGNOSIS IN ALZHEIMER'S DISEASE: A CRITICAL REVIEW AND COMPARATIVE ANALYSIS OF METHODOLOGY

Kathleen Joan Lesperance

Department of Epidemiology and Biostatistics
McGill University, Montreal, Canada
February 1997

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Master of Science

ACKNOWLEDGMENTS

I wish to acknowledge the generosity of Dr. James Mortimer of the Veterans Administration Medical Center, Minneapolis, Minnesota and of Dr. John Brooks, III of the Stanford Alzheimer's Disease Diagnostic and Treatment Center, Palo Alto, California for providing the data upon which the statistical analyses were performed. Marielle Olivier, Todd MacKenzie and Abdissa Negassa are to be thanked for their assistance in the analysis of data. I am also grateful to my thesis committee, consisting of Drs. Tina Wolfson, David Gayton and Michal Abrahamowicz for their enthusiasm and guidance. I particularly wish to thank my supervisor, Dr. Tina Wolfson, for her generosity, constructive feedback, and constant support.

I wish to thank the National Health Research and Development Program, the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche, and the Fonds de la Recherche en Santé du Québec, for their financial support. Lori Bell's helpful comments on an earlier draft of this thesis are greatly appreciated. Finally, I would like to thank my family and my partner, Simon Inwood, for their unfailing support and encouragement.

ABSTRACT

Knowledge of the clinical progression and prognosis of Alzheimer's disease (AD) is important for planning the care of afflicted individuals and evaluating the potential benefits of interventions. There is little consensus, however, regarding the prognostic importance of clinical and demographic characteristics investigated to date. This thesis examined the methodology of prognostic studies of AD through: 1) a critical review of published studies (1984-1995); 2) an assessment of the concordance among different methods of estimating annual rate of change; and 3) an evaluation of the assumption that decline in AD is linear.

A review of 59 eligible studies revealed considerable methodological diversity. The studies also varied in the extent to which they may have been influenced by several sources of bias. Despite this, the findings for some potential prognostic factors were fairly consistent across studies. Illustrative re-analyses of Mini-Mental State Examination (MMSE) data from two longitudinal cohorts of probable AD patients (N=65 and 46) indicated that annual rate of change estimates obtained from the two-point, adjusted two-point, and linear regression methods were comparable. Those of the trilinear model showed poorer concordance. Analyses of data from one cohort confirmed the presence of significant group and individual linear trends in MMSE scores over time and failed to provide evidence of a common quadratic trend.

These findings suggest that prognostic research in AD could benefit from more rigorous study design and further investigation of outcome instruments. Recommendations are made for future research.

RESUME

Une meilleure compréhension de la progression clinique et du pronostic de la maladie d'Alzheimer est importante dans la planification du traitement des personnes atteintes de la maladie et dans l'évaluation des bienfaits potentiels de tout traitement. Cependant, peu s'entendent sur la valeur pronostique des caractéristiques cliniques et démographiques étudiées jusqu'à nos jours. L'objet de cette thèse porte sur la méthodologie des études pronostiques de la maladie d'Alzheimer et propose 1) une analyse critique d'études publiées entre 1984 et 1995, 2) une évaluation de la concordance entre les différentes méthodes d'estimation du taux de variation annuelle et 3) une évaluation de la décroissance linéaire comme assomption dans la maladie d'Alzheimer.

Une revue de 59 études sélectionnées a mis en lumière une grande diversité méthodologique. Les études variaient aussi quant à l'importance qu'ont eu ou auraient pu avoir différentes sources de biais. Malgré tout, certains éléments pouvant servir au pronostic ont été retrouvés de façon plutôt constante parmi les différentes études. Des réanalyses de données provenant de MMSE (de l'anglais Mini-Mental State Examination) fait sur deux cohortes longitudinales de cas probables d'Alzheimer (N=65 et 46), nous indiquaient que le taux de variation annuelle obtenu par les méthodes «two-point», «adjusted two-point» et de régression linéaire étaient comparables. Les résultats provenant du modèle trilinéaire offraient peu de concordance. L'analyse des données tirées d'une cohorte nous a permis de confirmer la présence d'une tendance linéaire significative des résultats de MMSE d'individus et de groupes avec le temps mais ne nous a pas permis de confirmer la présence d'une tendance quadratique commune.

Les résultats issus de notre thèse nous portent à croire que la recherche sur le pronostic de la maladie d'Alzheimer profiterait d'une méthodologie plus rigoureuse des études ainsi que de plus de recherche sur les instruments. Des recommandations à cet effet ont été proposées.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS

AD	Alzheimer's disease
ADAS	Alzheimer Disease Assessment Scale
ADL	activities of daily living
ANOVA	analysis of variance
ARC	annual rate of change
ADDTC	Alzheimer's Disease Diagnostic and Treatment Center
BDS	Blessed Dementia Scale
CDR	Clinical Dementia Rating
CSHA	Canadian Study of Health and Aging
EPS	extrapyramidal signs
EURODEM	European Community Concerted Action on the Epidemiology and Prevention of Dementia
GDS	Global Deterioration Scale
mBIMC	modified Blessed Information-Memory-Concentration test
MIM	multiple interval method
MMSE	Mini-Mental State Examination
NINCDS-ADRDA	National Institute of Neurological and Communicative Disorders and Stroke, and the Alzheimer Disease and Related Disorders Association
RMSE	root mean square error
SD	standard deviation

CHAPTER 1 - INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative brain disorder producing global and progressive decline in higher intellectual functions. Although AD is characterized by loss of memory and other cognitive abilities, afflicted individuals typically experience difficulty in performing activities of daily living and may manifest various behavioral or psychiatric symptoms. The relentless deterioration associated with AD leads to the eventual loss of all verbal and psychomotor skills, and to premature death. The debilitating nature of this disease, its prevalence among the elderly, and its associated economic costs has led to the recognition of AD as an important public health problem (1-3). As a result, basic and epidemiological research on its occurrence, etiology, treatment, and clinical course have expanded significantly in recent years (4).

One growing area of inquiry is the study of prognosis in AD. Understanding the natural history of the disease and the factors influencing its progression is essential in providing patients and their families with accurate predictions of the disease course, in responding to demands on medical and social resources, and in designing studies of potential treatments. Longitudinal studies of AD have been consistent in finding a wide variability among individuals in the rate of disease progression (5). This heterogeneity has prompted researchers to try to identify factors which may be useful in better predicting the disease course of individual patients or groups of patients. Although numerous demographic and clinical features have been investigated, the prognostic importance of most remains controversial (5-8).

This lack of agreement in the literature may be due, in part, to the methodological diversity and limitations of prognostic studies of AD. Researchers have used a variety of instruments to monitor disease progression including various cognitive ratings scales, global staging instruments, and activities of daily living scales (5). Methods used to estimate progression and to assess the factors associated with it have also varied considerably. Inconsistent findings may also reflect variability among the studies in the extent to which they are affected by potential sources of bias. These include, among others, diagnostic inaccuracy, selective attrition, and the insensitivity of measurement instruments to some portions of the range of impairment.

The overall objective of this thesis is to examine the diversity and validity of the methods used in prognostic studies of AD in order to assess the potential for methodology to influence conclusions regarding a factor's predictive ability. Key issues related to study methodology are identified and examined through a critical review of studies assessing potential predictors of cognitive, functional, and disease stage progression. Two issues related to the estimation of disease progression are further examined through re-analyses of longitudinal data on AD patients: whether the assumption of linear decline on outcome scales is valid and whether different methods of estimating the annual rate of change of test scores produce comparable estimates.

1.1 Objectives

1. Through a critical review of published studies on prognosis of AD:
   a) To assess the degree of consistency in the results for individual prognostic factors.
   b) To examine specific features of study design and analysis in terms of the diversity of the methods used across studies.
   c) To examine issues of validity related to specific features of design and analysis.
2. Through re-analyses of data from two longitudinal studies of AD:
   a) To assess the validity of the assumption that decline in AD is linear.
   b) To assess the concordance among five different methods of estimating the annual rate of change of AD.

CHAPTER 2 - OVERVIEW OF ALZHEIMER'S DISEASE

2.1 Introduction

Alzheimer's disease (AD) is a progressive, dementing disorder whose symptoms emerge as early as the fourth decade of life (9). For many years it was considered a cause of dementia only among those with onset prior to age 65 (presenile dementia) (10). The neuropathological hallmarks of AD were subsequently found in the brains of demented elderly (senile dementia) as well (11). As a result of this discovery, and the promotion of a unitary concept of AD (1), the term Alzheimer's disease is now applied to all cases of the disease irrespective of age of onset. As such, it is the most common cause of dementia in the elderly, accounting for approximately 55% to 65% of cases (12-15).

2.2 Occurrence

Information on the prevalence and incidence of Alzheimer's disease (AD) is essential for the planning and allocation of health services to those afflicted with the disease and their families. Moreover, comparing the distribution of AD over time, across geographic areas, and by personal characteristics may provide clues to its etiology (16). Despite the importance of such data, it is only in recent years that research on the prevalence and incidence of AD has begun to proliferate.

Challenges in the study of the frequency of AD include the uncertainty of clinical diagnosis, the difficulty of estimating disease onset, the unsuitability of hospital or death certificate data (17,18), and the considerable time and expense required in population-based research.

2.2.1 Prevalence

Most of the information available on the occurrence of AD comes from prevalence studies (16). The reported estimates from these studies vary widely (19,20). This may be due to differences among the studies in case ascertainment procedures, diagnostic criteria, the inclusion of institutional samples, the age distribution of the populations studied, and the levels of severity sampled (19-22). It may also reflect real differences

in disease frequency. Evidence for the former is provided by a meta-analysis of 22 prevalence studies which found that reported estimates were significantly associated with several study characteristics (19). Whether the variation in prevalence estimates also reflects true geographic differences is difficult to assess given the diversity of study methods and their potential impact on the results.

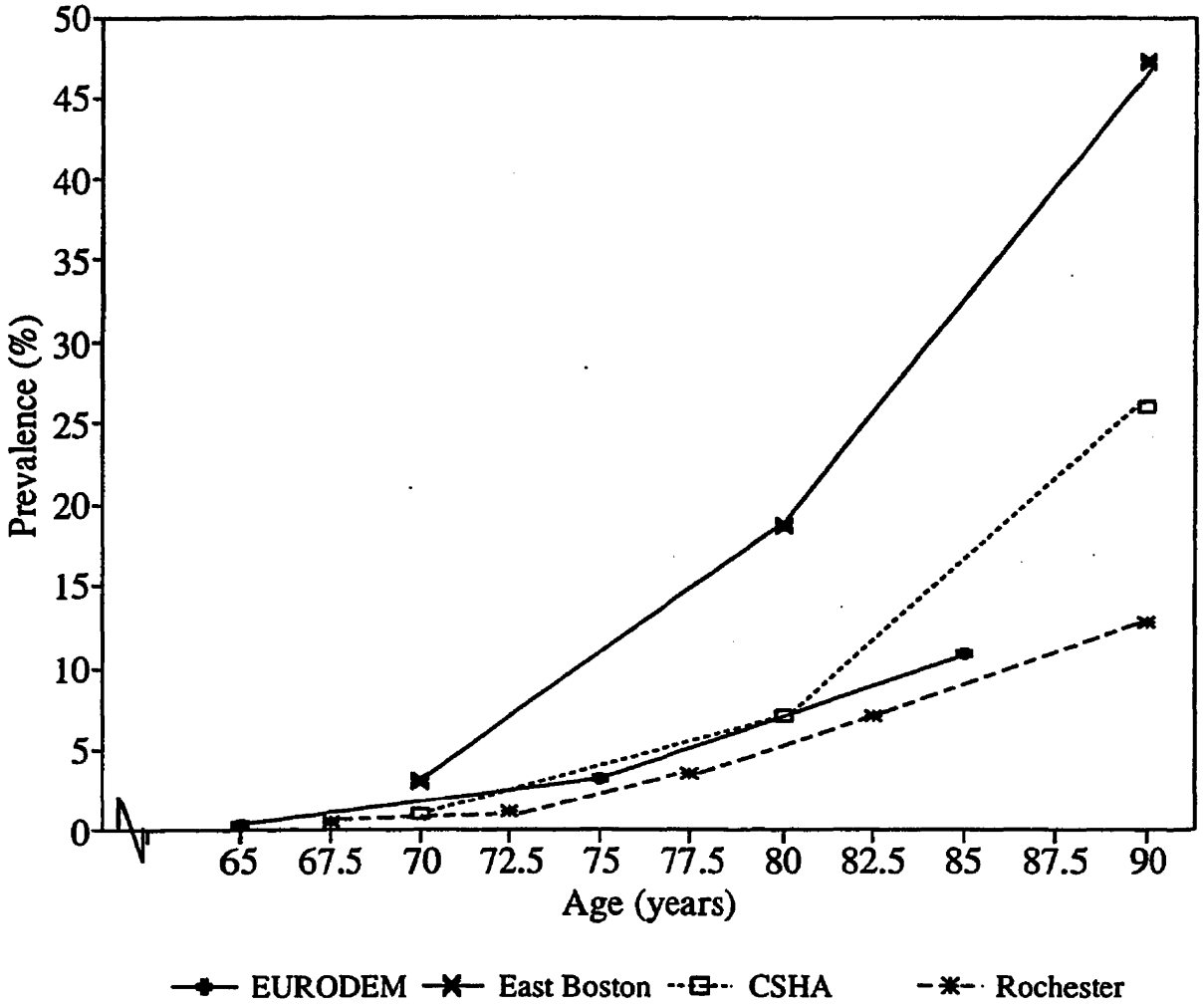
An appreciation for the potential of methodology to influence results in prevalence research is evident in recent reviews of the literature where only those studies meeting criteria designed to enhance comparability were included (21,22). One such effort was the EURODEM (European Community Concerted Action on the Epidemiology and Prevention of Dementia) collaborative re-analysis of European prevalence studies of AD conducted or published between 1980 and 1990 (22). Only six of the 23 studies identified were deemed to be eligible for comparison. Their individual estimates were compared and the original data pooled to provide age-specific prevalence estimates for Europe. No significant geographic differences in prevalence were noted once age and gender were accounted for. The overall European prevalence of AD for the age groups 60-69, 70-79, and 80-89 was reported to be 0.3%, 3.2%, and 10.8%, respectively (Figure 1).

Similar estimates have been reported by Kokmen et al. (23) for Rochester, Minnesota in a study based on the Mayo Clinic records-linkage system (Figure 1), by Bachman et al. (15) in the Framingham cohort, and by Fratiglioni et al. (13) in a Swedish cohort. In contrast, studies in a California retirement community (24) and in East Boston (25) found substantially higher prevalence estimates (Figure 1). The estimates of these latter studies, however, may have been inflated as a result of including cases on the basis of psychometric testing alone (20,26-28).

One of the few national prevalence surveys, the Canadian Study of Health and Aging (CSHA) evaluated elderly community and institutional residents living in five geographic regions of Canada (12). Sixty four percent of all cases of dementia diagnosed at clinical examination were found to suffer from AD, yielding an overall prevalence proportion of 5.1% for Canadians aged 65 and over. The corresponding age-specific estimates were 1.0% for ages 65-74, 6.9% for ages 75-84, and 26.0% for the 85 and over

Figure 1

Comparison of age-specific prevalence of Alzheimer's disease from selected studies*



* Selected prevalence studies: EURODEM (22), East Boston (25), CSHA (12), and Rochester (23).

age group (Figure 1). These values were intermediate to those of other studies.

In contrast to the sizable variation in their reported estimates, prevalence studies have been remarkably consistent in finding a strong positive association between prevalence and age (16). This relationship appears to persist even among the very elderly (13,29). In their 1987 meta-analysis, Jorm et al. (19) observed an exponential increase in the prevalence of AD with age, with a doubling every 4.5 years from age 60 to 90. Though less consistent, studies have also tended to report a higher prevalence of AD among females, a predominance of AD over other dementing disorders in Western Europe and North America, and a comparatively lower prevalence of AD with a predominance of vascular dementia in Japan and China (16,21,30). There is also some evidence to suggest that the prevalence of AD is higher among blacks than whites (31) and among those with little or no education (32,33). Results are sparse or inconsistent, however, and further research is needed to confirm these findings. With respect to secular trends, the overall prevalence of AD was found to be relatively stable from 1947 to 1972 in Lundby, Sweden (34) and from 1975 to 1980 in Rochester, Minnesota (35).

The public health implications of these prevalence estimates is particularly significant in light of the aging populations in most developed countries. Assuming constant incidence and survival, the CSHA investigators applied their 1991 age-specific prevalence proportions to Canadian population projections for the year 2021 and found an expected two and a half fold increase in the number of AD sufferers (12). Jorm et al. (36) applied a statistical model of prevalence to United Nations population projections for 29 developed countries. An overall increase in both the proportion of the elderly and of the demented was forecast for all 29 countries for the period 1980 to 2025. The extent of the projected growth varied, however, with four countries, including Canada, experiencing an increase of approximately 180% or more in the number of dementia cases.

2.2.2 Incidence

If true geographic differences in the prevalence of AD exist, these could reflect differences in incidence, survival or both. It is the comparison of incidence rates,

therefore, that is of greater etiologic value since these are not affected by survival (37). In spite of this, there are far fewer studies of AD incidence than prevalence. In addition to many of the difficulties involved in prevalence research, studies of AD incidence also face the challenge of establishing sizable dementia-free cohorts, of accurately assessing time of disease onset, and of ascertaining the disease status of subjects lost to follow-up (21). As with prevalence, incidence studies have been noted for their substantial variation in methodology and reported estimates (38,39).

A consistent finding, however, among studies of AD incidence is a steep increase in age-specific rates after age 60 (21,39). For example, incidence rates per 100,000 population per year in Rochester, Minnesota were reported to be 31.9, 221.3, 1271.0, and 2600.7 for ages 55-64, 65-74, 75-84, and over 84 for the period 1980 to 1984 (14). Similarly, Bachman et al. (40) observed a doubling of five-year cumulative incidence in successive five-year age groups for ages 65 to 89. It is not clear, however, whether AD incidence continues to increase in the very old or whether it may level off (39,41). Estimates of the lifetime risk of AD in Lundby, Sweden showed no evidence of levelling off up to age 89, at which point the lifetime risk for men and women was 25.5% and 31.9%, respectively (42).

Reports of gender differences in the incidence of AD are inconsistent, with some studies suggesting higher rates among women (34,43) and others reporting no difference (40,44). The comparison of estimates across different populations is hindered by the limited geographic and ethnic coverage of existing studies (21,39).

No significant secular trend was observed for AD incidence in Lundby, Sweden over the 25-year period from 1947 to 1972 (34) or in Rochester, Minnesota for the period 1960 to 1974 (43). A recent update of the latter study, however, noted a trend towards increasing incidence rates among the very elderly for the two quinquennial periods 1975 to 1984 (14). The authors speculated that this finding may have been the result of an ascertainment bias arising from the increasing awareness of AD among health-care providers during the study period.

Additional data on the incidence of AD is forthcoming from a new generation of dementia incidence studies now under way in Europe (38,45) and in Canada. The

European studies have incorporated a common core protocol in their respective designs which will improve the comparability of their data. By supplementing existing data, these studies may help fill the gaps in our present understanding of the incidence of AD and help resolve inconsistencies.

2.3 Etiology

Current knowledge regarding the risk factors for AD is derived primarily from case-control studies (46). These have generally been limited both in terms of internal validity and statistical power (21). With respect to the former, many studies have included prevalent cases which may lead to the identification of factors associated with disease survival rather than incidence (37). Moreover, their reliance on surrogate informants for exposure ascertainment, due to the nature of dementia, has also raised concerns regarding the accuracy of the information and the possibility of bias due to differential recall (47). The problem of low statistical power in individual studies has been addressed through meta-analyses, the most notable being the EURODEM collaborative re-analysis of case-control studies (48).

Most studies of risk factors for AD have investigated multiple hypotheses. As a result, a vast number of genetic, environmental, and psychosocial factors have been examined for their potential association with the occurrence of AD. Of these, only age and a positive family history of dementia are considered established risk factors (21).

An excess of dementia among relatives of patients with AD has been demonstrated in several comparative (49-51) and case-control studies (52,53). The EURODEM re-analysis yielded an odds ratio of 3.5 (95% confidence interval = 2.6-4.6) for AD among those with at least one first-degree relative with dementia (54). There is also some evidence to suggest that the risk of AD increases with the number of affected relatives (52,54) and that the association between AD and family history of dementia is stronger among those who are younger at onset (54).

A positive family history of Down's syndrome has also been identified as a risk factor for AD in some (49,55) but not all (56) studies. The EURODEM study showed a significant increase in risk for those with an afflicted first-degree relative (54). The

latter also indicated that a positive family history of Parkinson's disease was associated with AD, however, two subsequent studies have not confirmed this finding (52,53).

Genetic-linkage studies also support the importance of genetic factors in the etiology of AD. The study of familial clusters of AD has led to the identification of three AD loci on chromosomes 14, 19, and 21 (57). Most cases of AD, however, do not appear to be familial. More recently, a genetic association has been reported between the ε4 allele of apolipoprotein E and late-onset familial and sporadic AD (58,59).

Studies of late maternal age at index birth as a risk factor for AD have yielded contradictory findings (16,46). A re-analysis of four case-control studies indicated an increased risk for those whose mothers were older than 40 at their birth (60). An association between young maternal age and risk of AD has also been suggested (52,60).

The most commonly investigated environmental exposures include head trauma, smoking, alcohol consumption, and aluminum. Although several individual studies have shown no significant association, two meta-analyses, one of eight case-control studies and one of 11 case-control studies, have indicated that prior head trauma is a significant risk factor for AD (61,62). Subsequent studies have reported no clear association (53) and a marginally significant increased risk (52).

Individual studies of the relationship between smoking and risk of AD have produced inconsistent findings, with positive, inverse and no significant associations being reported (21). The EURODEM re-analysis found a significant inverse association, with risk decreasing as consumption increased (63). Results of investigations of alcohol intake and risk of AD have generally been negative (52,53,63). Epidemiologic evidence for aluminum as a risk factor for AD is inconclusive despite some indication of an association from laboratory and ecological research (21,46).

A link between low education and risk of AD has been suggested by some (52,64) but not all (56,65) studies. Various mechanisms have been hypothesized to explain this possible association, including the notion that higher education may lead to an increased brain reserve and thereby delay symptom onset (32,33).

Sparse and often conflicting evidence exists for a variety of other putative risk factors including a medical history of depression (52,66), thyroid disease (52,67), and

severe headaches and migraines (52,67). A history of arthritis and of nonsteroidal anti-inflammatory drug use have been associated with a low risk of AD (52,55). Though the evidence to date is inconclusive for most risk factors studied, it seems likely that several factors, both genetic and environmental, are important for disease expression (68,69).

2.4 Diagnosis

The antemortem diagnosis of AD is hindered by a lack of biological markers and unique clinical features (70,71). A definitive diagnosis must await autopsy since it relies on histologic examination of brain tissue. The examination involves confirming the presence of neuritic or senile plaques and neurofibrillary tangles, the neuropathologic hallmarks of the disease (72). Even these histologic features are not specific to AD, however, as they have also been found in the brains of nondemented elderly (73,74). The differentiation of the disorder from normal aging, therefore, relies on quantifying the neuropathology (72). Since such evidence is rarely obtained during a patient's life, the antemortem diagnosis of AD is most often a clinical one.

The clinical diagnosis of AD consists of first establishing the presence of dementia, a clinical syndrome characterized by intellectual deterioration sufficient to interfere with social or occupational performance (75). This is generally accomplished by means of clinical and mental status examination, and medical history taking, which together serve to identify abnormalities in cognition and establish a history of decline. Confirming the presence of dementia can be particularly difficult in the early stages where cognitive deficits are subtle and may be denied by the patient (71).

Once the presence of dementia is established, the differentiation of AD from other dementing disorders is attempted using the results of medical history, physical and neurological examination, laboratory tests, and brain imaging (4,76). These data are useful in identifying dementias of vascular origin, such as multi-infarct dementia, and secondary dementias such as those resulting from ethanol abuse, brain tumours, hydrocephalus, and chronic infections. When no specific cause of the dementia is identified, a presumptive diagnosis of AD is made (9,75). Consequently, the clinical diagnosis of AD is based on the exclusion of other potential etiologies.

This two-step diagnostic process has been formalized through the development of diagnostic guidelines. The most commonly used guidelines in AD research include those of the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) for primary degenerative dementia (77), the criteria of its revised version (DSM-III-R) for primary degenerative dementia of the Alzheimer type (75), and those for possible, probable, and definite AD developed in 1984 by a work group set up by the National Institute of Neurological and Communicative Disorders and Stroke, and the Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) (9).

Although the three sets of criteria are generally compatible, the NINCDS-ADRDA scheme was specifically established for research with an aim to improve the uniformity of diagnoses in different studies in order to allow for meaningful comparison of their results. Possible AD is diagnosed when patients with a dementia syndrome also have an atypical onset, presentation, or clinical course, or have concomitant disease sufficient to produce dementia but considered not to be the cause of it. A diagnosis of probable AD is made when the presence of a progressive dementing disorder of insidious onset is established and other potential causes are excluded. Definite AD requires histopathologic confirmation in the presence of a typical clinical picture (9).

The diagnostic accuracy of the NINCDS-ADRDA criteria has been assessed in a number of studies. Reported estimates of interrater reliability are generally comparable and moderate, with kappa statistic values ranging from 0.36 to 0.65 (78-80). Three studies have examined the sensitivity of these criteria: one reported an estimate of 92% (81); another observed values of 85% and 95% for each of two observers (82); and a third study's findings ranged from 81% to 85% depending on the neuropathologic criteria used (83). The corresponding specificities for the three studies were 65%; 13% and 33%; and 80% to 91%.

The validity of these criteria has mainly been investigated in terms of the positive predictive value of a clinical diagnosis of probable AD with neuropathology at autopsy serving as the gold standard. Estimates of this measure of validity range from 64% to 100% (81,83-87). These results indicate that as many as one third of patients with a clinical diagnosis of AD may not, in fact, have the disease. This has important

implications for AD research where the inclusion of nondemented subjects or demented subjects with other etiologies may obscure associations of interest.

2.5 Treatment

While research into potential treatments for AD has yielded some encouraging results, no agent has shown evidence of substantial efficacy to date (88,89). Challenges to the development of pharmacologic therapies for AD include uncertainty of diagnosis, difficulty of early diagnosis, accurate measurement of cognitive and noncognitive function, poor understanding of etiology and pathogenesis, variability in disease severity and progression, and uncertainty as to what constitutes clinically significant improvement (90-92). Fortunately, recent breakthroughs in molecular and cellular research are providing clues to the pathologic mechanisms in AD as well as new potential targets for treatment research (93).

Treatment strategies in AD can be categorized into four broad conceptual approaches: treating behavioral symptoms, treating primary or cognitive symptoms, attempting to slow or stop disease progression, and preventing or delaying the onset of illness (89). There appears to be some consensus that many behavioral disorders associated with AD are amenable to pharmacologic intervention, in spite of the limitations of research in this area (4,88,94). Commonly used drugs include antipsychotics, sedative-hypnotics, antidepressants and anticonvulsants (4). In addition, more effort is now being directed towards developing non-pharmacologic, behavioral interventions to manage these noncognitive problems (93).

Treatment research has focused primarily on improving cognitive function by restoring the neurotransmitter imbalances that result from neuronal death (90). The neurotransmitter deficit observed most consistently has been that of acetylcholine, an observation which has contributed to the cholinergic hypothesis of cognitive loss in AD (95). The most widely researched strategy for increasing brain levels of acetylcholine has involved using cholinesterase inhibitors such as tacrine (THA or tetrahydroaminoacridine) (89,96). This agent has demonstrated modest efficacy in improving cognitive symptoms in some patients (97,98) and was approved for use in AD by the U.S. Food and Drug

Administration in 1993. Whether tacrine leads to significant clinical improvement remains controversial (92).

Neurotransmitter replacement represents a therapeutic approach which is short-term and palliative. The longer-term goal of altering the disease course will require slowing or stopping the neurodegeneration in AD (90). Research aimed at modifying the disease process has been limited by poor understanding of its pathogenesis. Recent advances in this area, however, are generating optimism about the prospects of developing effective therapies that will intervene in the underlying pathology of the disease (21,93). There is also speculation that multiple treatment approaches may prove to be therapeutically optimal (89,99). Further understanding of the etiology and pathogenesis of AD may one day permit the early identification of those at risk and create the possibility of preventing or delaying disease onset (99).

In view of the limited success of pharmacotherapy, clinical management of AD has focused on supportive care which aims to keep patients socialized, and mentally and physically active in a safe environment (71,100). Provision of support services and personal counselling to caregivers may also be required to prevent or treat the physical and psychological problems often associated with the burden of caregiving (4,94,100).

2.6 Clinical course

Alzheimer's disease is a progressive, dementing disorder of insidious onset (9,75). Because initial changes are often subtle or imperceptible, time of onset is generally difficult to determine. As the disease evolves, diverse signs and symptoms emerge, some of which persist or progress as the disease advances while others regress (101).

The clinical manifestations of AD can be classified into five domains: cognitive, functional, behavioral, psychiatric, and motor. Cognitive symptoms are generally the first to emerge (102) and are considered cardinal since their appearance and progressive deterioration occur invariably in AD (103). The primacy of cognitive deficits is also supported by their association with the disease's neuropathology (73,84,104). While memory loss is the most prominent feature, language, orientation, attention, visual perception, and praxis (the ability to carry out purposeful movements) are also affected

(9,105). The functional domain consists of disability related to instrumental and basic (self-care) activities of daily living. The former includes difficulties with managing finances, travelling independently, and using a telephone, while problems with dressing, feeding, and toileting typify the latter.

Behavioral and psychiatric problems are common in AD, affecting at least 50% of patients at some point in their disease course (106-108). Delusions, depression, apathy, agitation, wandering, aggression, and sleep disturbance are among the most frequently observed. Motor signs include myoclonus, extrapyramidal signs, primitive reflexes, and seizures. Though many are prevalent, signs and symptoms belonging to these domains are considered associated disease features since they are not invariably manifested (9,103). Some of these features have been shown to be associated with proportionately greater neuropathology in specific brain areas (109,110). For many symptoms, however, it is unclear whether their manifestation is secondary to cognitive impairment or other symptoms, or the result of separate neural mechanisms or substrates (111,112).

Prevalence estimates of many associated features vary widely, due in part to differences between studies with respect to the source of subjects, definition of symptoms, levels of disease severity sampled, time frame of symptom assessment, and use of medical charts, caregiver reports, or patient interviews as data sources. For example, major depressive disorder was found in 86% of patients in one study (113) but was not observed in others (106,114). Estimates for anxiety and extrapyramidal signs are similarly varied, ranging from 12% to 76% (107,115) and 6% to 44% (116,117), respectively.

The typical sequential progression of these diverse clinical features is often described in terms of broad disease stages or phases. Global staging instruments further define distinct stages in the progression from no impairment to severe dementia, with each stage characterized by the presence and severity of specific signs and symptoms (118-121). Such scales have been widely used by clinicians and researchers to assess the global disease severity of a patient or patient group. Despite the usefulness of staging, there is little empirical evidence to support the presence of naturally occurring phases in the course of AD (122).

Alzheimer's disease has traditionally been described as having three broad phases (94,123,124). In the early phase, symptoms include memory loss, language problems, and

personality and behavioral changes. Patients initially lose memory for recent or everyday events. They may repeat themselves frequently, forget recent conversations or new acquaintances, or misplace objects. Judgement and abstract thinking are also impaired. Language symptoms generally appear as deficits in word and name finding (125,126). Difficulty learning and retaining new information, poor concentration, and difficulty performing novel or complex tasks may lead to decreased performance in work or social settings. Patients may withdraw from challenging situations, though social skills are often preserved.

Although patients may initially be aware of their cognitive deficits, they often lose insight as the disease progresses. They may neglect their personal appearance and hygiene, and experience difficulty managing their finances. Spontaneity and initiative are frequently diminished. Other behavioral problems that may emerge include withdrawal, apathy, irritability, paranoia, and hostility. Many are afflicted by depression and anxiety (127).

In the middle phase, cognitive functions decline further with continued deterioration of recent memory and the emergence of remote memory deficits. Language skills also diminish with the development of dysphasia. Patients commonly suffer from visuospatial deficits causing them to have difficulty dressing and to become lost, even in familiar surroundings. They also become disoriented to time. Other cognitive impairments include failure to recognize objects or familiar faces (visual agnosia) and difficulty executing purposeful movements (apraxia). Patients can no longer function independently, and require assistance with travelling and personal hygiene. Behavioral and psychiatric symptoms, such as delusions, agitation, aggression, and wandering are common and pose significant problems to caregivers.

In the late stage, patients are generally unaware of their environment and all recent life experiences. They may exhibit neurologic signs such as rigidity, postural abnormalities, myoclonus, and seizures. Profoundly demented, they lose both verbal and psychomotor abilities, eventually becoming mute, immobile, and incontinent. Death often results from complications of AD including pneumonia and other infections (128,129). Heart disease, stroke, and neoplasms are more common causes of death among patients who die early in the disease course (130).

The median survival following the onset of AD is estimated to be between five and nine years, with a range of one to 18 years (43,128,131). There is evidence to suggest that survival has improved during recent decades (132). Still, the life expectancy of AD patients is reduced relative to that of the age- and sex-matched general population (129,133,134). Five-year survival after diagnosis is approximately 50% of expected (128,135).

There is wide individual variability among patients with AD, not only in terms of disease duration and the emergence of associated disease features, but also in terms of when symptoms appear in the disease course. Although memory loss is the most common presenting symptom, approximately 10% of patients experience an atypical onset characterized by more focal neurologic deficits such as language problems, difficulty performing tasks, disorientation, or personality disturbance (71,136). The appearance of associated features relative to the course of AD is also quite variable. Studies examining the relation of motor signs (117,137-140), behavioral and psychiatric symptoms (106,113,114,127,141-146), and global behavioral measures (147-149) with level of cognitive impairment or disease stage have yielded mixed results. Those reporting significant associations generally found that associated symptoms, with the notable exception of depression, are more likely to occur with increasing disease severity (141,142). Still, predicting when these symptoms might appear is difficult since many features are quite prevalent even in mild disease.

Rate of disease progression, measured as decline in cognition, function, or disease stage, is similarly variable with some individuals deteriorating rapidly while others experience little or no decline for years (6,150-153). The heterogeneity among patients with AD in symptomatology, age at onset, and rate of progression has led to speculation and debate regarding the possibility of subtypes of AD (154-157). As yet, the existence of clinical subtypes with different etiologies or pathological mechanisms has not been demonstrated. The variable progression of AD, and the lack of subtypes to account for the diversity, has hindered clinicians' efforts to provide patients and their families with accurate prognoses. This, in turn, has prompted the search for factors that can reliably predict the future disease course of individuals with AD.

CHAPTER 3 - CRITICAL REVIEW OF PROGNOSTIC LITERATURE

3.1 Introduction

In this chapter, a critical review of prognostic studies of Alzheimer's disease (AD) meeting specified eligibility criteria is presented. The review focuses on studies investigating the ability of clinical and demographic factors to predict deterioration on cognitive and/or functional axes of the disease. The variability of mean progression rates across studies and of individual progression rates within studies is reviewed. The findings for individual prognostic factors are summarized and the consistency of the results is discussed. Particular emphasis is placed on the diversity of methods across studies and the potential impact of individual methods on the validity of the results.

3.2 Methods used to select and review studies

3.2.1 Identification and selection of studies

Prognostic studies of AD were identified through a MEDLINE search (1984-1995) using the following key words and medical subject headings: Alzheimer's disease, dementia, natural history, progression, prognosis, decline, clinical course, deterioration, disease course, longitudinal, and cohort studies. The search was restricted to articles published in English or French. The reference lists of relevant articles identified in this manner were perused for additional papers.

Studies of predictors of disease progression in AD satisfying the following criteria were included for review:

1 - original research study;

2 - prospective design;

3 - diagnosis of AD based on NINCDS-ADRDA criteria (9) for definite, probable or possible AD, DSM-III-R criteria for primary degenerative dementia of the Alzheimer type (75), DSM-III criteria for primary degenerative dementia (77), or their equivalent, either at intake or retrospectively;

4 - disease progression assessed using measures of cognition, function (activities of daily living), or disease stage; and

5 - clinical and demographic characteristics assessed as potential prognostic factors.

The prognostic value of electrophysiologic or radiographic tests was not examined in this review, nor was the literature related to the use of death or institutionalization as prognostic endpoints.

Only longitudinal studies that prospectively monitored the disease status of patients were eligible. Retrospective studies which estimated the disease progression of individual subjects or groups of subjects prior to entry into the study were excluded. The validity of such estimates is questionable given that they were based on retrospective estimates of subjects' test scores or on assumptions regarding their premorbid level of performance. Cross-sectional studies which estimated decline by comparing the performance of patients who were at different stages of the disease were also excluded. Empirical evidence suggests that this approach tends to underestimate the rate of individual disease progression (158).

The inclusion criterion 3 was applied in an effort to restrict the critical review to studies where subjects were most likely to have AD. The inclusion of patients with other causes of dementia, or other psychiatric or medical disorders, might spuriously increase the apparent heterogeneity of the disease course and bias results regarding the predictive value of potential prognostic factors. A clinical diagnosis of AD based on the guidelines specified above is highly predictive of a confirmation of AD at autopsy (81,86). The application of these guidelines therefore limits, as much as is currently possible, the inclusion of patients with other disorders. Studies reporting that their diagnostic guidelines were consistent with one of those listed above were considered eligible even when such assertions were only made in subsequent reports based on the same cohort.

3.2.2 Substantive review of studies

Key information regarding the results and methods of each of the eligible reports was abstracted using the form shown in Appendix 1. Each study was reviewed for its findings related to the sample's mean rate of progression over follow-up, the variability among subjects in their progression rates, and the ability of each factor examined to

predict decline on measures of cognition, function (activities of daily living), or disease stage.

There is evidence to suggest that disease progression along cognitive and functional axes in AD may be parallel but distinct (7). The evidence is based, in part, on the observation that correlations among cognitive and functional progression rates are on the order of 0.40 to 0.60 (7,159,160). The considerable amount of unshared variance suggests that cognitive and functional abilities may deteriorate at different rates and may therefore differ in their associations with potential prognostic factors. Accordingly, the prognostic findings for the three types of outcome measures were summarized separately since the failure to distinguish between them might, in itself, lead to the conclusion that the findings for individual factors are inconsistent.

This review of prognostic findings focused on the following: 1) the consistency of the findings for each factor both across and within different types of outcome measures; 2) issues related to the measurement or definition of individual factors; and 3) the discrepancy of the findings within individual studies and among different reports based on the same longitudinal cohort.

Several reports included in the review were based on different subsets of patients from the same longitudinal cohorts. The sample sizes, study methods, and/or prognostic factors varied sufficiently such that no two reports appeared completely overlapping. In several instances, however, the association of a given prognostic factor with a particular axis of disease progression was described by two or more reports generated from the same longitudinal study. The independence of such findings is questionable given the potential for considerable overlap of the subjects upon which they are based. The extent of overlap of the study samples was substantial for a few of the reports involved but could not be ascertained for most. The findings for each prognostic factor are therefore presented such that results derived from the same longitudinal cohorts are identified.

This review did not attempt to reconcile discrepant results among all studies examining a given factor by associating them with differences in study methods. Discrepant findings among reports based on the same longitudinal cohort were of interest, however, since these were likely related to study features other than the source or

selection of subjects. The consistency of results associated with the use of different methods (e.g., different measurement scales, estimates of progression, or statistical analyses) within the same report was similarly of interest. In these reports, conflicting results were often directly related to specific differences in study methods.

3.2.3 Methodological review of studies

Key issues related to the diversity and validity of the methods used in the studies were examined by assessing each report with respect to nine features of design and analysis. The evaluation of some of these features was guided, in part, by six criteria described by the Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre (subsequently referred to as the McMaster criteria) (161). These criteria were developed as guidelines for evaluating studies of clinical course and prognosis. They are listed below in relation to the study features to which they apply.

First, reports were classified into one of five categories according to the source of study subjects: (i) general population; (ii) volunteers (recruited through media announcements, primary physicians, or specialists); (iii) medical service (e.g., neurologic, psychiatric or geriatric hospitals services, psychiatric hospitals, referral practices); (iv) specialized clinic (memory disorders, dementia, or AD clinics); and (v) mixed (subject recruited from two or more sources among the preceding four and geriatric institutions). Studies were also examined with respect to whether the referral process by which patients entered the study was described (MCMASTER CRITERION).

Second, the selection criteria of each report were examined to determine whether subjects were assembled at an early and uniform point in their disease course (MCMASTER CRITERION). Identifying patients near the onset ("inception") of their disease is difficult in AD because time of onset is difficult to determine and many patients come to medical attention long after their first symptoms appeared. Choosing time of diagnosis or when patients first seek medical advice as the reference point (zero time) is also problematic since these are influenced by factors unrelated to the disease. One option is to restrict study entry to those rated as having mild disease according to a global staging

instrument or a narrow score range on a cognitive rating scale (162). The reports were therefore examined to determine whether this strategy was used and, if so, how mild disease was defined. Other selection criteria were also evaluated in terms of their potential impact on the representativeness of the sample.

Third, the assessment or definition of clinical and demographic potential predictors was examined with respect to issues of validity and consistency across studies. These issues are discussed in the review of the findings for individual prognostic factors. Issues related to the time frame of the assessment of clinical predictors were also considered. Studies were examined in terms of whether the assessment of clinical features corresponded to the time of study entry, some period prior to entry, and/or some portion of follow-up.

Fourth, studies were classified according to the type of measurement instrument that was used to assess disease progression: (i) cognitive rating scales; (ii) neuropsychologic tests; (iii) functional (activities of daily living) scales; and (iv) staging instruments. Studies that used more than one type of scale were included in each of the relevant categories. The reports were also assessed in terms of whether objective outcome criteria were used and whether outcome assessment was blind to prognostic status (McMASTER CRITERIA).

Fifth, each report was also examined with respect to the length of follow-up and attrition. Studies were classified according to whether the minimum duration of follow-up among subjects was less than one year, or one year or longer. The choice of a one year cut-off point was based on evidence suggesting that estimates of change, for some commonly used outcome scales, are less reliable when the length of follow-up is less than one year (160,163). The minimum length of follow-up was chosen because other measures of distribution (e.g., the percentage of subjects followed less than one year or the median length of follow-up) were not consistently reported. This classification served to identify studies where the duration of follow-up, of at least some subjects, was potentially inadequate. Reports were also categorized according to whether their loss to follow-up was 20% or less, greater than 20%, or not reported. The term loss to follow-up, in this thesis, refers to all attrition which precluded subjects' inclusion in the analyses.

The choice of a 20% cut-off is based on the McMaster group suggestion that a loss of more than 20% of the original cohort is unsatisfactory (MCMASTER CRITERION).

Sixth, studies were classified into three broad categories according to the approach used to estimate the disease progression of individual subjects: (i) computing the rate of change of test scores over follow-up; (ii) assessing the time to reach a given test score or clinical endpoint; and (iii) determining the presence or absence of decline over follow-up according to some pre-defined criterion for "change". Those using the rate of change approach were further subdivided according to the specific estimation method used. Studies that used more than one of the three broad approaches or methods of estimating rate of change were classified into each of the relevant categories. There were a few studies for whom this classification scheme was not applicable since the prognostic analyses performed did not require a separate estimation of disease progression.

Seventh, the classification of studies according to the statistical analyses used to identify significant predictors largely followed that for the estimation of disease progression. The analyses were categorized as follows: (i) rate of change or global change; (ii) survival; (iii) repeated measures; (iv) cross-sectional; and (v) analyses with no direct comparison of prognostic groups. Reports using more than one approach were classified into each of the relevant categories.

Eighth, studies were assessed according to whether they controlled for potential confounders (MCMASTER CRITERION). Baseline severity appears to be associated with disease progression and other potential predictors. Reports were therefore categorized according to whether an attempt was made to control for baseline severity through restriction, statistical adjustment, or matching. Studies in which control for severity was carried out for some prognostic factors but not others, and those for which it was felt that restriction had not been sufficient to preclude residual confounding were rated as having achieved partial control.

Lastly, the reports were categorized according to the number of study subjects included in the prognostic analyses. Evaluating the adequacy of the sample size of each report in terms of statistical power was complicated by the diversity of the follow-up schedules, measurement instruments, and prognostic analyses used among the studies.

Sample size calculations were therefore based on a "typical" study design that represented the intersection of the most commonly used follow-up scheme, measurement scale, and analysis. The features of this design consisted of two measurements separated by a one year follow-up, cognitive assessment using the Mini-Mental State Examination (MMSE) (164), and a comparison of the mean annual rate of change of two subject groups.

Calculations indicated that such a design would require about 40 subjects to detect a mean difference of three points/year between equal sized prognostic groups with 80% power and a 0.05 probability of type I error (165). The variance in the rate of change among individuals used in these calculations was 16. A difference of three points/year in the mean rate of change between groups was chosen as the *minimum*, clinically-meaningful difference to detect based on a review of the literature on the MMSE and on the clinical experience of a geriatrician familiar with the scale.

Because a number of studies examined prognostic factors whose prevalence was less than 20%, the calculation was repeated assuming a 15% prevalence. This yielded a required sample size of 80. Studies were therefore categorized according to whether the number of subjects analyzed was less than 40, 40 to 80, or greater than 80. The studies belonging to these categories were considered to have sample sizes which were possibly inadequate, likely adequate for more prevalent prognostic factors, and probably sufficient, respectively. The uncertainty of these appraisals reflects the fact that each report's statistical power was not computed directly, and that the calculation used did not account for multivariate analyses of multiple prognostic factors. This classification scheme provided, nonetheless, a general indication of the proportion of studies with potentially limited statistical power for detecting important prognostic factors.

3.3 Results and discussion of literature review

3.3.1 Overview of the studies

Fifty-nine prognostic studies met the eligibility criteria for review (section 3.2.1). A summary of the sample characteristics and outcome measures of these reports is shown in Appendix 2. Forty-six reports were based on studies conducted in the United States, five were from England, while Italy, Germany, Finland, Sweden, France, and Canada

accounted for one or two each. Approximately one third of the reports were based on multi-centre studies. These were more prevalent in recent years, however, reflecting a trend towards larger studies based on collaboration among several AD research centres.

3.3.2 Substantive findings of the studies reviewed

3.3.2.1 *Rate of disease progression*

In many of the studies, a subject's cognitive or functional decline was calculated as the average yearly change in test scores over the follow-up period. The sample means of these annual rates of change (ARC) are presented in Table 1 according to the cognitive or functional (activities of daily living) measurement instrument used. The reports using ARCs as estimates of disease progression are representative of all the reports reviewed in that they reflect a greater use of cognitive measures of decline. That most researchers chose to examine the deterioration of cognitive functions likely reflects the primacy of these symptoms in AD.

There is little consensus in the literature on what constitutes a "clinically significant" change in score for the scales shown in Table 1. There appear to be important differences, however, in the mean ARC values observed in different studies using the same scale. For example, the average rate of decline on the MMSE in one study (7) was 2.5 times faster than that of subjects in another study (166). Differences among the studies in the mean ARC value for a given scale may reflect variations in the study samples and/or methods.

In an earlier review of prognostic studies, Galasko et al. (5) observed that those reporting lower, intermediate, and higher mean ARCs on the MMSE had assembled subjects who were generally mildly, moderately, and more severely impaired at entry, respectively. This relationship has not been completely supported by subsequent research. For instance, two studies with subjects that were mildly to moderately impaired at entry have reported the highest mean ARCs for the MMSE (7,167). There is evidence, nonetheless, that ARC estimates are influenced by the severity of impairment at study entry. Consequently, differences among the studies in the distribution of subjects' baseline severity may explain, in part, the variability in mean ARCs.

Table 1

Annual rate of change on cognitive and functional scales in prognostic studies of Alzheimer's disease

Scale* (score range)	Study	*n*	Rate of change*		Comment
			Mean±SD	Range**	
Cognitive:					
MMSE (0→30)	Uhlmann et al. (1986)	156	2.6***		
	Reifler et al. (1986)	131	2.5***		
	Becker et al. (1988)	44	1.8***		
	Huff et al. (1990)	53	2.9***		
	Salmon et al. (1990)	55	2.8±4.3		
	Teri et al. (1990)	106	2.8±4.6		
	Boller et al. (1991)	33	3.2***		
	Burns et al. (1991)	85	3.5***	-3→15	
	Mortimer et al. (1992)	65	4.5***	0.2→15	
	Haxby et al. (1992)	16	4.4***		
	Morris et al. (1993)	430	3.9±3.7		
	Corey-Bloom et al. (1993)	244	3.1±3.9		
	Goldblum et al. (1994)	16	2.5***	1→5	
	Mielke et al. (1994)	25	4.2±3.6	-3.8→10.1	
	Kraemer et al. (1994)	81	3.1±2.3		
	Hogan et al. (1994)	135	3.6***		
mMMSE (0→57)	Stern et al. (1994)	236	6.7±6.0		
	Jacobs et al. (1994)	127	6.2***		
BIMC (0→37)	Lucca et al. (1993)	56	2.6±4.9		
mBIMC (0→33)	Katzman et al. (1988)	142	4.4±3.6		- 4 patient cohorts
	Thal et al. (1988)	40	4.5±3.2	0.4→12	
	Ortof & Crystal (1989)	54	4.1±3.0		
	Salmon et al. (1990)	55	3.2±3.0		
	Stern et al. (1992)	111	4.1±4.1		- four methods of calculating ARC
			3.9±4.0		
			5.2±5.4		
			4.0±3.6		
	Corey-Bloom et al. (1993)	190	3.1±5.0		
	Farrer et al. (1995)	186	4.2±3.1		
BOMC (0→28)	Morris et al. (1993)	430	3.8±4.3		
DRS (0→144)	Salmon et al. (1990)	55	11.4±11.1		
	Haxby et al. (1992)	16	19.1***		
CAMCOG (0→107)	Burns et al. (1991)	85	12.3***		

(table continues on next page)

Scale+ (score range)	Study	*n*	Rate of change* Mean±SD	Rate of change* Range++	Comment
ADAS					
(0→120)	Kramer-Ginsberg et al. (1988)	60	7.1±9.8		- cognitive & noncognitive scales
(0→70)	Stern et al. (1994)	111	9.6±8.2 8.4±6.0 11.1±9.9 11.2±9.9		- cognitive scale only - four methods of calculating ARC
Functional:					
BDS	Huff et al. (1990)	53	1.5**		
(0→28)	Lucca et al. (1993)	56	3.5±3.7		
	Stern et al. (1994)	236	2.4±2.6		
	Jacobs et al. (1994)	127	2.1**		
mBDS	Morris et al. (1993)	430	2.1±1.8		
(0→17)	Corey-Bloom et al. (1993)	70	1.8±6.7		
PSMS (6→30)	Green et al. (1993)	104	2.4±3.9	-9→18.6	
IADL (8→31)	Green et al. (1993)	104	2.1±3.3	-9.9→15.2	
PSMS + IADL (14→61)	Mortimer et al. (1992)	65	7.2**	-1.0→21	

* Expressed in points per year with positive numbers indicating deterioration.

** No SD reported for the full sample.

\+ MMSE = Mini-Mental State Examination; mMMSE = modified Mini-Mental State Examination; BIMC = Blessed Information-Memory-Concentration Test; mBIMC = modified Blessed Information-Memory-Concentration Test; BOMC = Blessed Orientation-Memory-Concentration Test; DRS = Dementia Rating Scale; ADAS = Alzheimer's Disease Assessment Scale; CAMCOG = cognitive portion of the CAMDEX; BDS = Blessed Dementia Scale; mBDS = modified Blessed Dementia Scale; PSMS = Physical Self-Maintenance Scale; IADL = Instrumental Activities of Daily Living Scale.

++ Minimum and maximum among individual subjects.

Katzman et al. (151), for example, compared the decline of four separate cohorts on the modified Blessed Information-Memory-Concentration Test (mBIMC) (168). The two cohorts whose subjects were more severely impaired at entry had lower mean rates of decline. Moreover, when all subjects were stratified according to their initial severity, the subset with greater than 24 errors (out of a maximum of 33) had a substantially lower mean ARC. The authors suggested that this most likely reflected a "ceiling effect"[1] of the scale. Once the more severe subjects were excluded from this study, the mean ARCs of the two cohorts who had initially been more severe increased to a level comparable to that of the other two cohorts.

The exclusion of subjects' data near the maximum impairment score also illustrates the potential influence of floor effects on ARC estimates. Stern et al. (169,170) applied several methods of estimating ARC to their sample in order to ensure that their conclusions would not depend on having chosen a particular approach (Table 1). Their comparison of mean ARCs generated from four approaches revealed marginally significant differences (169). The method which produced a mean ARC of 5.2 for the mBIMC was the same as that which yielded the 3.9 value except that the former did not use subjects' data after they scored over 30 errors. Other possible reasons for the variability among the mean ARC values across studies include differences in diagnostic error rates, length of follow-up and losses to follow-up, and small sample sizes that enhance the impact of sampling error (6,160).

The studies reviewed were also consistent in finding a wide individual variability in the rate of decline as indicated by the large standard deviations (SD) and ranges of the ARC values reported. In fact, the reported SD values were often equal to or larger than their corresponding estimates of the mean decline per year. In addition, the ranges of the

---

[1] Measurement scales are said to exhibit "floor" or "ceiling" effects when decline in function, below some level, is no longer detected. Thus, the measured rate of decline of individual subjects decreases as they approach the scale's maximum impairment score. Which term is used depends on whether the maximum impairment for a given scale is the lowest score (floor) or the highest (ceiling). The term floor effect will be used in the remainder of this thesis to refer to either situation.

ARC values indicated that some subjects even improved over follow-up. Studies using other approaches to estimating disease progression have also noted that some patients exhibited little or no decline for periods of three to four years (8,153).

This wide variability among subjects may be an artifact of sampling, poor measurement or other study limitations, or it may reflect true individual differences in the rate of disease progression. It is unlikely that the inadvertent inclusion of patients with disorders other than AD accounts for much of the observed variation since the ARC of subjects with subsequent autopsy confirmed AD is also highly variable (151,171). Another possibility is that ARC estimates derived from some scales are not reliable measures of change. There is evidence to suggest, however, that estimates of change for some instruments are fairly reliable when subjects are followed for more than one year (160,163).

A third possibility is that the rate of progression of individual patients may change over the disease course and that assembling subjects at different stages of their decline would inflate the apparent variability of the rates. This possibility is supported by observations that patients may experience a plateau phase early in their disease course (167,171) and that some scales appear to have a nonlinear pattern of decline (152,160,170). However, individual ARC estimates are still highly varied when periods of little or no change - initial plateau and floor phases - are excluded from the ARC estimation (167,171). This observation suggests that the wide variability of ARC estimates may reflect true interindividual differences in the rate of clinical progression. Whether such differences reflect natural variation in the disease process, differences among patients in their ability to resist or compensate for the neurodegeneration, or the existence of AD subtypes with different etiologies is unknown (8).

3.3.2.2 *Potential prognostic factors*

The prognostic findings of the eligible studies are presented in Table 2 for several of the more commonly investigated potential prognostic factors. At a glance, it is clear that the majority of the reports measured disease course in terms of the progressive loss of cognitive functions. Thirty-four of the 59 reports (58%) used only cognitive outcome

Table 2. Studies of clinical and demographic potential predictors of cognitive, functional, and disease stage progression in Alzheimer's disease

Potential predictors and their associations with decline	Study references by type of outcome measure[f]							
	Cognition (C)	n	Function (F)	n	C + F	n	Disease stage	n
Age at onset								
no association:	150,151,185,190,191,194*,195-197,207, 211, f-(169,170,251), d-192, g-174	16	6,8,188,195, d-192	5	--	0	185,216, i-(153,210)	4
younger at onset = ↑ decline:	6, d-193, g-167	3	d-193*	1	--	0		0
older at onset = ↑ decline:	--	0	--	0	172	1	--	0
Age at entry								
no association:	187,192,197,211,212,217	6	6,7,188,192	4	172	1	210,216	2
younger at entry = ↑ decline:	6,7,186*	3	--	0	--	0	--	0
Symptom duration								
no association:	6,150,186,187,190,191,193,194,212	9	6,8,193	3	172	1	216, i-(153,210)	3
shorter = ↑ decline:	196,197*	2	--	0	--	0	--	0
Gender								
no association:	7,150,186,187,190,191,197,207,211,212, d-(192,193), f-(169,170,251)	15	6,8, d-(192,193)	4	--	0	--	0
women = ↑ decline:	6	1	7	1	--	0	--	0
Education								
no association:	6,7,151,167,190,196,197,207,211,217	10	6-8,188	4	--	0	210	1
higher education = ↑ decline:	186	1	--	0	--	0	--	0
Family history of dementia								
no association:	150,167,190,191,193,194,196,207, f-(169,170,251)	11	8,193	2	--	0	--	0
+ve family history = ↑ decline:	197	1	--	0	--	0	--	0
no family history = ↑ decline:	--	0	204*	1	--	0	--	0
Extrapyramidal signs (EPS)								
no association:	7,116,188,191,197	5	7,8,188, e-208	4	--	0	--	0
EPS = ↑ decline:	173*,192,207*,208*, c-(175,190)	6	192, e-204*	2	--	0	180	1

(table continues on next page)

Table 2 (continued from previous page)

Potential predictors and their associations with decline	Study references by type of outcome measure[‡]							
	Cognition (C)	n	Function (F)	n	C + F	n	Disease stage	n
Language impairment								
no association:	197, a-166	2	7	1	--	0	--	0
more impaired = ↑ decline:	7,191, a-(188,211), c-(190,213)	6	183*,188,195	3	--	0	i-(182,183,210,214)	4
more impaired = ↓ decline:	212	1	--	0	--	0	--	0
Psychosis								
no association:	190, a-188	2	8,188, e-208	3	--	0	--	0
psychosis = ↑ decline:	181,192•,208*, a-(178•,215)	5	181,192•, e-204•	3	--	0	181	1
Delusions								
no association:	7,115,116,207,221	5	7	1	--	0	--	0
delusions = ↑ decline:	--	0	--	0	--	0	216	1
Hallucinations								
no association:	7,115,116,187	4	--	0	--	0	--	0
hallucinations = ↑ decline:	207•,222	2	7	1	--	0	216	1
Depression								
no association:	7,179,186,223, b-(177,187), c-(115,190)	8	7,8	2	--	0	--	0
Agitation								
agitation = ↑ decline:	115,186•,187,207••	4	--	0	--	0	--	0
Previous rate of change								
no association:	191,217, f-152	3	8,152	2	--	0	--	0
predicts subsequent decline:	167•, f-170	2	--	0	--	0	--	0

+ Reports in which the authors' conclusion, regarding the association indicated, was based on a non-significant trend.
* Report in which the association indicated was only found in mild or moderate dementia subgroups but not both.
• Reports in which the use of more than one measurement scale, measure of progression, or method of analysis led to mixed results.
§ Reports prefixed by the same letter in a given cell of the table are based on the same longitudinal cohort study.

measures while another 12 examined cognitive decline in conjunction with decline in activities of daily living (function) and/or disease stage. Functional decline and progression of disease stage were assessed, alone or in combination with other measures, in 15 and 11 reports, respectively. One report used the sum of two scales, one cognitive and one functional, as its measure of disease progression, and is listed separately in Table 2 (172). Reports also varied widely in the number and type of clinical and demographic factors investigated, with many examining a single factor (159,173-185) while others studied 15 or more (7,8,186-188). The findings for given prognostic factors based on different reports from the same longitudinal study are indicated where applicable (see footnote to Table 2).

Age at onset:

Age at onset was assessed as a potential prognostic factor or confounding variable in 25 reports, making it the most frequently investigated among clinical and demographic potential predictors. Early studies by Heston et al. (49) and Seltzer and Sherwin (189) found shorter survival times among patients who were younger than 55 and 65 years of age at onset, respectively. These findings suggested that early onset of illness might be associated with more rapid disease progression. Subsequent investigations of the effect of onset age on cognitive, functional, or disease stage progression, however, provided little support for this hypothesis.

A lack of significant association between age at onset and disease progression was observed in all or most reports for each type of outcome measure. Included among these were studies that adjusted for other covariates, such as baseline severity, and whose sample size likely provided adequate statistical power (169,190-192). Lucca et al. (6) reported an association between younger age at onset and more rapid cognitive, but not functional, decline. Haxby et al. (167) also observed a faster rate of cognitive decline among subjects who were younger at onset, unlike an earlier report based on the same longitudinal cohort which found no significant association (174). Once the two subjects with disease onset prior to age 50 and the fastest rates of decline were excluded, however, the correlation was negligible and no longer significant (167).

Different conclusions were also reached by another set of reports based on subsamples from a single longitudinal cohort (192,193). Both reports used the same analytic approach but the one which found a significant association was based on a smaller subsample of the cohort with longer follow-up (193). The latter report was characterized by discrepant results in that a multiple linear regression analysis demonstrated a significant influence of age at onset on functional decline while a repeated measures ANOVA did not (193).

The treatment of age at onset as both continuous and dichotomous within the same study has also led to different conclusions. Ortof and Crystal (194) found no significant difference in the mean progression rates of presenile (less than 65) and senile (65 or older) onset groups ($P<0.27$), and concluded that age at onset had no significant influence on cognitive decline. Other authors, however, point to this study's finding of a modest, but significant correlation ($r=-0.38$, $P<0.05$) as supportive of a faster progression among patients who are younger at onset (6). Only one report observed a faster rate of decline in senile onset subjects (172). The authors dismissed this finding, though statistically significant, because of the "considerable overlap among values for the two patient groups".

The time of onset of AD is difficult to determine given the gradual emergence of early symptoms. It is generally estimated from information provided by family members or other informants since patients may no longer be able to accurately recall their early symptoms when they first present for medical evaluation. Only two studies reported information on the reliability of their method of assessing age at onset: one noted a high agreement among different informants (195) while another found a high concordance between independent estimates obtained in the neurologic and psychiatric portions of the patient evaluation (188). Though informants and clinicians may agree, the validity of the estimates is uncertain. The ability of those affected to compensate for early deficits and their tendency to deny symptoms may limit an informant's ability to recognize, and later pin-point, the time of symptom onset. Thus, the potential exists for misclassification of age at onset. Its lack of predictive value in most studies may reflect a true lack of association or a nondifferential misclassification bias.

Age at study entry:

Three of the 13 studies investigating the predictive value of a subject's age at time of enrolment found that a younger age was significantly associated with faster cognitive decline (6,7,186). Two of the three also examined the influence of age at study entry on functional decline but found no association (6,7). Lucca et al. (6) noted that age at entry and age at onset were significant independent predictors but dismissed the former because of its high correlation with the latter and its "lower informative value". Mortimer et al. (7) speculated that the predictive value of age observed in their study may have reflected a selection bias whereby older subjects with faster disease progression were less likely to have participated. The third study reporting an influence of age at entry did so on the basis of a significant association between age and rate of cognitive decline as measured by the Dementia Rating Scale (DRS) (186). The association was not significant, however, when cognitive decline was assessed using the MMSE. The authors suggested that differences in sensitivity between the two scales may have contributed to the inconsistent findings.

Symptom duration:

Estimates of a subject's symptom duration at study entry are obtained from the same process described for age at onset. Their validity, therefore, are similarly uncertain. Most of the 16 studies examining symptom duration at study entry found that it was not predictive of future disease course. A tendency for subjects with a shorter history of symptoms to experience more rapid deterioration was observed in two studies of cognitive decline (196,197) but was statistically significant in only one of the two (196). Thal et al. (196) interpreted their finding as suggesting that individuals with slower progression presented later in the disease course.

Gender:

Several studies have reported an association between male gender and shorter survival in AD suggesting that men experience more rapid disease progression (132,198-201). Most of these, however, did not account for the greater life expectancy of women in the general population. Gender was not found to have predictive value in most of the 17 studies examining cognitive or functional decline. Moreover, the two

studies that reported significant associations found that women experienced a faster decline. It is noteworthy that one of these studies observed faster cognitive, but not functional, progression among women (6) while the reverse was noted in the other (7). Estimates of association in both of these studies were adjusted for other covariates including baseline severity.

Education:

An individual's level of education was found to have no significant value in predicting future disease course in all but one of the 14 studies investigating this factor. The study in question observed a faster cognitive decline among more highly educated subjects (186). The authors proposed that the ability of highly educated subjects to perform well on cognitive screening instruments allowed them to escape early detection, thereby delaying the onset of overt dysfunction to a later point in the disease process where decline was accelerated. This hypothesis is consistent, in part, with the observation that some cognitive rating scales are sensitive to level of education (202,203) and with the theory that increased educational attainment imparts a cognitive reserve that delays the onset of clinical symptoms in AD (32). The influence of education on cognitive test performance raises concerns regarding the validity of assessing this potential prognostic factor using these instruments to measure decline.

Family history of dementia:

Only two of 14 studies found a significant association between family history of dementia and disease course. Burns et al. (197) observed a faster cognitive decline in patients whose mother or father suffered from dementia. The second study, however, reported that the absence of a family history of dementia was predictive of a shorter time to moderate impairment in instrumental, but not basic, activities of daily living (204). The presence of a dementing illness among subjects' first degree relatives was generally ascertained from family members or other informants. This method of investigating familial aggregation is common and has been shown to be reliable across different informants (205). However, the classification of patients as having a positive or negative family history of dementia using this method is only a proxy for the presence of a genetic etiology.

Farrer et al. (159) examined the potential prognostic importance of a genetic etiology in AD using an algorithm that incorporated information from a genetic model and from the subject's family. A higher probability of having a major genetic AD locus was associated with faster functional decline only among male subjects. No significant association was found for cognitive decline in men or women. Carriers of the apolipoprotein ε4 allele, a gene implicated in the etiology of AD, experienced a similar level of cognitive decline to non-carriers (184). Taken together, these studies provide little evidence that genetic etiology predicts disease course in AD.

Extrapyramidal signs:

Bradykinesia, rigidity, gait disturbance, or other signs of parkinsonism are frequently observed in individuals with AD. They emerge as early as one year after estimated disease onset (206), affecting approximately 10% of those with mild disease (192,207). Extrapyramidal signs (EPS) are more prevalent in the later stages of AD (137,138) with an estimated 50% of patients affected by six years post-onset (206). Studies of the prognostic value of the presence of EPS at initial evaluation have yielded inconsistent results. Eight of the 14 reports observed faster decline in cognition, function, or disease stage among subjects with these signs while others found no significant association. Differences in study methods may, in part, account for this incongruity.

The multidimensional nature of the Blessed Dementia Scale was proposed as a possible reason for the lack of association with functional decline in a report by Stern et al. (208). These investigators repeated their analysis in the same sample using four factors, identified from a factor analysis of this scale, as independent outcome measures (204). The presence of EPS was significantly associated with a shorter time to moderate impairment only in the basic self-care factor, suggesting that these signs may predict decline in some functional abilities but not others. Soininen et al. (173) made a similar observation in that EPS predicted faster decline on neuropsychologic tests of some cognitive functions but not others.

Other possible reasons for the discrepant findings include differences between studies in whether patients with potentially drug-induced signs were excluded from the analyses and how the presence of EPS was defined. The manifestation of idiopathic EPS

in AD patients has been associated with Parkinson's disease pathology and disproportionate degeneration of the substantia nigra (110,180). The presence of these neuropathological changes, superimposed on those of AD, might accelerate the clinical progression of those individuals affected. Subjects with potentially drug-induced EPS, however, were not found to have these changes on autopsy (180). Consequently, the inclusion of these subjects in prognostic analyses might dilute the predictive value of these signs. Only five of the 14 reports mentioned that subjects whose EPS were potentially drug-induced were excluded from the relevant analyses (175,192,204,207,208). All five found the presence of EPS to be predictive of faster cognitive and/or functional deterioration.

The definition of EPS was not specified in some studies (116,190,191,197) and varied among the others. For instance, some reports defined the presence of EPS based on six or more parkinsonian symptoms (8,192,204) while others considered only one or two (173,207). These varied approaches to defining EPS may also have contributed to inconsistencies in the results.

The findings of one report are particularly noteworthy. Chui et al. (207) observed that the presence of EPS was significantly associated with faster cognitive decline in subjects with moderate dementia severity but not among those with mild dementia. They similarly found that the presence of hallucinations and agitation were predictive of faster cognitive decline only in the mild dementia subgroup. These findings suggest the potential for effect modification by disease severity where the presence of certain factors may be predictive of decline from some stages of the disease but not others. The assessment of potential prognostic factors in subgroups of patients defined by dementia severity has not been examined in other studies and warrants further investigation.

Language impairment:

Language is one of the main cognitive processes affected in AD (9). The loss of language begins in many individuals with decreased verbal fluency and word finding difficulty (126,209). As the disease progresses both receptive and expressive language abilities deteriorate, often reaching a level of impairment, in later stages, that precludes cognitive testing (125,209).

Eleven of the 14 reports found that greater language disturbance at study entry predicted faster cognitive, functional, or disease stage progression. The general concordance among these studies is remarkable in view of the diverse ways in which this prognostic factor was assessed. Nine reports evaluated language disturbance using a composite score of various verbal tests (7,166,182,183,210-212) or a general definition of aphasia (190,213). Six of these observed a significant association between greater impairment and more rapid disease progression (7,182,190,210,211,213). Other reports investigated specific aspects of language function (188,191,195,211,212,214). Among these, naming impairment proved to be the most consistent independent predictor of faster deterioration (188,191,211,214).

There were, however, some inconsistencies in studies of the prognostic importance of language. Mortimer et al. (7) observed that greater language impairment was significantly associated with more rapid cognitive, but not functional, decline. They found, instead, that faster functional deterioration was predicted by poorer performance on nonverbal cognitive tests. The authors proposed that these findings may reflect differential involvement of the two brain hemispheres in the clinical manifestations of AD. Other reports examining functional decline, however, found language impairment to be a significant predictor (188,195).

Three reports based on a single longitudinal cohort also reached different conclusions regarding the predictive value of language disturbance (166,188,211). The first report observed no association between greater lexical/semantic impairment and cognitive decline (166). The two subsequent reports found significant associations but used a different measure of cognitive decline and different definitions of language impairment (188,211).

Goldblum et al. (212) reported that severity of language impairment did not differentiate between fast and slow cognitive decliners. Instead, slow decliners were found to have a broader range of language deficits. The authors proposed that this unexpected finding may have been related to the high weighting of language-mediated tasks in the scale they used to measure cognitive decline. The validity of evaluating language impairment as a predictor of cognitive decline, given the dependence of

cognitive performance on language function, has also been questioned by other researchers (195). Although this issue has not been resolved, the prognostic importance of language impairment is supported by the significant findings of studies using functional or disease stage measures of decline.

Psychotic symptoms:

Approximately 20% to 40% of patients with mild AD experience hallucinations, delusions, or other psychotic symptoms (106,141,181). Studies of the association of these symptoms with disease severity are divided. Several have reported a higher prevalence in moderate or severe dementia (106,116,141,145,181,215) while others have found no association (7,115,127,181,207). Psychotic symptoms have been investigated both collectively and individually as potential predictors of disease course. Nine studies using the former approach varied in their definition of "psychotic symptoms". Seven defined them as the presence of hallucinations and/or delusions (8,178,181,192,204,215,216) while the remaining two provided no definition (188,190). Four of the seven also specified the presence of illusions (204), misidentifications (181,216), or paranoia (8) as indicative of psychosis. The diagnosis of these symptoms was generally based upon interviews with the patient and/or a family member or other informant.

The presence of psychosis, hallucinations, or delusions at study entry was predictive of faster disease progression in 10 out of 17 reports. Two of these reported mixed results, with psychosis predicting more rapid cognitive, but not functional, decline in one study (208) while the opposite was observed for hallucinations in another (7). A reanalysis of the former report, using four independent factors derived from the functional scale, revealed that the presence of psychosis predicted a significantly shorter time to moderate impairment in instrumental activities of daily living but not in basic self-care activities (204).

While the findings for psychosis and hallucinations as potential predictors were generally mixed, all but one of the studies investigating delusions reported no association. The presence of delusions in the latter, however, was not independently predictive of decline once adjusted for the presence of hallucinations (216). Huff et al. (188) pointed to their brief assessment of psychotic symptoms and the low prevalence (13%) of these

symptoms in their sample as possible reasons for their negative findings. The latter speaks to a problem of limited statistical power.

Inconsistent findings within studies also suggest an influence of study methods on results. Lopez et al. (178) found the presence of psychotic symptoms to be predictive of significantly faster cognitive decline on the MMSE but not on a neuropsychologic test battery. In another study, psychotic symptoms were significantly associated with a faster rate of cognitive decline but not a shorter time to moderate cognitive impairment (192). Furthermore, the association between psychosis and these two definitions of progression were reversed for the functional outcome. These findings suggest that conclusions regarding a factor's prognostic value may differ according to the outcome measure chosen or the method of estimating disease progression (rate-of-change versus time-to-endpoint).

Depression:

Depressive disorders must be excluded in making a clinical diagnosis of AD since depression itself can impair cognitive functioning (9). A diagnosis of AD is compatible, however, with the presence of depressive symptoms when these are judged not to be etiologically important. In fact, the prevalence of depression in AD patients has been reported to be as high as 45% (8,115,127,143). It differs from other behavioral and psychiatric manifestations associated with AD in that depression appears to be more prevalent in milder disease (113,127,141). The nine reports examining the prognostic value of the presence of depression at study entry were remarkably consistent. None found a significant association with cognitive or functional decline even though the definition of depression ranged from tearfulness and depressed mood (7) to syndromal major depression (179). Included among these were studies that adjusted for potential confounding by initial cognitive impairment (7,179,190).

Agitation:

Agitation is also common among AD patients with prevalence estimates of 30% to 40% in mild disease and 60% to 70% in severe AD (106,141). All four reports investigating the presence of agitation at study intake found it to be a significant predictor of faster cognitive decline. There were inconsistencies, however, in the findings within two of these studies (186,207). In one study, agitation was a significant predictor of

cognitive decline as measured by the MMSE, but not by the DRS (186). The other study found that agitated patients with mild dementia declined by six points on the MMSE significantly sooner than mildly impaired patients who were not agitated (207). No association was observed, however, when the time to an MMSE score of eight was used as the endpoint. These discrepancies further demonstrate the potential impact of the choice of outcome instrument and the definition of disease progression on conclusions regarding a factor's predictive ability.

Previous rate of decline:

Six reports examined the possibility that a subject's rate of decline in a given period might predict their subsequent progression. Four found little evidence to support the predictive value of an individual's previous rate of decline. Salmon et al. (217) calculated two progression rates for each subject, one corresponding to the first year of follow-up and one to the second. These two estimates of decline were not significantly correlated for any of the three cognitive scales used (r= 0.17 to 0.29). The authors speculated that the lack of association reflected the variable progression of AD within individuals, or poor reliability of the cognitive scales. The latter possibility is supported by the observation that estimates of change based on some scales are less reliable when the length of follow-up is one year or less (160,163).

Salmon et al.'s (217) use of the 12-month data point in calculating both change scores may have underestimated the true correlation because of regression to the mean. In order to avoid this problem, two studies calculated the first change score using the baseline and 12-month data points while the second was based on the 6-month and 18-month scores (152,170). While both studies reported significant correlations between the two change scores, Green et al. (152) dismissed theirs as "quite small" (r = 0.37). This approach is also problematic because the considerable overlap in the time windows upon which the change scores are based may induce a spurious positive association.

Two reports examined the ability of retrospective estimates of decline to predict progression over the study period (8,191). Neither found that these estimates were predictive of subsequent decline. The validity of the estimates is doubtful, however, since they were based on estimates of disease duration at initial evaluation and on the

assumption that all subjects had the same premorbid level of function at disease onset.

Another approach involved fitting an initial plateau phase followed by a decline phase to each subject's trajectory at the end of their follow-up (167). The rate of progression in the initial portion of the decline phase (nine to 15 months) was found to be significantly correlated with that of the later portion, for two of the three cognitive scales examined (r=0.66, 0.67, and 0.29). These results imply that early decline may be predictive of future progression once a subject's initial plateau phase is factored out. The transition from the plateau phase to the decline phase was not obvious for most subjects and it is unclear, from this approach, how one would prospectively determine when it occurs. Thus, while this finding is interesting from a theoretical perspective, its practical utility may be limited by the difficulty of establishing prospectively when a subject's decline phase begins.

These approaches to assessing the consistency of decline within subjects were varied and often limited by potentially unreliable or invalid estimates of change. It was also unclear how some approaches would translate into clinical practice. There was evidence, nonetheless, that rate of decline may itself have predictive value. This possibility merits further research.

Disease severity:

Twenty-four studies examined the association between disease severity at enrolment and cognitive, functional, or disease stage progression. Their findings are discussed according to the method of estimating disease progression (i.e., global progression on staging instruments, time to some pre-defined score or clinical endpoint, and rate of change of test scores) since these lend themselves to somewhat different interpretations.

Four reports examined the association of baseline performance on cognitive, functional and/or neuropsychologic tests with global decline in disease stage (153,210,214,218). All four were based on the same longitudinal cohort. Decline was defined as progression from the mild stage of the Clinical Dementia Rating scale (120,121,219) to moderate or severe stages over follow-up. One report found no significant association (218). The other three, however, observed that subjects who

progressed from mild to moderate or severe stages were more impaired on all or some of the severity measures at entry than those who remained at the mild stage (153,210,214). These findings suggest that even within a single disease stage some subjects may be more impaired than others and therefore more likely to progress to the next level.

Drachman et al. (8) postulated that the time to reach some endpoint for individual subjects may be influenced by their disease severity at study entry (i.e, how far their dementia has already progressed) and their rate of decline. Their analyses revealed that measures of cognitive impairment at entry predicted the time to dependence in basic activities of daily living (ADL) and time to incontinence. Similarly, Galasko et al. (220) found that subjects who reached milestones in basic and instrumental ADLs over two years of follow-up had greater cognitive impairment at entry than those who did not. Greater initial cognitive impairment was also found to predict a shorter time to cognitive endpoints (190,207). One study, however, found that poorer cognitive function was a significant predictor of the time to reach an MMSE score of eight but not the time to decline by six points on MMSE (207). Overall, the predictive ability of initial cognitive impairment was consistently demonstrated in the few studies using time-to-endpoint measures of disease progression.

Sixteen studies examined the association between subjects' initial scores and their subsequent rate of decline on cognitive and/or functional scales (Table 3). The findings from these studies may be interpreted in terms of the pattern of decline on individual scales: a lack of association suggests a constant decline throughout the scale while the presence of a significant relationship implies nonlinear decline over some portion of the scale. For most scales, the association between initial performance and rate of decline has only been examined once. Thus, little can be said of the consistency of the findings across studies for these scales. The association for the MMSE and mBIMC has been assessed in six and four studies respectively. The results for these two scales are inconsistent.

Overall, most studies found that initial performance on a scale predicted subsequent rate of progression. The nature of the reported associations are inconsistent however. Several studies observed that progression was more rapid among those less

Studies of the association between baseline severity and rate of progression on cognitive and functional outcome scales

Scale[§]	Study references by type of association reported			
	No association	Faster decline in less impaired	Faster decline in more impaired	Faster decline in moderately impaired
Cognitive:				
MMSE	217	7, 212[+]	191, 211[+]	197
mMMSE		193		
BIMC	6			
mBIMC	217, 169	151[*], 196		
DRS			217	
CAMCOG				197
ADAS		251		
ADAS (C)				170
Functional:				
BDS	6, 193	188		
PSMS				152
IADL		152		
PSMS+IADL	7			
Cognitive + Functional:				
BDS+BIMC	172			

+ Reports in which the association indicated was based on a marginally significant trend.
* One of the four cohorts included in this report was the study sample from Thal et al. (196).
§ ADAS = Alzheimer Disease Assessment Scale (cognitive and noncognitive subscales); ADAS (C) = ADAS (cognitive subscale only); BDS = Blessed Dementia Scale; BIMC = Blessed Information Memory Concentration scale; mBIMC = modified Blessed Information Memory Concentration scale; CAMCOG = cognitive portion of the CAMDEX; DRS = Mattis Dementia Rating Scale; IADL = Instrumental Activities of Daily Living scale; MMSE = Mini-Mental State Examination; mMMSE = modified Mini-Mental State Examination; PSMS = Physical Self-Maintenance Scale.

impaired at entry. This association was generally interpreted as indicating the presence of a floor effect in the scale. The finding was dismissed in one report, however, as a regression to the mean effect between the initial score and the difference between scores at entry and at one year (6). Fewer studies found that greater impairment at entry predicted a faster rate of decline. Furthermore, this relationship was observed for only one of three cognitive scales used in one study (217). The authors interpreted this finding to mean that the DRS had greater sensitivity to change in more severely demented patients.

Most studies treated initial severity as a continuous or dichotomous variable in their analyses. Burns et al. (197), however, categorized subjects according to mild, moderate, or severe impairment at entry, and found that the moderately demented group declined faster than the other two. The same trend was observed in two of three studies using a different approach (152,169,170). These studies analyzed the relationship of baseline severity with rate of change using linear, quadratic, and cubic polynomial regressions. Two reported a significant quadratic component where subjects who were moderately impaired initially experienced faster decline (152,170). The third study found no significant association (169). Another two studies using this same analytic approach also reported significant nonlinear associations (160,186). The interpretation of their findings is unclear, however, since the association examined was that of rate of decline with the *mean* severity over follow-up. Moreover, the use of a measure of severity based on the entire follow-up is not relevant to the goal of predicting future disease progression.

Taken together, these observations suggest that the pattern of decline on interval scales used to monitor progression in AD may be nonlinear over some portion of the scale. Whether observed nonlinearity reflects the true pattern of change in the disease process or is simply an artifact of the scale's composition (poor sensitivity to change) is unknown. Conclusions regarding the existence and nature of nonlinear trends for individual scales must await further research however.

Many factors may have contributed to the overall inconsistency of reported associations between initial severity and rate of decline including differences in study methods. For instance, almost half of the studies restricted entry to subjects with mild

or moderate disease, and/or excluded data at or near the maximum impairment score. Those which excluded severe subjects were no longer examining the association over the full range of the scale and may consequently have missed a trend in the severe range. Study results may also have differed according to whether data at or near the maximum impairment score was used to estimate rate of change. Where floor effects exist, the use of such data may attenuate rate of change estimates and consequently influence the study findings. Thus, the association observed may depend on the particular composition of the measurement instrument used and the portion of the scale over which subjects are observed.

3.3.3 Methodological characteristics of the studies reviewed

3.3.3.1 Source of subjects

Methodological characteristics of the 59 reports reviewed are shown in Table 4. Only one of the studies used a population-based sample where AD subjects were identified and recruited through a community survey of the elderly population in a defined geographic area. The nine reports classified under "volunteers" were derived from a single longitudinal study that recruited subjects through media announcements and physicians. Forty-four reports (75%) were based on convenience samples recruited solely or largely from general and/or specialized medical facilities. The case ascertainment of two of these studies was unusually comprehensive: one identified all patients who had been in contact with the only two psychiatric hospitals serving a defined catchment area (197,221-224) and the other used a surveillance network of primary care physicians and other medical services (186). Five reports (8%) did not describe the source of their subjects.

Subjects identified from medical facilities, particularly from specialty clinics, may not be representative of individuals with AD in the general population (47,225). Clinic-based samples are often highly-selected and reasons for self-selection or referral by family physicians may be related to prognosis. Patients with more rapidly progressive courses may be more likely to be referred to a specialty clinic. Certain patient or disease characteristics may also influence referral. For instance, those who are younger at disease

Table 4

Methodological characteristics of 59 prognostic studies reviewed

Characteristic	Reference numbers	n (%)*
Source of subjects:		
General population	184	1 (2)
Volunteers	153,180-183,210,214,218,226	9 (15)
Medical service	160,176,177,187,194-197,204,208,216, 220-224	16 (27)
Specialized clinic	8,115,116,152,159,166,169,170,172,175, 178,179,188,190,191,211,213,215,217,251	20 (34)
Mixed	6,7,151,185,186,192,193,207	8 (14)
Not reported	150,167,173,174,212	5 (8)
Outcome measure:		
Cognitive rating scales	6,7,115,116,150,151,159,160,167,169,170, 172,175-179,181,184,186-188,190-197, 207,208,211-213,215,217,221-224,251	42
Neuropsychologic test batteries	160,166,173,174,178,179,185	7
Activities of daily living (functional) scales	6,7,152,159,160,172,181,183,188,192,193, 195,204,208,220	15
Staging instruments	153,180-183,185,210,214,216,218,226	11
Clinical endpoint	8	1
Loss to follow-up:		
20% or less	6,8,160,180,182,185,193,195,210,212,220	11 (18)
Greater than 20%	116,153,166,173,181,184,197,214,218, 221-224,226	14 (24)
None specified	7,115,150-152,159,167,169,170,172, 174-179,183,186-188,190-192,194,196, 204,207,208,211,213,215-217,251	34 (58)

Characteristic	Reference numbers	n (%)*
Estimates of progression:		
Rate of change		38
Difference score	6,177,197,211,212,215,216,221-224,251	
Adjusted difference score	6	
Two-point method	116,150,151,160,166,169,170,172,176, 184,188,191,196,217	
Linear model	7,115,159,169,170,174,175,186,192-194, 213	
Bilinear model	167	
Multiple interval method	152,169,170	
Time to some endpoint	8,190,192,195,204,207,208	7
Global measure of change	153,180-183,185,210,214,218,220,226	11
Not applicable	173,178,179,181,185,187,193,215	8
Prognostic analysis:		
Rate of change or global change	6,7,115,116,150-153,159,160,166,167, 169,170,172,174-177,180-186,188, 191-194,196,197,210-214,216-218, 220-224,251	47
Survival	8,190,192,195,204,207,208	7
Repeated measures	178,179,185,187,193	5
Cross-sectional	173,181,215	3
No direct comparison	173,215	2
Not reported	170,226	2
Control of confounding:		
Yes	6-8,116,159,169,173,176,178,179,187, 190-193,210,214	17 (29)
Partial	180-183,185,186,188,207,208,211,213, 226	12 (20)
No	115,150,151,166,167,172,174,175,177, 184,194-197,204,212,215-217,221-224, 251	24 (41)
Not applicable	152,153,160,218,220	5 (8)
Not ascertainable	170	1 (2)

Characteristic	Reference numbers	n (%)*
Sample size analyzed:		
Less than 40	150,153,167,173,174,178-184,211,212, 214,215,218,226	18 (31)
40 to 80	6-8,115,166,172,188,194,196,204,208, 210,213,216,217,221-224,251	20 (34)
More than 80	116,151,152,159,160,169,170,175-177, 185-187,190-193,195,197,207,220	21 (35)

* Percentage values are not provided for those features of design or analysis where some studies were classified into more than one category.

onset, who manifest behavioral or psychotic symptoms, or who have a strong family history of dementia may be preferentially referred for assessment at specialized clinics. In addition to limiting the generalizability of the results, the preferential referral of AD patients with certain characteristics may bias the assessment of potential prognostic factors. Such a selection bias would occur if referral was differential according to factors of prognostic interest and rapidity of disease progression.

Whether such selection factors were operating in the prognostic studies reviewed was difficult to assess since only six of the 44 reports using clinic-based samples provided some description of the referral process (116,187,191,195,196,216). Two of these mentioned that subjects were solely or mostly referred by other physicians (191,196). Referral by family members or community support groups, in addition to general physicians, was reported by another two studies (116,187). The presence of psychiatric disturbances (216) and cognitive or behavioral disturbances (195) were also specified as reasons for referral. The general lack of information on factors that may have influenced referral to the study centres precludes any insight into the potential for selection bias in these studies. The issue of the generalizability of clinic samples, however, was examined in one study. Hogan et al. (191) found that dementia subjects attending their clinic were younger, had milder dementia, were more likely to be community-residing, and were more liable to have AD than dementia cases identified in a contemporaneous population-based prevalence study. The authors therefore concluded that their sample was not representative of the dementia population as a whole.

3.3.3.2 Selection of subjects

Investigators used a wide variety of eligibility criteria in the studies reviewed, many of which were clearly intended to improve the likelihood of a correct diagnosis of AD. Even with current diagnostic guidelines, however, as many as 30% of subjects in clinical samples may not, in fact, have AD (83,86) (section 2.4). The contamination of study samples by individuals with other disorders may have biased results regarding the true predictive value of potential prognostic factors.

Studies varied in the extent to which specific diagnostic criteria were reported. Some referred only to the diagnostic guidelines used (e.g., 175,194,197,212,216) whereas others detailed numerous eligibility criteria (e.g., 167,185,192,208,226). It is not clear whether this reflects differences in reporting or in the operationalization of the diagnostic guidelines. Many of the specific criteria were designed to eliminate subjects with other known causes of dementia (e.g., a history or clinical evidence of stroke, depression, or alcoholism). Other criteria, however, were aimed at excluding individuals with any of a number of health problems that might affect brain function (e.g., renal disease, diabetes, cancer, multiple sclerosis). These exclusion criteria produce samples of "clean" AD cases (225). It is not clear, therefore, that prognostic findings based on such samples can be generalized to individuals with AD and coexisting medical illnesses.

Some of the eligibility criteria used in the studies were based on logistical considerations (e.g., English speaking, informant available). In particular, over one third of the reports only included subjects with a specified minimum number of data points and/or follow-up. This restriction reflects the need for two or more measurements taken over some period of time to estimate disease progression. The post-hoc application of these criteria, however, effectively eliminates from the report any patients who died, dropped-out or were lost before having had the requisite number of visits or follow-up. Attrition in longitudinal studies of AD is likely related to the rate of disease progression, and may well be influenced by the presence of other disease symptoms of prognostic interest. Assessing the potential for selection bias due to nonrandom attrition is precluded, however, in reports using these criteria because information on the number and characteristics of subjects excluded is generally lacking.

Mixed zero times:

Fourteen reports (24%) specified criteria that restricted entry to subjects with mild and/or moderate disease severity. The criteria that were used, however, may not have been restrictive enough to ensure that those included were at a uniform point in their disease course. Five reports required subjects to have a minimum MMSE score of 14 to 18 for inclusion (179,190,192,207,213). The disease severity of individuals with AD scoring at or above these cut-offs still varies considerably however. In fact, two of the

studies referred to their cut-offs as being indicative of "mild to moderately" severe dementia (190,207). The remaining nine reports (153,180-183,210,214,218,226) were all based on the same longitudinal cohort which only included subjects rated as being in a mild stage of dementia using the Clinical Dementia Rating (CDR=1) (120,121). The range of performance of CDR=1 subjects on clinical and psychometric measures is considerable however (210). It is unclear, therefore, to what extent patients with a CDR=1 rating are homogeneous in their level of disease severity.

These observations suggest that most reports, if not all, assembled subjects who were at various points in the disease course (mixed zero times). This fact is also reflected in the variable symptom duration of subjects at entry into the studies (Appendix 2). In addition, most reports examined the prognostic importance of the presence of clinical features at study entry. Whether a subject has certain symptoms of prognostic interest (e.g., language impairment) at enrolment will likely depend on how far he/she has progressed into the disease. Rate of deterioration may also vary according to subjects' disease severity at intake (section 3.3.2.2). One consequence of assembling subjects with mixed zero times, therefore, is the potential for baseline severity to confound the association between some prognostic factors and disease progression (section 3.3.3.8). Even in the absence of restriction, this bias can be controlled for through multivariable analyses that incorporate measures of subjects' disease severity at study entry.

Another issue raised by mixed zero times relates to the practical implications of finding that a symptom's presence at or near time of entry is predictive of future decline when *entry* was at different points in the disease course for different subjects. Does the finding apply to all patients irrespective of how far they have progressed into the disease? Should a patient's prognosis be modified if and when he/she develops the symptom, regardless of when it emerges? These questions highlight an implicit assumption in assembling subjects with mixed zero times, namely that a symptom's presence will be equally predictive at different points in the disease course. The validity of this assumption is uncertain.

The only study to report separate results for mild and moderately impaired subjects found that three clinical features were significantly associated with faster cognitive decline

for either mild or moderately demented subgroups but not both (207). Whether the estimates of association were meaningfully different in the two subgroups could not be assessed since these were only reported for those associations which were statistically significant. Still, this finding suggests that disease severity may also be an effect modifier of the prognostic importance of clinical symptoms. The notion that a symptom's presence may be predictive of decline from some stage of the disease but not others has some intuitive appeal. For instance, the early manifestation of language impairment may reflect a more aggressive disease process and/or greater host vulnerability, and might therefore signal a poor prognosis. Its emergence in more severe disease, however, may simply be a consequence of global neural degeneration and therefore not predictive of subsequent decline.

If a clinical feature's prognostic importance depends on when it emerges in the disease course, examining its presence at study entry among subjects at various levels of severity might obscure its true predictive value. A study's findings would therefore depend on its sample's distribution of zero times relative to the prognostically relevant period of the disease course. Differences in the distribution of zero times across studies might therefore explain, in part, the inconsistent findings regarding the predictive value of most clinical features. The potential for effect modification should be assessed in future studies by performing subgroup analyses or by modelling the interaction between the presence/absence of clinical features and disease severity, duration, or stage at entry. How one defines homogeneous points in the disease course is open to debate.

3.3.3.3 Assessment of potential prognostic factors

Several issues related to the assessment of individual prognostic factors were discussed in the review of study findings (section 3.3.2.2). A broader issue related to the evaluation of disease symptoms as potential predictors is whether their assessment in the studies captured their true prognostic importance. A clinical feature's predictive value may depend on when it emerges in the disease course, as discussed in the preceding section. Severity, frequency, or persistence may also be prognostically relevant dimensions of a symptom. Prognostic research to date, however, has rarely addressed

these possibilities. More generally, empirical evidence and conceptual models on the mechanisms by which specific symptoms might be predictive of disease course are lacking. The absence of a theoretical basis to guide the assessment of clinical features is reflected by the variety of approaches that have been used by researchers.

The operational definition of language impairment, psychosis, depression, and extrapyramidal symptoms was inconsistent across studies (section 3.3.2.2). The time frame corresponding to a symptom's assessment also varied widely. Most investigators evaluated the prognostic value of disease features present at entry to the study. It is unclear, however, whether the cross-sectional assessment of a symptom's presence at or near study entry is prognostically relevant. Such an evaluation does not take into account when the symptom first appeared in the disease course, its duration prior to entry, or whether it emerged earlier in the disease but has since regressed. The latter consideration is particularly important in assessing behavioral and psychiatric symptoms which may be episodic in nature.

Only four studies reported examining the presence of clinical symptoms over some window of time prior to study entry. The time windows considered, however, varied notably among the four. The presence of psychiatric symptoms was assessed for the week prior to enrolment (204), for the six months preceding enrolment (7), for the year preceding enrolment (221,222) and for any time since onset (221,222). In each of the latter two reports, an additional 5% of the sample was rated as having the symptom of interest when the time frame of assessment was extended to any time since onset. These observations suggest that the time frame of the assessment of psychiatric symptoms, at least among a few studies, differed to such an extent as to limit their comparability.

Seven reports examined the presence or severity of clinical features at entry and over some portion of follow-up. This approach accounted for the development of symptoms of prognostic interest during follow-up among subjects who did not have them at entry. In one study, three prognostic groups were defined according to whether subjects had a given symptom at entry, developed it over follow-up, or never developed it during follow-up (213). Another four compared one or both of the former two groups with the latter (180,181,215,216). A sixth study added the severity scores for behavioral

and psychiatric symptoms taken over the first year of follow-up (7). The remaining study treated clinical predictors as time-dependent covariates in survival analyses (192).

Disease progression in each of these studies was based on the full follow-up period. Consequently, the time frame of symptom assessment was the same as that of disease progression for all or several subjects in each of these studies. The findings of these studies should therefore be interpreted in terms of correlates of disease progression and not predictors. Though this approach may provide insight into the natural history of the disease, it does not address the goal of predicting future disease course.

3.3.3.4 Outcome measures

A variety of outcome measures were employed in the 59 prognostic studies reviewed (Table 4). Among the five classes of measures identified, cognitive rating scales were used most commonly, followed by activities of daily living (functional) scales, global staging instruments, and neuropsychologic test batteries. Only one study used clinical endpoints as its measure of the progression of AD (8). Twenty-six studies (44%) used more than one outcome measure: 15 of these used measures from different classes while 17 used multiple measures from the same class. The use of more than one cognitive or functional scale in some studies led to different conclusions regarding the predictive value of one or more factors, indicating that prognostic findings may vary according to the outcome scale chosen (173,178,186,204).

Cognitive rating scales are multi-item measurement instruments that assess several areas of cognitive function. Ten such scales were used in 42 of the prognostic studies reviewed. The Mini-Mental State Examination (MMSE) was chosen as an outcome measure in 28 reports, making it the most commonly used scale (164,227). A modified version of the Blessed Information-Memory-Concentration Test (mBIMC) (168), and the Cambridge Cognitive Examination (228), used in six and five reports respectively, were the next most frequent. The other seven scales served in only one to three reports each (Appendix 2).

Several of these rating scales were designed to screen for the presence of dementia and quantify its severity (229). Each evaluates cognitive functions typically impaired in

AD including memory, language, orientation, attention, and praxis. The number and type of abilities assessed, however, vary from scale to scale, with some scales focusing on memory and orientation only (203), while others tap a broader range of functions (228,230,231). With the exception of the Alzheimer Disease Assessment Scale (ADAS) (230,232), each of these scales is limited to the evaluation of cognitive abilities. The ADAS also includes a noncognitive section which measures mood, behaviour, and psychiatric symptoms. This scale's lack of specificity for cognitive functions may reduce the comparability of results based on it with those based on instruments that assess only cognition. An additional limitation of at least two of these scales, the MMSE (202) and the Short Portable Mental Status Questionnaire (203), is that education appears to influence test performance. The evaluation of education as a potential prognostic factor may be biased as a result of this influence.

In comparison to mental status examinations, neuropsychologic tests are generally more detailed assessments of specific cognitive functions. In order to assess the variety of intellectual impairments in AD, individual tests of different functions are typically combined into test batteries. The seven studies using this class of measures administered between four and 22 tests each. Of these, three used each of the tests as separate outcome measures (160,173,185) while the other four created one or more composite scores based on several tests each (166,174,178,179). The number and type of tests used, and the different ways in which their scores were combined varied substantially among these studies, making any comparison between them difficult. Examples of neuropsychologic tests used in these studies include the Boston Naming Test (233), and subtests of the Wechsler Memory Scale (234) and the Wechsler Adult Intelligence Scale (235).

The use of individual neuropsychologic tests as independent outcome measures allows one to examine whether potential predictors are differentially associated with specific cognitive functions. The clinical utility of predicting the decline of a single cognitive function may be limited, however, particularly if that ability is lost over a relatively short portion of the disease course. This limitation may also apply to the use of individual factors (identified through factor analyses of cognitive or functional scales) as separate outcome measures where these assess a narrow range of abilities.

Many activities of daily living (ADL) scales rely upon information provided by relatives or other informants regarding the subject's level of function. They are especially useful in the later stages of the disease when cognitive testing may no longer be possible. The Blessed Dementia Scale (BDS) (73,236) was used in nine of the 15 studies examining progressive impairment in ADL. This scale assesses personality changes in addition to daily functioning. Its multidimensional nature, however, reduces its specificity as a measure of functional status. Consequently, two reports used a modified version of the BDS which excludes the items evaluating personality (160,220), while another used four factors, derived from the BDS, as separate outcome measures (204). Other functional scales used in the prognostic studies were restricted to measuring basic and/or instrumental ADLs only (Appendix 2).

One of two staging instruments was used in 11 of the prognostic studies reviewed: the Clinical Dementia Rating (CDR) (120,121,219) or the Global Deterioration Scale (GDS) (118). They provide a global measure of the stage of dementia based on ratings of several cognitive and functional skills. Unlike cognitive or functional scales, these instruments have a limited number of gradations with only three (CDR) and five (GDS) possible ratings for subjects with clinically diagnosed AD. As a result, these instruments are likely unable to detect smaller but clinically important changes in function, and long follow-up periods may be required before subjects move from one stage to another. These instruments are therefore limited as measures of the progression of AD.

Many of the cognitive rating and activities of daily living scales have demonstrated good criterion validity in that they discriminate between demented and nondemented elderly (160,203,228,232,237,238) and correlate significantly with the neuropathology of AD (73,84,104,239). Furthermore, correlations among scales within the same symptom domain are typically moderately high to high, reflecting the considerable overlap in the constructs they measure (convergent validity) (e.g., 104,164,240-242). Test-retest reliability also tends to be quite high (e.g., 152,203,240,243,244), as is interrater reliability for the few scales for which it has been assessed (152,228,230,245). Though many are valid and reliable measures of the presence and severity of dementia, only a few of these scales have been evaluated for their ability to detect clinically important *changes* in the disease status of individuals over time.

Scales used to characterize disease progression should be reliable and sensitive measures of change. Estimates of change derived from the MMSE (160,163) and the modified Blessed Dementia Scale (163) have been found to be fairly reliable when change is measured over a period of greater than one year. The reliability of change scores for other cognitive and functional scales has not been assessed. The evaluation of sensitivity to change has been hindered by the lack of a gold standard measure of the progression of AD (other than change in neuropathology), against which estimates of change could be validated. Consequently, the various approaches that have been used to assess sensitivity to change are, to varying degrees, limited.

A common approach to assessing the sensitivity to change of cognitive and functional scales has been to compare the rate of change from different levels of baseline severity. When the average change among severe subjects is significantly less than that of mildly or moderately impaired individuals a scale is said to exhibit a floor effect. The presence of floor effects are generally interpreted as reflecting poor sensitivity to change of the measurement scale rather than a true lack of decline. Such a conclusion assumes that the amount of *true* change over a given period is constant throughout the scale and that more or less *observed* change reflects greater or lesser sensitivity to change. To the extent that this assumption is valid, several scales used in the prognostic studies may be limited by poor sensitivity to change in severe disease (see Baseline severity in section 3.3.2.2).

Only one study specified that the assessment of disease progression was blinded to patients' prognostic status (214). The extent to which a clinician's evaluation of disease severity at follow-up may be influenced by his/her knowledge of the patient's prognostic status depends on the degree of judgement involved in assigning test scores. Most of the cognitive scales described above generally assess a subject's ability to answer specific questions and to perform various mental and motor tasks. Consequently, these scales are fairly independent of clinical judgement. Most activities of daily living scales and staging instruments, however, are at least partly based on information provided by informants and may be more susceptible to observer bias.

3.3.3.5 Length and completion of follow-up

The duration of follow-up in prognostic studies of AD should be sufficiently long to allow for measurable change in disease severity. The proportion of subjects that reach a disease endpoint will also depend on how long they are followed. Moreover, the reliability of estimates of change has been found to vary according to length of follow-up. Research findings suggest that progression rates on some commonly used scales are less reliable when subjects are followed one year or less (160,163).

The minimum length of follow-up was less than one year in 13 of the 59 studies (22%). Unfortunately, the proportion of subjects with less than one year of follow-up was rarely reported. One study with 20% of the sample followed less than one year (187), and five studies with minimum follow-up periods of one year (7,160,186,192,193), used weighted regression analyses to assign less weight to subjects with a shorter follow-up or a larger standard error of the slope (rate of change estimate). This strategy does not reduce the potential for misclassification resulting from unreliable measures of change. It does, however, limit the influence of these estimates in the assessment of prognostic importance.

Attrition in longitudinal studies of AD is problematic given the progressive debilitating nature of the disease and the advanced age of those afflicted. Moreover, it is likely that attrition is selective such that those who progress more rapidly are more likely to succumb quickly, to be institutionalized, or to simply not return for re-assessment. Consequently, subjects with less severe disease courses may be over-represented in the analysis. Loss to follow-up may also be influenced by the presence of certain prognostic factors of interest (e.g., behavioral problems). Differential attrition according to both rate of decline and prognostic factor status creates the potential for selection bias.

Thirty-four studies (58%) made no mention of losses to follow-up (Table 4). Most of these, however, required a minimum number of data points and/or minimum duration of follow-up for inclusion in the report. Any subjects lost to follow-up before meeting these criteria would therefore have been excluded from the report. Since information on the number and characteristics of subjects excluded from these studies was not generally reported, the potential for selection bias could not be assessed.

Eleven studies reported losses of less than 20% while 14 reported attrition rates of greater than 20%. Many of these 25 studies did not describe the reasons for the losses to follow-up. Several, however, did examine the association between prognostic factors and death during follow-up. This strategy does not account for other types of losses and the statistical power to detect an association was often limited by the small number of deaths. Six studies found a significant association or marginally significant trend between at least one prognostic factor and death (173,181,184,197,222,224). Other studies examined differences in clinical and/or demographic predictors between all losses and those analyzed. Some of these, however, did not appear to have compared the subject groups on all of the prognostic factors investigated (116,195,246). Others that did found no significant associations (6,193,214). One study did report, however, that subjects who failed to return to the clinic for follow-up had a greater progression of disease stage over two years than those who did not (185).

Selective attrition is potentially a big problem in the prognostic studies reviewed but was difficult to assess in most since the number and characteristics of those lost were often not reported. Most studies can be faulted for failing to adequately report, describe, and analyze attrition.

3.3.3.6 Assessment of disease progression

Three general approaches were used to estimate the disease progression of individual subjects: 1) computing the annual rate of change (ARC) of test scores over follow-up; 2) assessing the time to reach some specified test score or clinical endpoint; and 3) determining the presence or absence of decline over follow-up according to some pre-defined criterion for "change" (Table 4). Of these, the annual rate of change estimation was the most commonly used approach. This method and the time-to-endpoint approach were typically applied to test scores from cognitive and functional scales, while the "global measure of change" approach was used in conjunction with staging instruments.

Six different methods were used to estimate a subject's annual rate of change (Table 4). Eleven studies calculated the difference between the first and last score over

a fixed follow-up of one or two years. Another study that used the simple difference score, however, appeared to have variable follow-up (177). Comparing the difference scores of subjects with varied lengths of follow-up is problematic since the amount by which a subject declines depends on the length of observation. This problem can be avoided by using the two-point rate, calculated as the difference between the first and last score divided by the interval in time between them. This approach was used in 14 studies. Although the two-point rate has the advantage of allowing for different lengths of follow-up, it is similar to the difference score in that neither makes use of any interim data points that may be available.

The psychometric literature has been critical of the use of simple difference scores as a measure of change because of the spurious influence of baseline scores on their corresponding difference scores, due to regression to the mean (247). Various strategies have been proposed to deal with this artifact including one by Fleiss (248) which statistically adjusts for the influence of baseline severity. The computation of the "adjusted difference score" consists of subtracting, from the difference score, the portion of change that can be predicted from the baseline score alone. Consequently, this approach removes the entire influence of a baseline score on its corresponding difference score. This would be justified only if one assumes that the entire association between the two is due to regression to the mean, an assumption that may be extreme. This strategy was used in a single prognostic study (6).

In 12 prognostic studies, ARC estimates were obtained through a least-squares regression of each subject's test scores on time of assessment. The slope of the resulting linear function is the rate of change estimate. Unlike the first three methods, the linear model makes use of all available outcome measurements. The regression, however, requires at least three data points per subject. This model and the first three methods described are based on the assumption that decline on the outcome instrument used is linear over the period of assessment.

One study estimated ARC assuming a bilinear model of decline in AD (167). The approach involved fitting an initial plateau phase followed by a decline phase to each subject's trajectory and required a minimum of four data points. The ARC was estimated

from the slope of the decline phase. The bilinear model produced a better fit than the linear model only for the subset of subjects with isolated memory problems at entry (i.e., believed to be in an earlier stage of their illness).

Lastly, three studies used the multiple interval method (MIM) which divides each subject's follow-up into consecutive one year intervals and calculates the difference score for each interval. This approach assumes that change within intervals is linear. Several ARC estimates may be produced for each subject thus time intervals, rather than subjects, are the unit of analysis. The treatment of these change scores as independent observations in conventional statistical analyses, however, violates the underlying assumption of independence. It should be noted that each of these six methods of calculating ARC is susceptible to floor effects on outcome scales.

The comparability of ARC estimates obtained with the two-point, linear regression, and MIM methods was assessed in two studies by applying all three methods to each subject's outcome data (169,170). One study compared the mean ARC values from each method and found no significant differences (169). Comparing only the means, however, has limited value since small differences between the means of ARCs produced by different methods may mask large individual differences. The second study performed separate prognostic analyses for ARC estimates from each method and found no important differences in the results (170).

Time-to-endpoint estimates were used to assess disease progression in seven studies. Six of these used time to a given test score as the endpoint: three were based on scores that investigators felt were indicative of moderate severity (192,204,208), two reflected severe impairment (195,207), and one used the first of two consecutive maximum impairment scores (190). One limitation of choosing scores at or near the maximum impairment score is that subjects are more likely to be lost before reaching it (fewer events). A second limitation is that estimates of the time taken to reach the specified score may be exaggerated if subjects' performance tends to plateau in the floor of the scale.

Criteria proposed for the choice of endpoints include that they: 1) are clinically meaningful, 2) are unambiguous, 3) reflect disease severity and not factors unrelated to

disease, 4) occur commonly in the disease course, and 5) are stable (unlikely to be reversed) (220). Studies using test scores as endpoints may satisfy criteria 2), 3), and 4), but 1) requires an arbitrary decision of what a clinically meaningful cut-off score is, and 5) will be influenced by fluctuations in patients' scores and poor reliability of the scores. One study looked at the percentage of subjects reversing various endpoints after reaching them and found that 18% reversed the severe stage of the Clinical Dementia Rating scale (CDR=3), 17% reversed the loss of five or more instrumental activities of daily living, 10% reversed the loss of toileting, but only 2% reversed an MMSE score below 10 (220). These findings suggest that some endpoints may be more stable than others.

Ten of the 11 reports using global definitions of disease progression used this approach in conjunction with global staging instruments (153,180-183,185,210,214,218, 226). Researchers in these studies monitored the progression of patients from one stage of the instrument to more severe stages. As discussed in the review of outcome measures (section 3.3.3.4), the use of global staging instruments is limited by the small number of gradations used to describe the disease course and the long follow-up periods that may be required before subjects move from one stage to another.

Studies classified as "not applicable" in Table 4 used statistical analyses to identify predictors that did not require a separate estimation of disease progression.

Assumption of linearity:

The different methods of estimating annual rate of change assume that an individual's decline in test performance is linear over the period of follow-up upon which the estimate is based. For most studies using the difference score, adjusted difference score, two-point rate or linear model, this period was the entire length of follow-up. The multiple interval method, however, assumes only that decline is linear within each measured interval. Various studies have assessed the validity of the linearity assumption by examining the pattern of change in scores over *time* within subjects or the relationship between the rate of change on a scale and level of *severity* across subjects.

A few studies have made observations which suggest that decline may not be linear over time. Four studies examined the correlation between two change scores calculated for each subject based on different periods of follow-up (152,167,170,217).

Two of the four reported only modest associations (152,217) (see Previous rate of decline in section 3.3.2.2). These findings were interpreted as suggesting that the rate of progression within subjects was not consistent over time. The remaining two, however, reported statistically significant associations (167,170). The approaches used in these studies were potentially limited by regression to the mean and/or poor reliability of the change scores.

The finding that rate of change differs according to baseline severity has also been interpreted as suggesting nonlinear decline on a scale (see Baseline severity in section 3.3.2.2). Three studies analyzed the relationship of baseline severity with estimates of change using linear, quadratic, and cubic polynomial regressions (152,169,170). One found no significant association between initial severity and rate of decline (169), but the remaining two reported significant linear (152) and quadratic components (152,170). The results of the latter two suggest the presence of nonlinear trends in decline. Two other studies using the same analytic approach also found significant nonlinear associations for some scales (160,186). The interpretation of the findings from the latter two is unclear, however, since the association examined was between rate of decline and *mean* severity over follow-up rather than *baseline* severity.

Two studies, using a different approach, examined nonlinear patterns of decline in subjects' test scores over follow-up through least-squares regression analysis (167,169). One study investigated, in a stepwise fashion, the successive improvement in fit to data provided by a linear model with a common slope for all patients, a linear model with separate slopes for each patient, a quadratic model with a common parameter for all patients, and a quadratic model with separate parameters for each subject (169). Though statistically significant, the quadratic trend was small and highly individual. The second study used the same first two models described above, but the third model tested was a bilinear model consisting of plateau phase followed by a decline phase fitted to each subject's data (167). The bilinear model produced a significantly better fit than the linear model but only for a subset of subjects with isolated memory problems at entry. While the analyses used in these studies required several measurements for each subject, the likely inter-correlation of repeated observations within subjects was not accounted for in the estimation of the common linear slope.

A final study reported examining linear and nonlinear trends in decline over time through "between-group trends testing" (193). The actual analysis used is not clearly reported but appears to have been a repeated measures ANOVA. The authors reported that the analysis confirmed the presence of a linear trend but tests of nonlinear, higher-order trends were not significant. Overall, the results of these studies have been inconsistent with respect to confirming the presence of linear and nonlinear trends in decline. Further research is needed to resolve this issue for individual scales given the implications of nonlinear decline for the validity of ARC estimates of disease progression.

3.3.3.7 Prognostic analyses

The statistical analyses used to identify significant predictors generally followed from the type of progression estimate calculated (Table 4). These consisted of bivariate or multivariable analyses of ARC estimates (e.g., *t*-test, multiple linear regression, analysis of variance), of subject groups categorized according to whether or not they progressed over follow-up (e.g., chi-square test, discriminant analysis), and of time-to-endpoint estimates (e.g., Kaplan-Meier product limit method, Cox proportional hazards model). Many of the studies in each of these three categories used bivariate analyses only.

Despite having followed subjects longitudinally, three studies performed cross-sectional comparisons at one or more visits after enrolment. One of the three compared test scores at one year post-entry, adjusting for baseline severity (173). The other two made no adjustment for possible differences (181,215) despite evidence in one that the disparity in baseline scores between the two prognostic groups was substantial (215). Unless initial scores are very similar among the prognostic groups being compared, this approach is flawed given that differences observed at some point in follow-up, or the lack thereof, may simply reflect differences at baseline rather than rate of progression.

Two of the studies using cross-sectional analyses also performed analyses which involved no direct comparison of prognostic groups. One compared change scores in the year before and after the onset of psychosis and found that decline after onset was significantly faster (215). The lack of comparison group in this approach raises the possibility that the trend observed may have reflected natural fluctuation in the disease

course. The second report examined decline on several neuropsychologic tests in the two prognostic groups separately (173). Consequently, it could not be determined whether observed differences in decline between the two groups were statistically significant.

The most sophisticated approaches used were those which analyze repeated measures directly. Two studies used repeated measures ANOVA (179,193) and one (187) used a growth curve approach to modelling change developed by Laird and Ware (249). Both analyses make use of all available follow-up data but the latter allows for different numbers of data points per subjects and variable intervals between assessments. Two studies included in this category reported using two-factor ANOVAs where the interaction of "time" and "groups" factors provided evidence for a between-group difference in the rate of decline (178,185). Neither of these reports specified, however, that these were repeated measures ANOVAs. The use of the classical ANOVA framework in this case is inappropriate since no account is taken of the dependency of repeated measurements within subjects.

Two studies used more than one type of analytic approach (192,193). One study found that age at onset was a significant predictor of functional decline according to a multiple linear regression analysis but not a repeated measures ANOVA (193). Similarly, the second report found that the presence of psychosis was significantly associated with a shorter time to the functional endpoint but not with the rate of functional decline (192). These associations were reversed for the cognitive scale. These findings suggest that conclusions regarding a factor's predictive ability may depend on the analytic approach used.

Fifteen or more statistical tests were performed to identify prognostic factors in 15 of the 59 studies reviewed (25%). The large number of tests typically resulted from the investigation of numerous prognostic factors, the use of multiple outcome measures, and/or the application of different types of analyses. Given that significance testing was used to identify predictors of deterioration, the inflation of the probability of type I error in these studies may have been serious. Only seven of the studies, however, made any mention of correcting the α level to reflect the multiple testing (8,170,173,180,192,197).

3.3.3.8 Control of confounding

There is evidence to suggest that disease severity at entry may be associated with estimates of disease progression (section 3.3.2.2) and with disease features of prognostic interest (sections 2.6 and 3.3.2.2). Patient characteristics considered unrelated to disease severity in the general population of AD patients may also be associated with severity at entry due to factors influencing when patients present for medical attention. For instance, some studies have observed that more highly educated subjects were less impaired at entry (8,250). This finding suggests that those with higher educational attainment may be more likely to seek medical attention sooner. These sampling artifacts are associated with the use of clinic samples and would not be expected in community-based samples.

These findings indicate that disease severity at entry is a potential confounder for the association of many, if not all, potential predictors and disease course. Thus, control for baseline severity is necessary in order for prognostic results to be valid. Only 17 reports (29%) were judged to have controlled for baseline severity (Table 4). All but two of these achieved control by including measures of baseline severity in multivariable analyses. The remaining two matched AD subjects with and without the prognostic factor of interest on baseline scores (178,179). Another 12 (20%) were considered to have partially controlled for baseline severity. Although many of these reports restricted the baseline severity of subjects at study entry or in the analyses, none were considered to have restricted severity enough to have fully controlled for potential confounding.

Almost half of the studies (41%) failed to control for initial severity and, consequently, the validity of their results is in question. Some of these compared prognostic groups with respect to baseline severity and, when no significant differences were found, concluded that confounding was not a problem. Moreover, in some cases there were statistically or clinically meaningful differences in the severity scores of the groups being compared (204,215,224,251). A lack of significant results was not considered proof of the absence of a confounding effect by baseline severity, particularly as the significance tests were often limited by inadequate statistical power due to small sample sizes. Studies classified as "not applicable" in Table 4 were those which only examined measures of disease severity as potential prognostic factors.

3.3.3.9 Sample size and power

The number of subjects analyzed in the 59 reports ranged from 16 to 430, with a median of 65. Approximately one third of the reports had sample sizes which exceeded 80. These reports likely had adequate statistical power to detect clinical or demographic features with important predictive abilities. Another third, with sample sizes between 40 and 80, are also likely to have had sufficient power but only for prognostic factors whose prevalence in the study sample was greater than about 15%. The adequacy of the sample sizes of the remaining third (less than 40 subjects) is doubtful.

Depending on study characteristics such as the number of measurements per subject and between-subject variation in rates of change, some reports may have had greater power, in truth, than their classification here would suggest, and others less. It is not obvious, however, that this misclassification would have been so differential to have meaningfully changed the distribution of the reports with respect to the adequacy of their sample sizes. Thus, the results suggest that as many as two thirds of the 59 reports may have had insufficient power to detect less frequent but important prognostic factors, while half of these were likely unable to discern even more prevalent predictors. This finding may explain, in part, the inconsistency in the results from different studies in as much as conclusions regarding the predictive value of factors was based on significance testing.

3.4 Summary of literature review

This review of prognostic studies of AD confirms earlier observations regarding the wide variability of disease progression among individuals with AD reported in the literature (5). In contrast to expectation, the findings for several of the potential prognostic factors investigated are largely consistent across studies. There is little or no evidence of a predictive ability for any of the demographic factors examined or for depression. The trends for negative results were also observed among the subset of studies that adjusted for other covariates and had sample sizes suggesting adequate statistical power. Most reports evaluating language impairment found that poorer function was predictive of future disease course. Agitation also signalled a poorer prognosis in the

few studies that examined it and warrants further investigation. Potential prognostic factors for which the results were inconsistent include psychotic symptoms, extrapyramidal signs, previous rate of decline, and initial severity. The trends in the results across the different types of outcome measures were generally consistent for each of the factors reviewed. Overall, the heterogeneity in the progression of AD remains largely unexplained (5).

The use of different methods within several studies led to inconsistent findings regarding the statistical significance of a factor's predictive value. These include the use of different measurement scales, different methods of estimating of disease progression, and different statistical analyses. Whether the estimates of association were meaningfully different for the alternate approaches could not be assessed since these were not generally reported for associations that failed to reach statistical significance. Still, differences in the statistical significance of findings are important since most studies relied on significance testing to identify predictors of decline. These observations suggest, therefore, that conclusions regarding a factor's predictive value may depend on the method chosen and that differences in methods across studies may have contributed to observed inconsistencies in the reported findings.

Upon closer examination of individual features of study design and analysis, considerable diversity was noted in the methods used across studies. Several measurement scales were used within the different classes of outcome measures and these differed in the constructs they assessed. Analytic approaches also varied, with many studies relying on bivariate analyses only while others used multivariate statistical modelling. Reports differed in their operational definition of several clinical prognostic factors and the time window of their assessment. The literature was also characterized by different methods of estimating the annual rate of change of subjects' test scores over follow-up. The agreement among these diverse approaches has generally not been investigated in a quantitative manner, nor on theoretical grounds taking into account the analytical properties of various methods.

The prognostic studies reviewed were also characterized by several potential sources of bias. These include nonrepresentative study samples, selective attrition,

diagnostic inaccuracy, confounding by baseline severity, the use of mixed zero times, floor effects or nonlinear patterns of decline on outcome measures, and inadequate follow-up. In addition, many studies were likely limited by insufficient statistical power. The inadequate reporting of several design features, particularly the referral process and the numbers and characteristics of subjects lost to follow-up precluded a full examination of the potential for bias associated with these methods. The extent to which study results were affected by these various sources of bias is not known. Possible differences in the methodological rigour of the studies, however, may have contributed to inconsistencies in their results.

CHAPTER 4 - DATA AND METHODS

4.1 Introduction

Perhaps the most important methodological issues identified in the review of prognostic studies of AD are those related to the accuracy of estimating disease progression (section 3.3.3.6). The most common approach to estimating a subject's clinical deterioration has been to calculate the annual rate of change (ARC) of their test scores over follow-up. Various methods have been used to compute ARCs and it is unclear whether the estimates they produce are comparable. Furthermore, each method is based on the assumption that decline on the outcome measure used is linear over the portion of follow-up upon which the estimate is based. Few studies have examined the concordance among the different methods estimating ARC or the validity of the assumption of linear decline. Of those that have been conducted, many are limited by the use of inappropriate statistical analyses.

In this chapter, I further examine these two issues through separate re-analyses of data from two longitudinal studies of AD. Data on the two cohorts were provided by Dr. James Mortimer of the Veterans Administration Medical Center, Minneapolis, Minnesota (subsequently referred to as the Minneapolis sample) and Dr. John Brooks, III of the Stanford Alzheimer's Disease Diagnostic and Treatment Center (ADDTC), Palo Alto, California (subsequently referred to as the Palo Alto sample). The recruitment and selection of the study samples are detailed first, followed by a description of the data made available for each sample. Finally, statistical analyses of the presence of linear and quadratic trends in decline on the Mini-Mental State Examination (MMSE) and of the concordance among ARC estimates produced by five different computational methods are described.

4.2 Recruitment and selection of study subjects

4.2.1 Minneapolis sample

The Minneapolis sample consists of subjects recruited from the Minneapolis Veterans Administration Medical Center, and through physicians and drug studies, to

participate in a prospective, longitudinal study of primary degenerative dementia. Participants were recruited between August 1986 and December 1988. Details of the screening and selection process have been reported by Mortimer et al. (7). Ninety-three potential participants were screened to determine their eligibility for enrolment. Screening evaluations consisted of medical and family history, physical and neurologic examinations, clinical laboratory tests, neuropsychologic assessment, and computed tomography scan of the head no more than two years prior to recruitment. The eligibility criteria adopted for this thesis research were the same as those for the prognostic study of Mortimer et al. (7). Individuals were included if, at the time of recruitment, they:

- satisfied DSM-III criteria for primary degenerative dementia (77);
- were community dwelling with an identifiable and cooperative informant;
- had a score of 7 or more on the Mini-Mental State Examination (MMSE) (164);
- had mild to moderate cognitive impairment as defined by Reisberg's Global Deterioration Scale (GDS) stages 3, 4, 5 or 6 (118,119); and
- had a score of 15 or less on the Hamilton Depression Rating Scale (252).

Subjects were excluded if they:

- had a history of drug or alcohol abuse less than five years prior to evaluation (alcohol abuse was defined as a score of 3 or more on the Short Self-Administered Michigan Alcoholism Screening Test (253) which was verified with an informant); or
- had a score of 4 or more on the Modified Ischemia Scale (254) (to minimize the likelihood of including individuals with significant cerebrovascular disease).

Seventy-nine of the 93 subjects screened met the eligibility criteria and were enroled in the study. Seventy-six of these were found, retrospectively, to also satisfy NINCDS-ADRDA criteria for probable AD (9). At the time the study was published, three of 14 subjects studied postmortem were found not to have AD at autopsy. These three subjects, and eight for whom data were available for fewer than three time points, were excluded from the analyses (7). The study population for the present investigation consists of this final sample of 65 eligible subjects with a clinical diagnosis of probable AD.

4.2.2 Palo Alto sample

The Palo Alto sample was drawn from subjects identified and recruited from the Stanford ADDTC to participate in a prospective, longitudinal study of dementia. The Stanford ADDTC is located at the Veterans Administration Medical Center in Palo Alto. Participants underwent clinical diagnostic evaluations including assessments of depression using the Hamilton Depression Rating Scale (252), of possible ischemic disease using the Hachinski Ischemia scale (255), and of history of alcohol abuse using the Short Self-Administered Michigan Alcoholism Screening Test (253). Clinical diagnoses of AD were made according to NINCDS-ADRDA criteria (9) and represented a consensus of three to six clinicians.

Data for the Minneapolis sample had already been made available at the time Stanford researchers agreed to provide data on their cohort. Consequently, the eligibility criteria applied to the Palo Alto cohort, for the purposes of the present study, were specified such that the data available would be as comparable as possible to those of the Minneapolis group.

Individuals were included in this thesis research if:

- they satisfied NINCDS-ADRDA criteria for definite or probable AD;
- they had a minimum of three MMSE data points; and
- select demographic and clinical data were available for them, including data related to age, gender, education, time of onset, and to neurologic, behavioral, and psychiatric symptoms

Forty-six subjects participating in the longitudinal study were found to satisfy these criteria and thereby formed the present study's Palo Alto sample.

4.3 Data available on study subjects

4.3.1 Minneapolis data

Subjects were followed at six-month intervals until death or loss due to other reasons: three subjects voluntarily withdrew from the study. Up to three and a half years of follow-up data were made available for the current study (Table 5). The outcome data used to estimate AD progression consist of baseline and semi-annual assessments of

Table 5

Data available for the two study samples

Minneapolis sample (n=65)	Palo Alto sample (n=46)
MMSE scores	MMSE scores
age at entry	age at entry
gender	gender
education	education
age at onset	age at onset
duration of symptoms	duration of symptoms
delusions/paranoia	delusions
hallucinations	hallucinations
activity disturbances	agitation
aggression	emotional lability
sleep disturbances	wandering
affective disturbances	depressed mood
anxiety/phobia	
family history of dementia	
Global Deterioration Scale stage	

cognitive function using the MMSE (164) (Appendix 3). As mentioned in the review of outcome measures used in prognostic studies of AD (section 3.3.3.4), the MMSE is a brief cognitive rating scale which provides an index of global intellectual functioning. It consists of a mixture of questions and tasks intended to elicit information about orientation, registration, attention, calculation ability, recall, language, and praxis. The total score ranges from 0 to 30, with lower values representing greater impairment.

The subject characteristics provided include age at entry to study, gender, family history of dementia, and education. Age at onset and symptom duration at initial evaluation were estimated from historical information provided by the family. The disease stage at entry, as measured by the Global Deterioration Scale, served as a global measure of baseline severity. The presence and severity of various psychiatric and behavioral symptoms was assessed at baseline and semi-annually using the Behavioral Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD) (107). This scale yields a total score in addition to scores for each of seven subscales: paranoid and delusional ideation, hallucinations, activity disturbances, aggressive behaviour, sleep disturbances, affective disturbances, and anxieties or phobias. Ratings were based on information provided by a caregiver regarding the subject's behavioral problems during the preceding six months.

4.3.2 Palo Alto data

Palo Alto subjects were re-evaluated semi-annually at which time cognitive status and other disease symptomatology were assessed. Up to nine and a half years of follow-up data were made available for the present study including repeat assessments of cognitive function using the MMSE (Table 5). Data related to subject characteristics included age at entry to study, gender and education. Age at onset and symptom duration at entry were estimated from caregivers' recall of the time of symptom onset. Data from baseline and semi-annual evaluations of the presence of delusions, hallucinations, emotional lability, agitation, depressed mood, and wandering were also provided. These associated disease features were rated as present, absent or questionable using the California State Department of Health Services AD Diagnostic and Treatment Center form.

4.4 Statistical analyses

All analyses were performed using the SAS statistical software package (SAS Institute Inc., Cary, N.C., USA) except where indicated otherwise.

4.4.1 Assessment of linearity of decline

The validity of the assumption that disease progression in individuals with AD is linear was assessed through statistical analyses of MMSE scores for the Minneapolis sample. Group and individual trends in subjects' scores over time were analyzed. Establishing the presence of significant linear trends and the absence of nonlinear ones would support the assumption that cognitive decline in AD, as measured by the MMSE, is linear.

Unbalanced repeated measures analyses were performed using BMDP module 5V (BMDP Statistical Software Inc., Los Angeles, CA, USA) in order to assess the magnitude and significance of common linear and common quadratic trends in decline while accounting for the interdependence of the repeated MMSE scores for each subject (256,257). The initial model estimated a common linear rate of decline across subjects while controlling for baseline severity as measured by the Global Deterioration Scale (GDS). In order to avoid collinearity problems, baseline severity was treated as a dichotomous variable with less severe subjects defined by GDS stage 3 or 4, and more severe patients by stage 5 or 6. This division resulted in minimal loss of information since only two subjects were at GDS stage 3 and only one was at stage 6.

Subsequent models were formed by the stepwise addition of quadratic and interaction terms to the initial model. These evaluated the presence and significance of a common quadratic trend, and of effect modification of the linear and quadratic trends by baseline severity. To avoid variance overlap and, thus, maximize power, the common quadratic term was specified as orthogonal to the common linear.

Modelling longitudinal data requires specification of the covariance structure of the errors. Two covariance structures, the random effects model and the first-order auto-regressive model, were expected a priori to be suitable for AD data. The random effects model is based on the assumption that the covariance of residuals can be predicted from

the variance and covariance of the slopes and intercepts of individual linear trends over time. The first-order auto-regressive model assumes that the strength of the correlation between two scores for the same patient is determined by the distance in time between the scores. These two models use different numbers of parameters. Therefore, the Akaike Information Criterion (AIC) was used to compare their fit to data (258). The AIC is a general criterion for choosing between different models applied to the same data when generalized least-squares or maximum-likelihood estimation is used. The AIC statistic reflects a model's fit to data adjusted for the number of parameters used.

The repeated measures analysis was followed by a multiple linear regression analysis, using SAS statistical software, to assess the significance of individual linear trends in decline. This was accomplished by estimating the improvement in fit obtained by adding individual linear terms (slopes) to a linear regression model with a common slope and individual intercepts (see model equation below). The value of the common linear parameter in these models was restricted to the estimate obtained for the equivalent parameter in the initial repeated measures model. The rationale for this restriction was that a repeated measures estimation of the common slope accounts for the inter-correlation among repeated scores within individuals while a linear regression estimation does not. The number of measurements per subject was too small to allow for the assessment of individual quadratic trends, in addition to linear trends.

$$\hat{Y}_{it} = \sum_{i=1}^{n-1} \hat{\alpha}_{0i} * D_i + \hat{\beta}_0 + \hat{\beta}_1 * t + \sum_{i=1}^{n-1} \hat{\alpha}_{1i} * D_i t$$

where i = 1, 2, ..., n

$\hat{\beta}_0$ = common intercept

$\hat{\beta}_1$ = common slope parameter which is restricted to the estimate obtained for the equivalent parameter in the repeated measures model

t = time

D_i = dummy variable where $D_i = 1$ for the ith subject and 0 for all others

$\hat{\alpha}_{0i}$ = individual intercepts

$\hat{\alpha}_{1i}$ = individual slopes

Baseline severity was expected, a priori, to be an important predictor of subsequent MMSE scores. Consequently, the linearity analyses were performed on the Minneapolis data first since it was the only sample for which an independent measure of baseline severity was available. The analyses of the Minneapolis data confirmed the predictive importance of baseline GDS scores and, since no such measure of severity was available for the Palo Alto sample, the linearity analyses were not performed on the latter.

4.4.2 Calculation of annual rates of change

Several methods of estimating the annual rate of change (ARC) of test scores for individual subjects are reviewed in section 3.3.3.6. Of these, the two-point method, the adjusted two-point method, the least squares regression slope of the linear model, and the multiple interval method were applied to the MMSE data from subjects in both samples. Because the length of follow-up in each of the two samples was variable among subjects, a simple difference in scores between the first and last assessments would not have generated an annual rate, and was therefore not considered further. A fifth method called the trilinear model was also applied to the data. This approach was proposed by Brooks et al. (259) for the study of decline in AD and has since been applied to longitudinal data from AD patients to describe their disease progression (171).

The annual rate of change in cognitive function generated from each of these five methods was calculated using some or all of the serial MMSE measurements. Since lower MMSE scores indicate poorer performance, negative ARC values indicate decline in cognitive function. In order to avoid a floor effect, MMSE scores obtained subsequent to a subject attaining a value of zero on the scale were not used in any estimation of ARC.

The two-point method, as its name implies, produces an ARC estimate based on only two data points. It was computed, for each subject, by subtracting the baseline MMSE score from the last available score and dividing this difference by the time elapsed, in years, between the two measurements.

The adjusted two-point method is a strategy that was proposed by Fleiss (248) and used by Lucca et al. (6) to control for the spurious association between baseline scores

and their corresponding simple difference scores, due to regression to the mean. The approach involves removing the influence of the baseline score on its associated difference score through statistical adjustment. This method is extended in the present study to the situation where the measure of change is the two-point ARC. The formula for the adjusted two-point ARC for the i-th subject is

$$\Delta_i = D_i - \hat{\beta} (Z_i - \overline{Z}) \qquad i = 1, 2, ..., n$$

where

Δ_i = adjusted two-point rate for i-th subject

D_i = unadjusted two-point rate for i-th subject

Z_i = baseline score for i-th subject

$\overline{Z}$ = mean baseline score for the sample

$\hat{\beta}$ = slope of the least squares regression of D on Z.

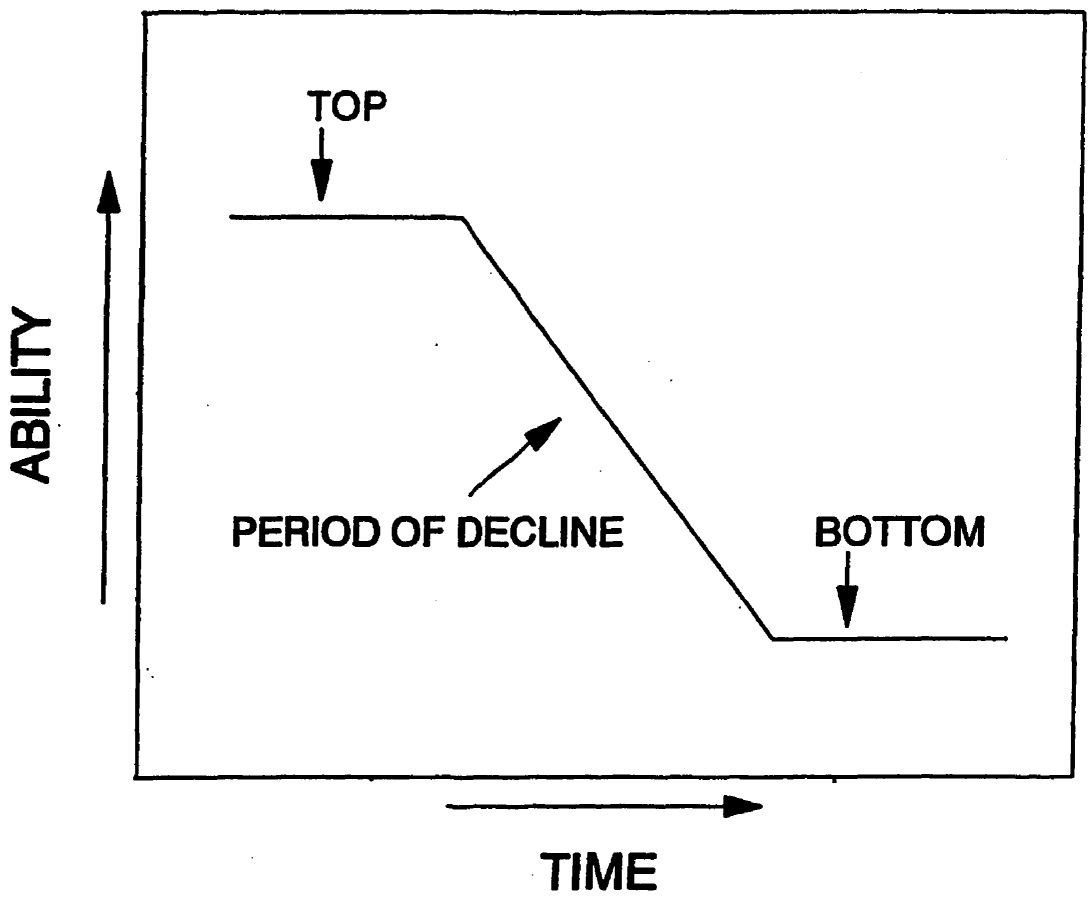
The quantity $\hat{\beta} (Z_i - \overline{Z})$ represents the portion of a subject's two-point ARC estimate (D_i) that can be predicted from his/her baseline score. The resulting adjusted rate (Δ_i) can be interpreted as the i-th subject's "expected" ARC assuming that his/her baseline score was equal to the mean baseline score of the sample. While this approach has the advantage of making the adjusted rates comparable for subjects with different levels of baseline severity, it effectively removes the *entire* influence of baseline scores on the rates of change. This would be justified only if one assumes that the entire association is an artifact of regression to the mean, which may be an extreme assumption.

The linear regression method (linear model) generates an ARC estimate for each subject through least-squares regression of their MMSE scores on time, in years, since baseline assessment. The regression coefficient, or slope, expressing this linear function provides an ARC estimate based on all available data points for a given patient.

The trilinear model of disease progression in AD postulates three phases: an initial period of stability or "top" in which there is no perceptible decline, a period of decline where rates of deterioration vary for individual subjects, and a final period of stability termed the "bottom" (Figure 2) (259,260). When applied to the longitudinal data of individual subjects, this analytic model estimates whether a subject's trajectory over follow-up encompasses one, two or all three of these phases. The model determines if

Figure 2

An illustration of the trilinear model



and when the phases occur by applying all possible combinations of the two cut-points (transition points where decline begins and ends) to a subject's set of outcome measurements. The combination producing the best fit is selected using least-squares analysis.

The ARC calculation for the trilinear model also consists of a least-squares regression of MMSE scores on time. Unlike the linear model, however, this regression is based only on those data points corresponding to the portion of follow-up estimated to be the decline phase. Brooks et al. (259) maintain that the trilinear model more closely reflects the pattern of observed decline in AD than does the linear model. By excluding periods of relative stability, the trilinear model, they argue, provides a more accurate ARC estimate than those of methods that assume linear decline throughout AD. This purported advantage applies whether the periods of stability represents a "true" plateau or an artifact due to poor sensitivity of the outcome instrument. The trilinear model requires a minimum of five data points per subject. It was applied to each subject's MMSE data using software developed and supplied by Dr. John Brooks (259).

The multiple interval method (MIM) of estimating ARC, like the linear model, uses all available data points for each subject. It is unique, however, in that several ARC values may be generated for a single subject. Each subject's total follow-up is divided into consecutive one year periods. An ARC value is computed for each one year interval by subtracting its initial or baseline score from the MMSE score corresponding to the end of the interval. This approach requires that subjects be reassessed at uniform time intervals. Furthermore, it assumes that change within intervals is linear and that change scores across intervals within a given subject can be analyzed as though they constitute independent observations.

4.4.3 Comparison of annual rates of change

The concordance among ARC values generated from the various approaches selected for comparison was examined in both an aggregate and pairwise fashion. Aggregate comparisons included plotting the ARC values from all methods for each subject in the two samples. The *within-subject* variability among ARC estimates from different methods was quantified by determining the range among all ARC values, and

the range among the ARC estimates generated by the multiple interval method alone.

The multiple interval method was excluded from pairwise comparisons since it produced more than one ARC value for 69% and 52% of Minneapolis and Palo Alto subjects, respectively. The rates derived from this method cannot be meaningfully combined as they are intended to represent independent snapshots of a subject's disease course. Nor could the results of one-on-one comparisons of each of these ARC values with those of the other methods be meaningfully combined to assess their overall concordance.

Pairwise comparisons consisted of computing the mean and standard deviation of ARC estimates for each method. The statistical significance of the difference between mean ARCs produced by the different approaches was assessed using paired *t* tests. Since moderate violations of the normality assumption were observed for some comparisons, the Wilcoxon sign rank test was also used. Difference scores were also computed for all six pairwise comparisons between the four methods (method-pairs) and their distributions were graphed. Other measures of agreement examined for each method-pair included the Pearson product moment correlation and the root mean square error (RMSE). The RMSE was calculated as a measure of the mean absolute difference between each method-pair of ARC estimates for individual patients.

The intra-class correlation coefficient (ICC) was also considered since it represents the correlation among ARC estimates generated by two different methods, accounting for any systematic differences between them. The ICC was calculated according to the method of Bartko using two-way mixed-effects ANOVA where "methods" constituted a fixed factor while "subjects" were considered a random factor (261). The presence of interaction between the "subjects" and "methods" factors could not be assessed. Therefore, both lower and upper bounds of each ICC were computed. The lower bound represents a conservative estimate where interaction is assumed to be zero (261).

The results of these pairwise comparisons prompted further analyses to explore the differences observed. These included plotting each of the linear - trilinear, two-point - trilinear, and adjusted two-point - trilinear difference scores against the trilinear decline portion of total follow-up. Linear and trilinear models were also fitted to the progression curves of select subjects for whom the corresponding ARC values differed substantially.

CHAPTER 5 - RESULTS

5.1 Introduction

In this chapter, the two study samples are compared with respect to demographic features, disease duration and severity at study entry, and length of follow-up. The results of analyses performed to assess the presence of linear and quadratic trends in cognitive decline are presented next. Finally, the variability and concordance among annual rate of change estimates generated from five different computational methods are described.

5.2 Characteristics of the study samples

Subjects in the Minneapolis sample were slightly older at entry and at symptom onset, were more likely to be male, and had a lower level of education than Palo Alto subjects (Table 6). The preponderance of males in these two samples, compared to those of most prognostic studies, likely reflects the fact that both studies were conducted in Veterans Administration Medical Centers. The median level of cognitive functioning, as measured by the MMSE score at study entry, was higher in the Palo Alto sample: 50% of Palo Alto subjects had an initial MMSE scores above 20 compared to 26% of Minneapolis subjects. The two groups were identical with respect to the median duration of symptoms at study entry.

The median length of follow-up for the Palo Alto sample was more than double that of Minneapolis subjects. This did not translate, however, into a proportionately higher number of MMSE data points, reflecting the fact that Palo Alto subjects, in general, missed more of their follow-up visits. Still, nine Palo Alto subjects (20%) had more than eight data points which was the maximum number in the Minneapolis sample.

5.3 Linearity of decline

Of the two covariance models considered in the analysis of repeated measures, the first-order auto-regressive provided slightly better fit to data (lower Akaike Information Criterion value) than the random effects. Therefore, the results presented here are those of models that assume this structure of the covariance matrix.

Table 6

Demographic, clinical, and study characteristics of the two study samples

Characteristics	Minneapolis sample (n=65)	Palo Alto sample (n=46)
Median age in years (range)	69 (53-90)	67.5 (55-88)
% male	73.8	63.0
% high school education or less	53.8	28.3
Median age at onset in years (range)	64 (49-89)	61.5 (47-81)
Median symptom duration in years (range)	3 (0-13)	3 (0.5-9.5)
Median MMSE score (range)	17 (7-28)	20.5 (2-29)
Median length of follow-up in years (range)	2 (1-3.5)	4.5 (1-9.5)
Median number of data points (range)	5 (3 - 8)	6 (3 - 14)

The common linear rate of decline, as estimated by the initial BMDP model, was 3.8 MMSE points/year ($P<0.0001$; 95% CI= 3.3, 4.4). Baseline severity was also highly significant ($P<0.0001$), indicating that a subject's initial severity rating was highly predictive of subsequent MMSE scores. Neither the common quadratic term ($P=0.19$) nor the interaction between the common linear and baseline severity terms ($P=0.48$) were found to be significant when added separately to the initial model. The addition of the interaction terms between baseline severity and each of the linear and quadratic variables failed to provide strong evidence of effect modification by the former ($P=0.12$ and $P=0.10$, respectively).

A modest non-linear trend, with some tendency for mean scores to level off, was observed after two years of follow-up (Figure 3)[1]. This trend was more pronounced among the subgroup of patients more severely demented at baseline[2]. It must be interpreted with caution, however, since the observed MMSE means are based on the subset of subjects available at each visit, with those of the last three visits based on seven, seven, and three subjects respectively in the more impaired baseline group. Selective attrition of more severe subjects may also have contributed to the observed levelling of mean scores. Those with no more than two years of follow-up had MMSE means of 3.2 to 5.5 points lower at the first five visits than those followed for longer than two years.

The equivalent graph for the 11 subjects who attended all eight visits did not suggest the presence of a non-linear trend in later visits (Figure 4). The fact that the mean progression curves corresponding to the two baseline severity groups are essentially parallel in each figure also suggests the absence of effect modification by baseline severity. The improvement in the log likelihood afforded by the models with and without the two interaction terms, relative to the null model, was virtually identical (202.74 and 202.76, respectively), providing further support for a lack of interaction.

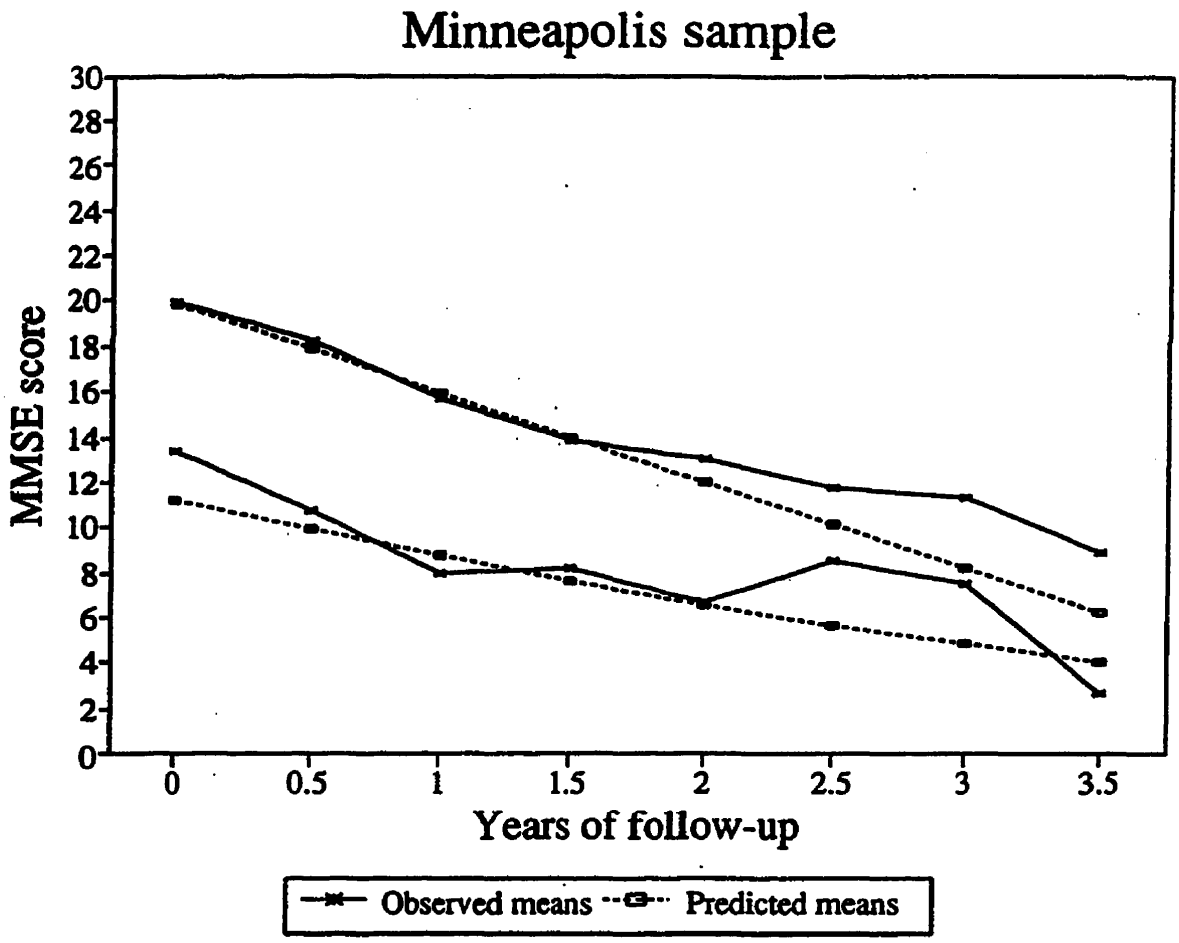
---

[1] The predicted MMSE means in Figure 2 are based on the repeated measures ANOVA model including all parameters: linear, baseline severity, quadratic, and the interaction terms between baseline severity and each of the linear and quadratic terms.

[2] The more severe baseline subgroup was defined by GDS=5 or 6 (lower MMSE scores) and the less severe subgroup by GDS=3 or 4 (higher MMSE scores).

Figure 3

Predicted* and observed mean MMSE scores by baseline severity group**



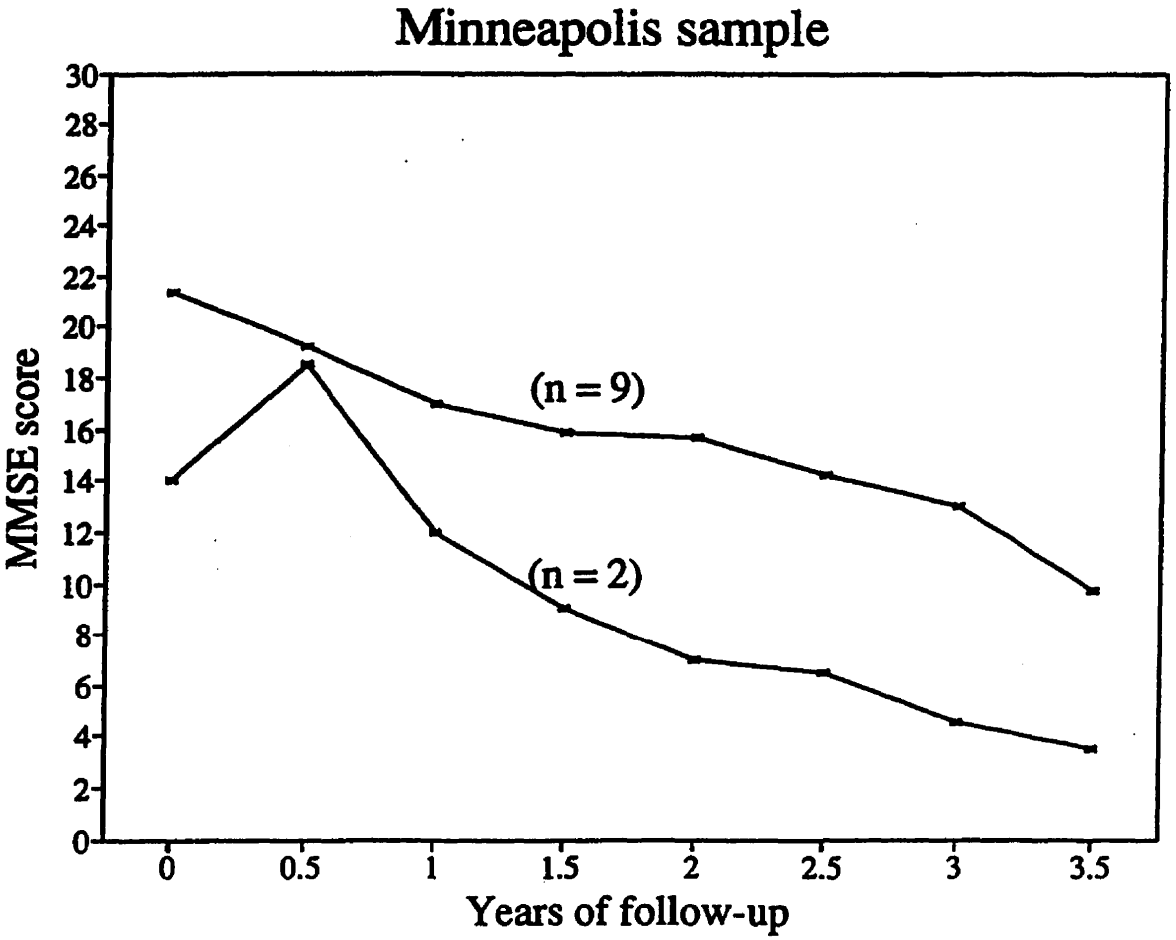
Year of follow-up	0	0.5	1	1.5	2	2.5	3	3.5
n (less severe)	39	39	39	37	29	20	16	10
n (more severe)	26	26	25	18	17	7	7	3

* Predicted means are based on the repeated measures ANOVA model including all parameters: linear, baseline severity, quadratic, and the interaction terms between baseline severity and each of the linear and quadratic terms.

** The more severe baseline subgroup was defined by GDS=5 or 6 (lower MMSE scores) and the less severe subgroup by GDS=3 or 4 (higher MMSE scores).

Figure 4

Observed mean MMSE scores by baseline severity group* for subjects attending all follow-up visits (n=11)



* The more severe baseline subgroup was defined by GDS=5 or 6 (lower MMSE scores) and the less severe subgroup by GDS=3 or 4 (higher MMSE scores).

The least squares regression model with a common linear slope and individual intercepts explained 84% of the variance in MMSE scores. The introduction of individual linear terms to this model accounted for an additional 9% of the total variance ($P<0.001$).

5.4 Comparison of annual rates of change

5.4.1 Sample sizes available

Given that a minimum of three data points was required for inclusion in the study, ARC values could be calculated for all subjects using the two-point, adjusted two-point, and linear regression methods. The trilinear method, however, requires a minimum of five data points. Consequently, trilinear estimates could only be calculated for 44 (68%) and 31 (67%) subjects in the Minneapolis and Palo Alto samples respectively. Furthermore, since the pattern of follow-up visits for seven Palo Alto subjects was such that no two visits fell one year apart, a MIM ARC estimate could not be computed for these subjects. At least one MIM ARC value could be calculated for each Minneapolis subject as only five missed one visit and no subject missed more than one.

5.4.2 Within-subject variability among annual rates of change

Annual rate of change (ARC) values obtained from four of the five methods selected for comparison are shown for six subjects from each sample in Figures 5 and 6. These subjects were selected from each sample as follows: 1) the distribution of the range of ARC estimates within subjects was computed for the subset of subjects who had an ARC estimate from each of the five methods; and 2) the distribution was then divided into six equal portions and one subject was randomly selected from each of the portions. Due to software restrictions in the number of values which could be displayed per subject, a maximum of three out of a possible nine MIM ARC values were plotted for the Palo Alto sample. The adjusted two-point ARC values were not graphed for the same reason (as will be shown later, the adjusted two-point and two-point methods yielded comparable results for all subjects). For both samples, the within-subject variability among ARC estimates produced by the different methods was marked for several subjects. ARC values generated from the MIM method alone also appeared to be widely spread in many cases.

Figure 5

Annual rate of change estimates generated from selected methods

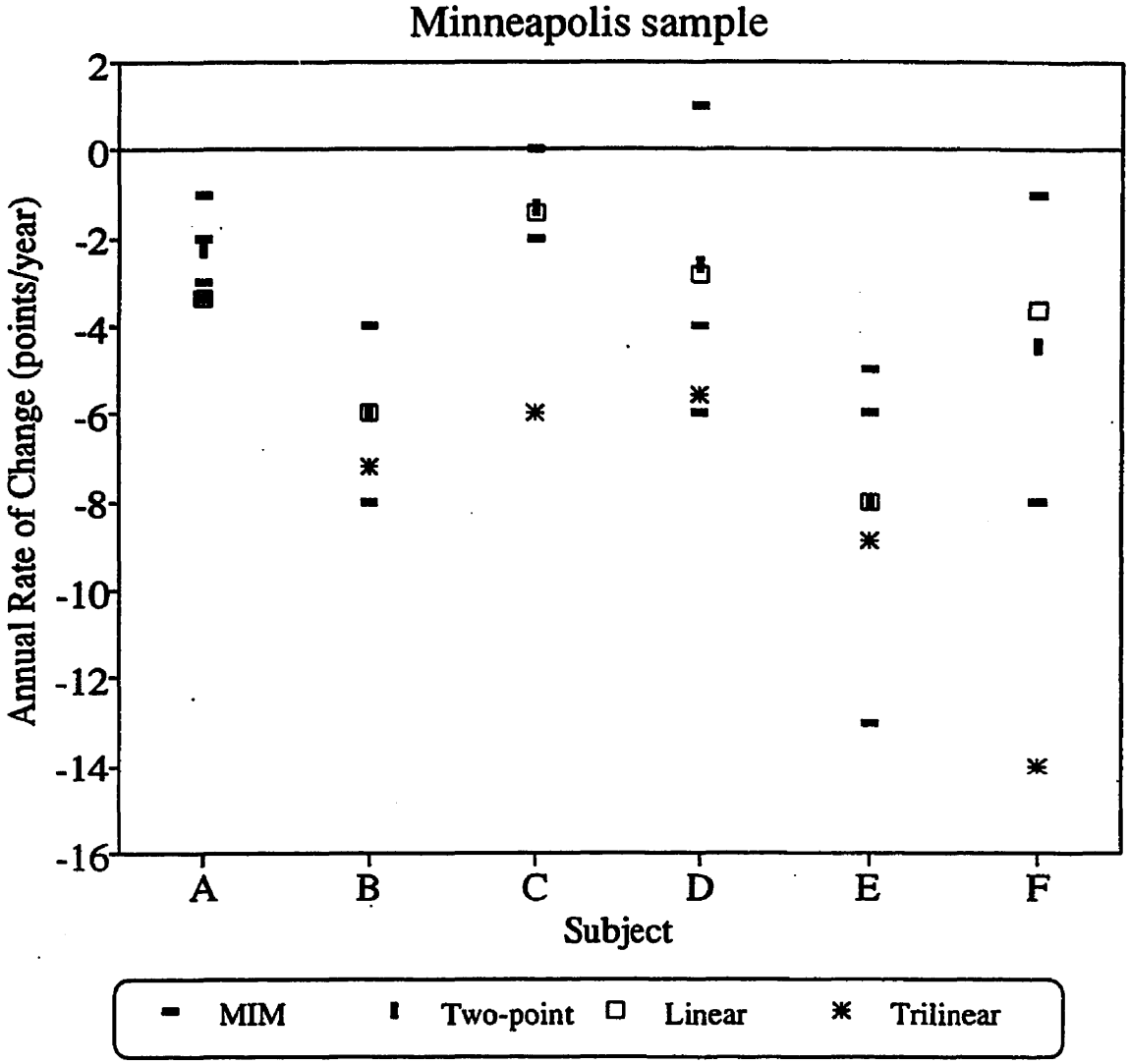
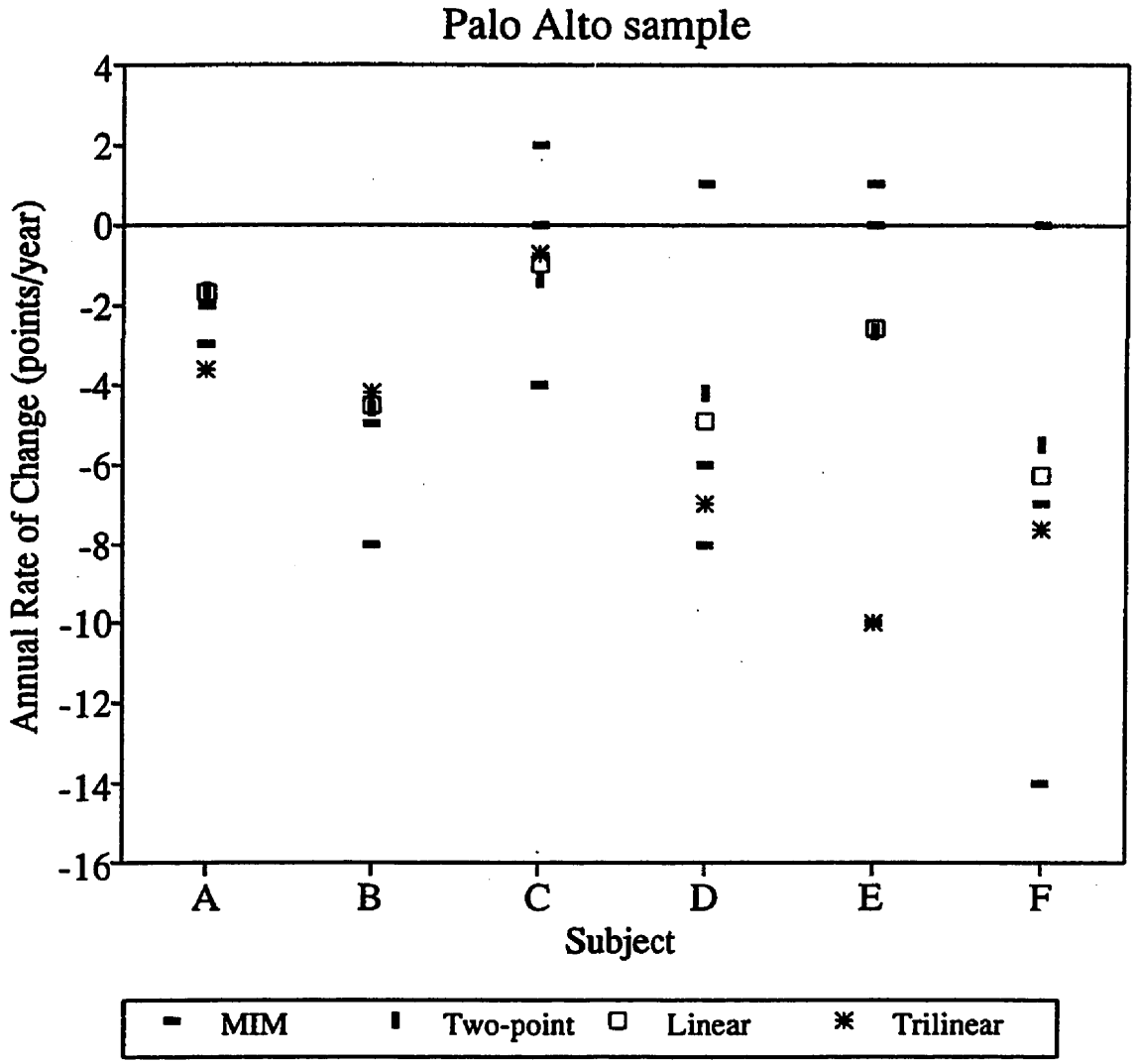


Figure 6

Annual rate of change estimates generated from selected methods



Distributions of the range of ARC estimates within subjects (i.e., the difference between the highest and lowest among estimates generated by different methods for the same subject) are shown in Figure 7. Fifty-seven percent of the Minneapolis sample and 46% of the Palo Alto sample had a range of four or more MMSE points/year among their various ARC estimates. In addition, 17% and 13% respectively had ranges greater than or equal to 10 MMSE points/year. Distributions of the range of ARC values generated by the MIM alone also showed large within subject variability (Figure 8). Among those subjects with greater than one MIM ARC value, 73% in the Minneapolis group and 67% in the Palo Alto group had ARC ranges of four or more MMSE points/year.

5.4.3 Pairwise comparisons of annual rates of change

5.4.3.1 *Means and standard deviations*

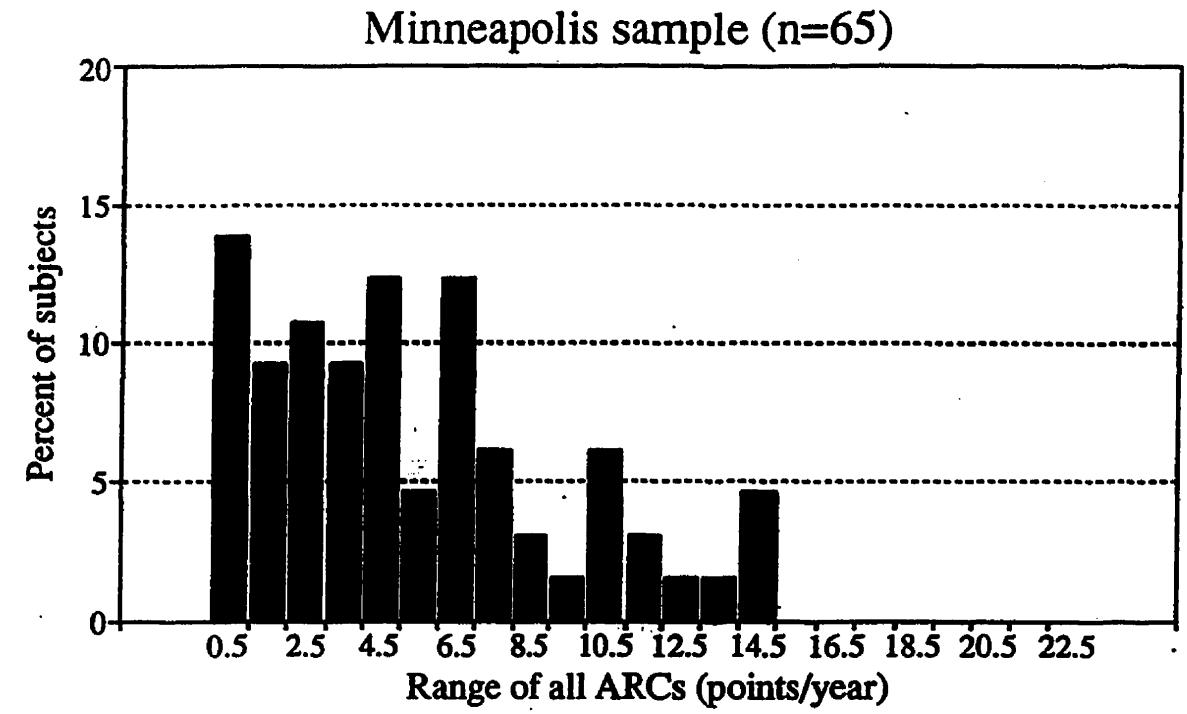
The ARC means and standard deviations produced by the methods selected for comparison are presented in Table 7. The two-point, adjusted two-point, and linear regression methods generated remarkably similar means and standard deviations within each of the data sets. None of the corresponding paired *t* tests reached statistical significance at the 5% level. It should be noted, however, that the ARC means and standard deviations of the Minneapolis group are approximately a full MMSE point higher than those of the Palo Alto data for each of these three methods.

For both samples, the mean trilinear ARC was significantly larger than those of the other three methods ($P<0.05$). Among subjects with ARC values for all four methods, the trilinear mean was larger by 2.3 to 2.4 MMSE points/year for the Minneapolis sample and by 1.9 to 2.0 for the Palo Alto sample. The trilinear ARC means differed less between the two data sets than did the ARC means of the other three methods. It is noteworthy that, in both samples, the variation in the trilinear ARC estimates among subjects was of the same order of magnitude as that of estimates derived from other methods. This finding suggests that the large variability in rate of change among individuals with AD is not simply a consequence of having included subjects in a stable phase (factored out by the trilinear model) along with those who are declining.

Since one of the Palo Alto subjects had an extreme trilinear ARC value (-26 MMSE points/year), the analyses were repeated for this sample excluding the subject.

Figure 7

Distribution of the range of all annual rate of change (ARC) estimates within subjects



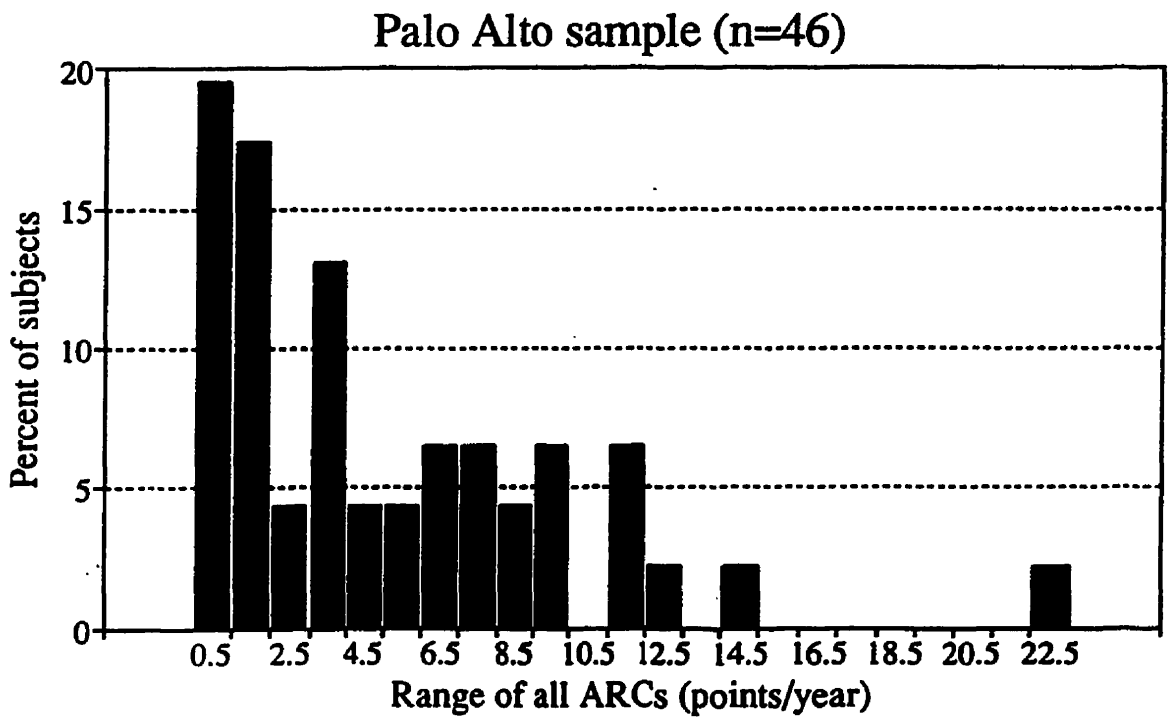
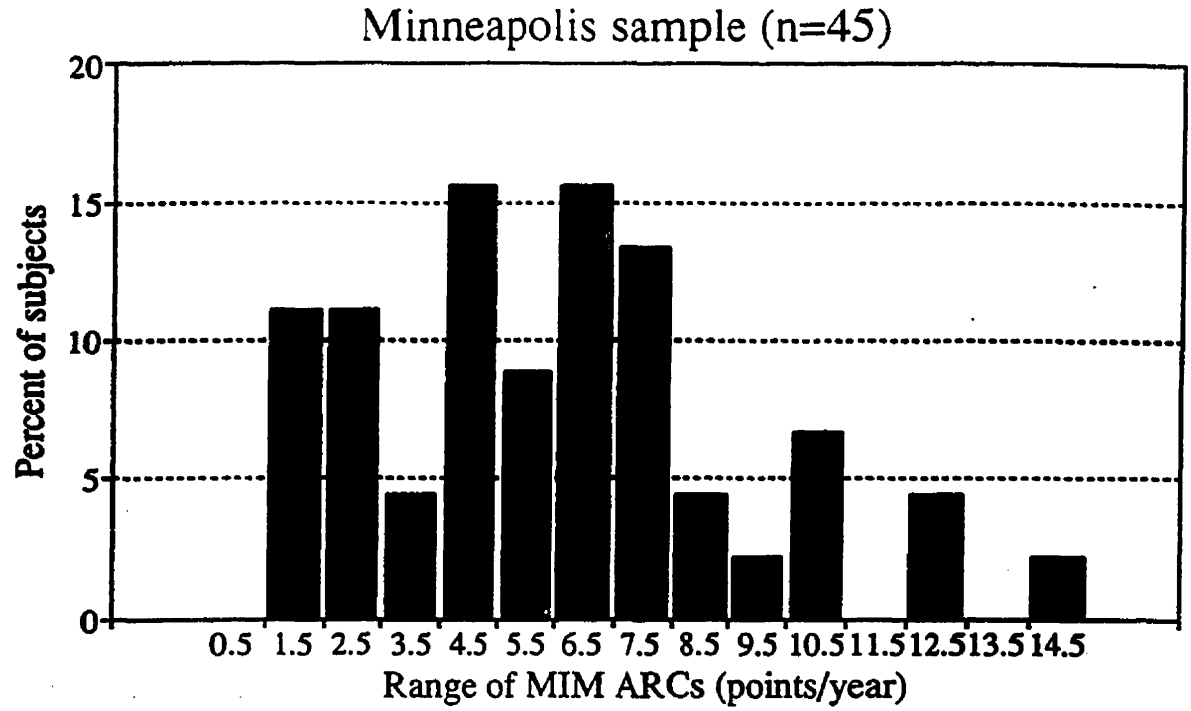


Figure 8

Distribution of the range of multiple interval method (MIM) annual rate of change (ARC) estimates within subjects



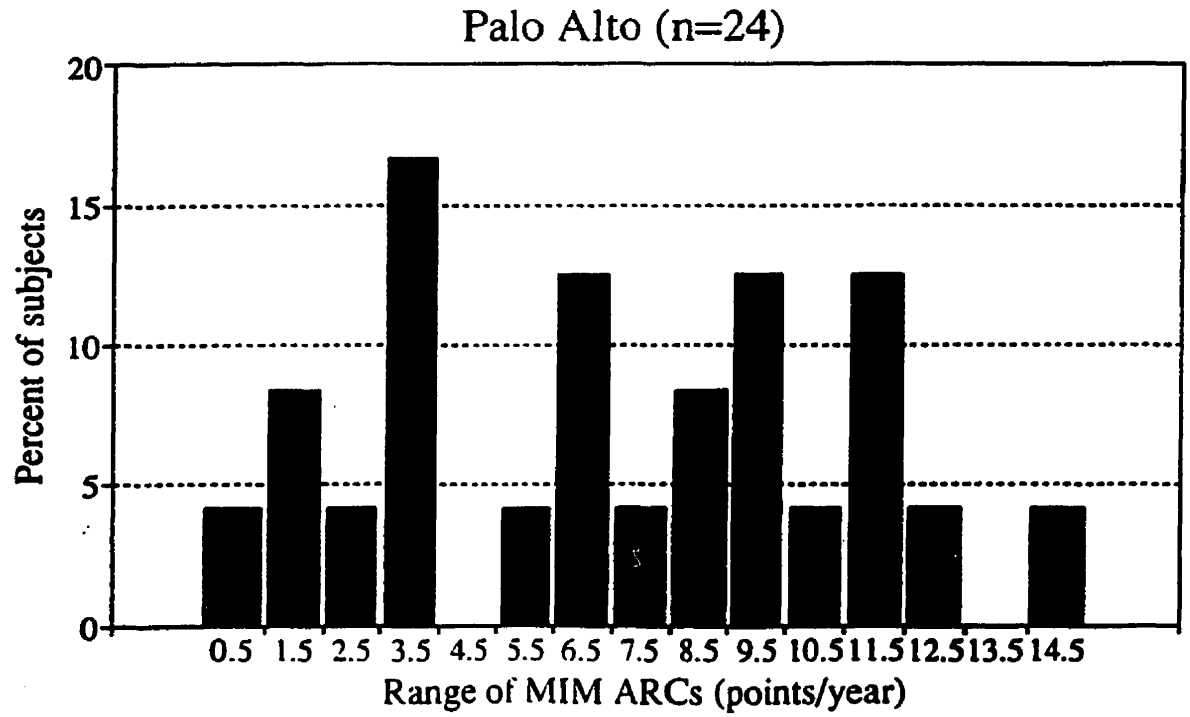


Table 7

Mean annual rate of change (ARC) estimates generated from selected methods

A) Minneapolis sample:

Method of calculating ARC		Mean [SD]*	*P*-values of paired *t* tests (n)			
			1	2	3	4
1	Two-point	-4.5 [3.2]	—			
2	Adjusted two-point	-4.5 [3.2]	1.00 (65)	—		
3	Linear regression	-4.4 [3.1]	0.22 (65)	0.22 (65)	—	
4	Trilinear model	-6.1 [3.6]	0.0001 (44)	0.0001 (44)	0.0001 (44)	—

B) Palo Alto sample:

Method of calculating ARC		Mean [SD]	*P*-values of paired *t* tests (n)			
			1	2	3	4
1	Two-point	-3.7 [2.0]	—			
2	Adjusted two-point	-3.7 [2.0]	0.94 (46)	—		
3	Linear regression	-3.7 [2.0]	0.90 (46)	0.89 (46)	—	
4	Trilinear model	-5.7 [4.4]	0.02 (31)	0.01 (31)	0.02 (31)	—

* Standard deviation.

The trilinear ARC mean and standard deviation dropped to -5.1 and 2.3 MMSE points/year respectively. This trilinear ARC mean was still significantly larger than those of the other three methods ($P<0.05$). With the outlier excluded, however, the difference in the trilinear ARC means between the two samples was consistent with the sample difference in ARC means for the other three methods.

Because of moderate violation of the assumption of normality for estimates obtained with some methods being compared, the Wilcoxon signed rank test was also computed. The statistical significance of the results remained unchanged.

5.4.3.2 *Correlations*

Correlations among the various ARC values displayed the same pattern of agreement as that for the means (Table 8). Near perfect product-moment correlations were observed for the estimates of change produced by the two-point, adjusted two-point, and linear regression methods for both samples. Correlations among the values for these three methods and those of the trilinear method were moderate for the Minneapolis data (r=0.51 to 0.56), while the corresponding correlations for the Palo Alto sample were low (r=0.19 to 0.26). These calculations were also repeated with the extreme value from the Palo Alto sample omitted. The resulting correlations between the trilinear method's ARC estimates and those of the other methods increased to a level comparable to that of the Minneapolis group (r=0.44 to 0.53).

5.4.3.3 *Intra-class correlation coefficients*

The agreement between the methods, as measured by the intra-class correlation coefficient (ICC), also demonstrated a similar pattern of results (Table 9). Near perfect agreement was observed among the ARC values generated by the two-point, adjusted two-point and linear regression methods as indicated by ICC coefficient bounds above 0.94. The upper and lower bounds of ICC estimates for the comparison of trilinear ARC values to those of the other methods were considerably lower and more widely spread. This was especially true for the Palo Alto sample. Although these bounds represent a considerable

Table 8

Correlations* among annual rate of change (ARC) estimates generated from selected methods

A) Minneapolis sample:

Method of calculating ARC	1	2	3	4
1 Two-point	1.00 (65)			
2 Adjusted two-point	0.99 (65)	1.00 (65)		
3 Linear regression	0.98 (65)	0.98 (65)	1.00 (65)	
4 Trilinear model	0.56 (44)	0.57 (44)	0.51 (44)	1.00 (44)

B) Palo Alto sample:

Method of calculating ARC	1	2	3	4
1 Two-point	1.00 (46)			
2 Adjusted two-point	0.99 (46)	1.00 (46)		
3 Linear regression	0.95 (46)	0.95 (46)	1.00 (46)	
4 Trilinear model	0.26 (31)	0.26 (31)	0.19 (31)	1.00 (31)

* Pearson product-moment correlation coefficients.

Sample size upon which the correlation is based in parentheses.

Table 9

Intra-class correlation coefficients* of annual rate of change (ARC) estimates generated from selected methods

A) Minneapolis sample:

Method of calculating ARC		1	2	3	4
1	Two-point	1.00 (65)			
2	Adjusted two-point	0.99 (65)	1.00 (65)		
3	Linear regression	0.98 (65)	0.98 (65)	1.00 (65)	
4	Trilinear model	0.50-0.60 (44)	0.51-0.60 (44)	0.45-0.57 (44)	1.00 (44)

B) Palo Alto sample:

Method of calculating ARC		1	2	3	4
1	Two-point	1.00 (46)			
2	Adjusted two-point	0.99 (46)	1.00 (46)		
3	Linear regression	0.96 (46)	0.95 (46)	1.00 (46)	
4	Trilinear model	0.16-0.41 (31)	0.16-0.40 (31)	0.11-0.38 (31)	1.00 (31)

* Reported as the upper and lower bounds of the intra-class correlation coefficient. A single estimate is given where upper and lower bounds do not differ as of the second decimal place.

Sample size upon which the intra-class correlation coefficient is based in parentheses.

range of possible values for the ICC estimates, their low values signify poorer agreement between the trilinear method and each of the other three in terms of the rate values they produced.

5.4.3.4 *Root mean square errors*

The root mean square error (RMSE) of ARC values for each method-pair is shown in Table 10. For the Minneapolis sample, the RMSE for method-pairs involving the trilinear method were of the same order of magnitude as each of the individual methods' standard deviation (SD). The equivalent RMSE values for the Palo Alto data were the same order of magnitude as the trilinear method SD, and twice the magnitude of the other methods' SDs. In contrast, the RMSE values for the two-point, adjusted two-point and linear regression method-pairs were several orders of magnitude smaller than the individual method SDs in both samples. These results indicate that the average "between-method" difference in ARC estimates produced by the trilinear method compared to each of the other three is at least equivalent to the mean "between-subject" difference in ARC values for any of the given methods.

5.4.3.5 *Distributions of difference scores*

For all subjects in both samples, the two-point and adjusted two-point methods produced very similar results: differences in their ARC values did not exceed an absolute value of one MMSE point/year. The small impact of the statistical adjustment used to generate the adjusted two-point estimate reflects a lack of influence of baseline scores on the two-point ARC. The correlation between the two-point rate of change and baseline MMSE scores was -0.04 and -0.09 for the Minneapolis and Palo Alto samples, respectively. The two-point and adjusted two-point methods also generated comparable ARC values to that of the linear regression slope. The pairwise ARC differences fell within one MMSE point/year for at least 89% of subjects in both samples and none exceeded a difference of two MMSE points/year.

Differences in ARC values for the trilinear method compared to the other three, however, were substantial in a marked portion of Minneapolis subjects (Figures 9, 10 and

Table 10

Root mean square error of annual rate of change (ARC) estimates generated from selected methods

A) Minneapolis sample:

Method of calculating ARC		SD	Root Mean Square Error (RMSE)			
			1	2	3	4
1	Two-point	3.2	—			
2	Adjusted two-point	3.2	0.1 (65)	—		
3	Linear regression	3.1	0.5 (65)	0.5 (65)	—	
4	Trilinear model	3.6	3.7 (44)	3.7 (44)	3.9 (44)	—

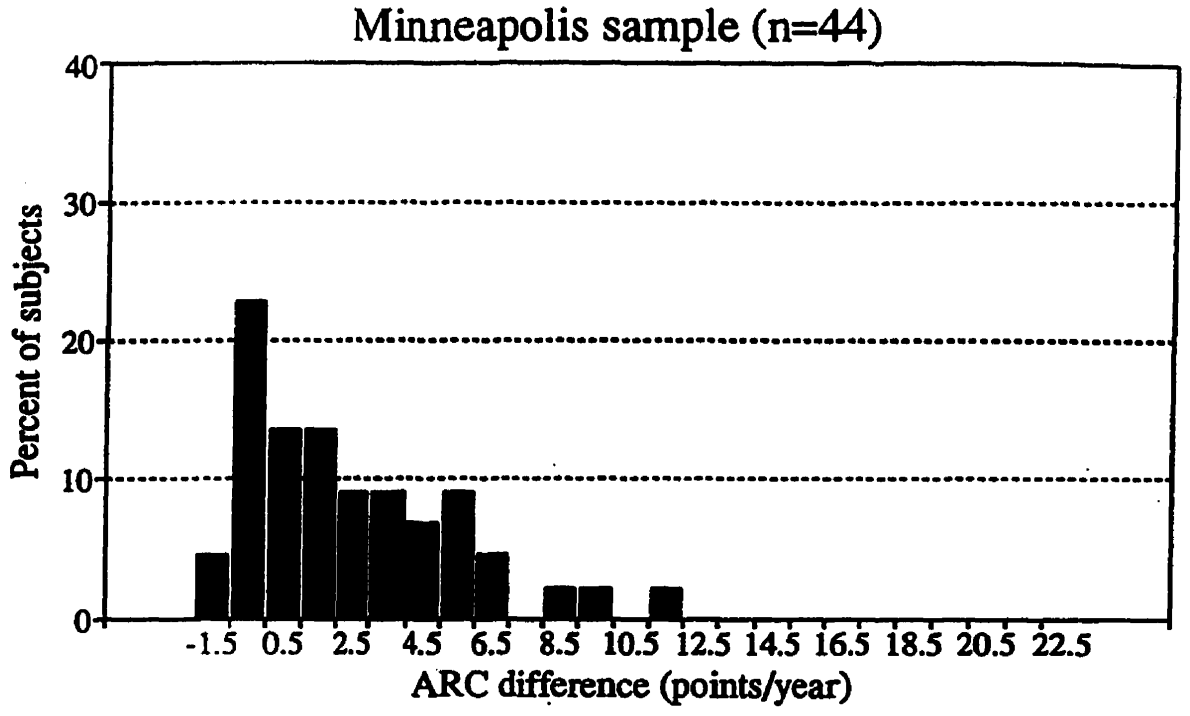
B) Palo Alto sample:

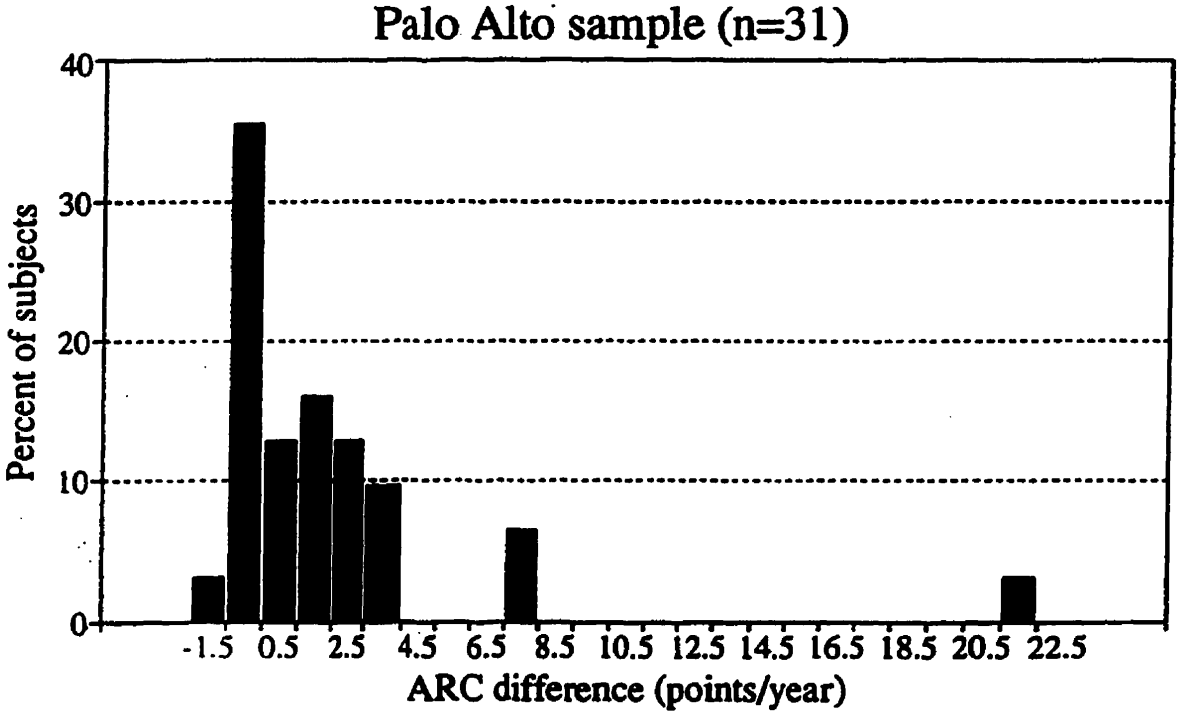
Method of calculating ARC		SD	Root Mean Square Error (RMSE)			
			1	2	3	4
1	Two-point	2.0	—			
2	Adjusted two-point	2.0	0.2 (46)	—		
3	Linear regression	2.0	0.6 (46)	0.6 (46)	—	
4	Trilinear model	4.4	4.6 (31)	4.6 (31)	4.7 (31)	—

Sample size upon which the RMSE is based in parentheses.

Figure 9

Distribution of the difference in annual rate of change (ARC) estimates generated by the two-point and trilinear methods*

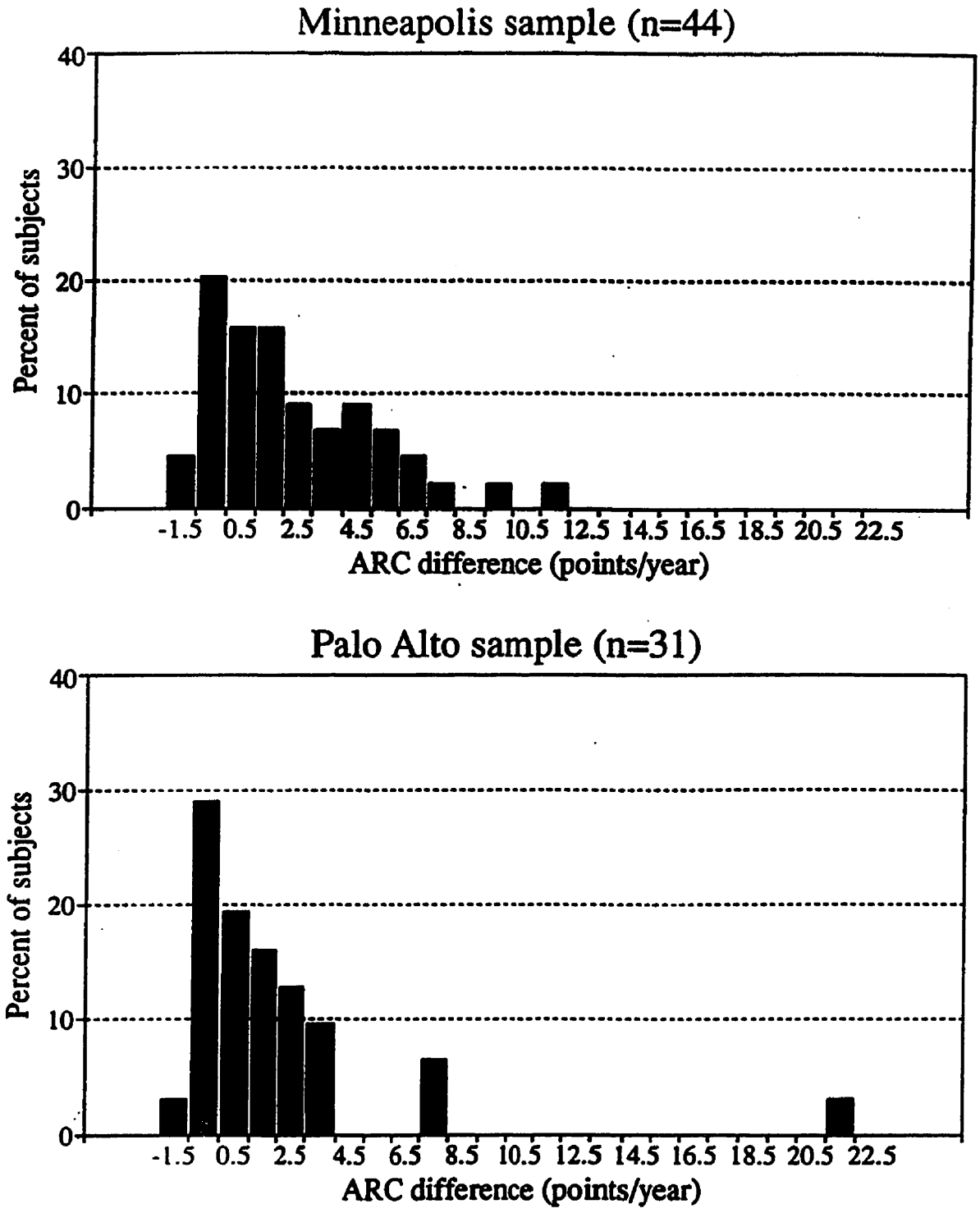




* Two-point ARC minus trilinear ARC.

Figure 10

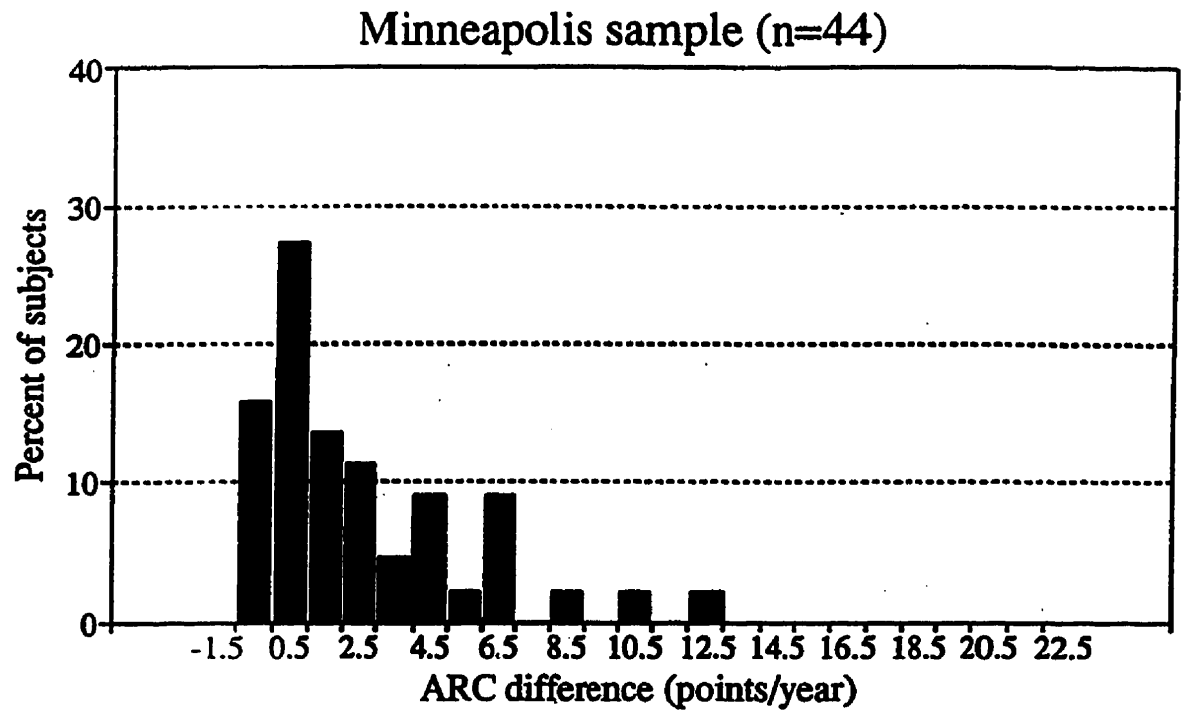
Distribution of the difference in annual rate of change (ARC) estimates generated by the adjusted two-point and trilinear methods*

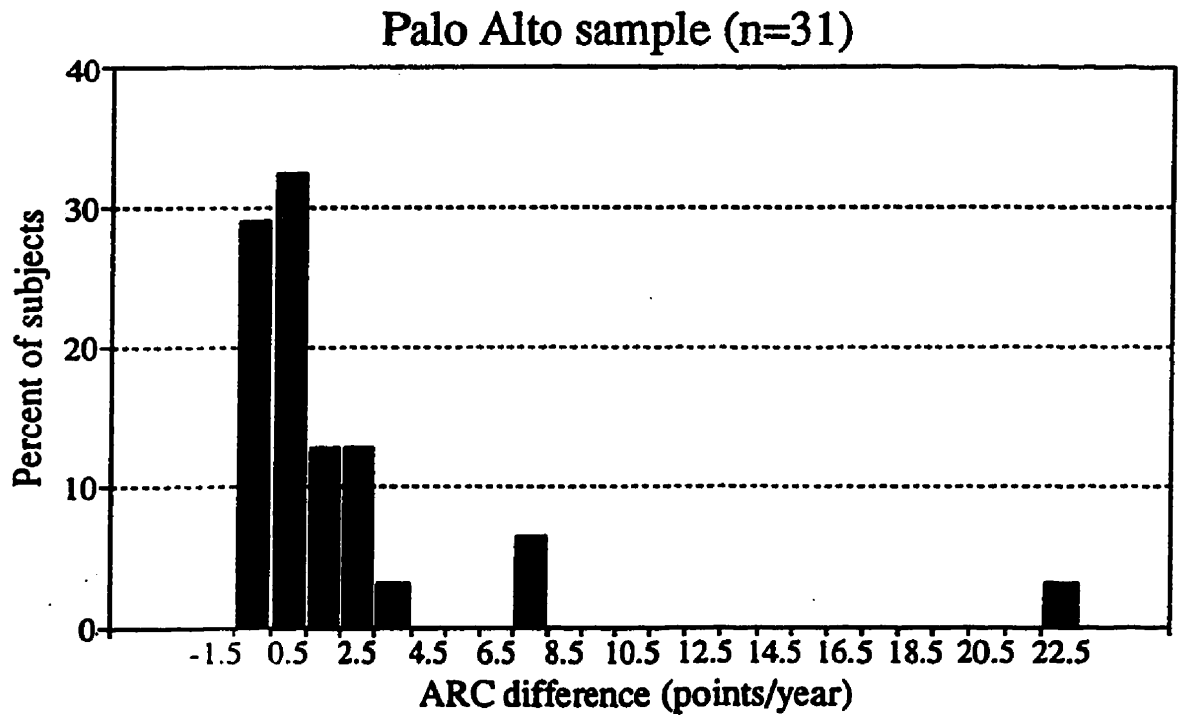


* Adjusted two-point ARC minus trilinear ARC.

Figure 11

Distribution of the difference in annual rate of change (ARC) estimates generated by the linear and trilinear methods*





* Linear ARC minus trilinear ARC.

11). For example, of those Minneapolis subjects with a trilinear ARC value, 12 (27%) had a difference of four MMSE points/year or more between this value and that produced by each of the other three methods. Conversely, the Palo Alto sample yielded sizeable differences for only three subjects (10%) (Figures 9, 10 and 11). For one of these subjects, the difference between the trilinear method ARC and each of the other three methods exceeded twenty MMSE points/year. In both samples, the distribution of the ARC difference scores between the trilinear method and each of the other three was skewed to the right. This indicates that in the vast majority of subjects for whom these ARC values diverged, the trilinear method had estimated a faster rate of decline.

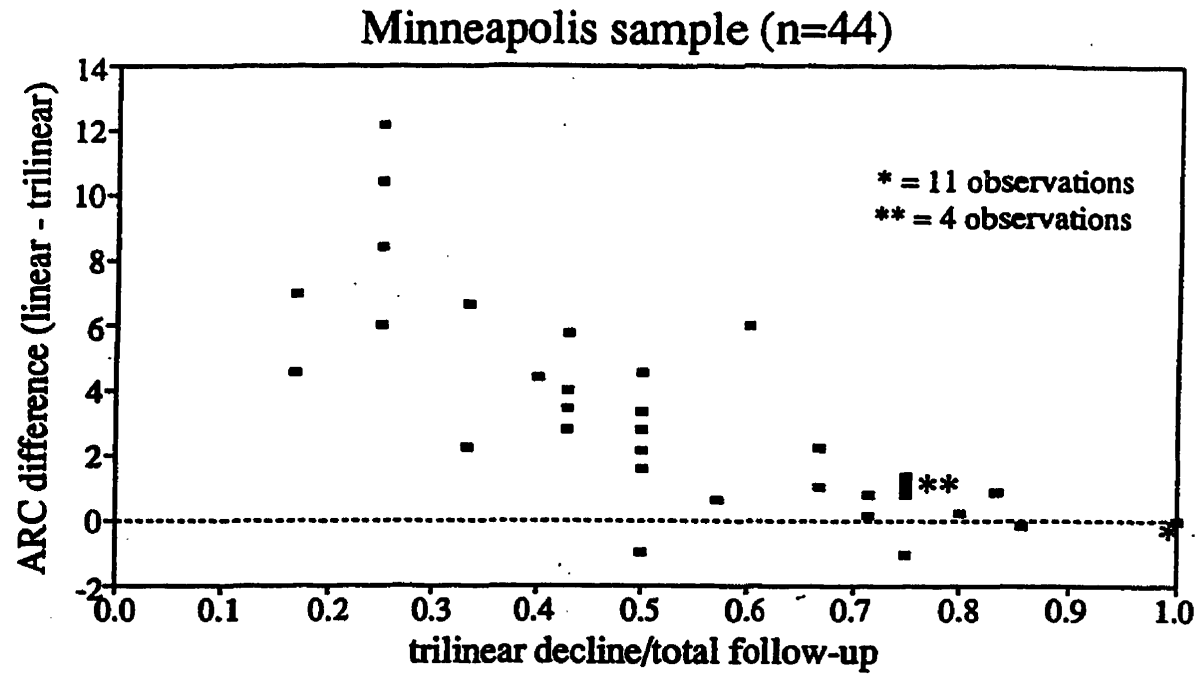
There were a few exceptions to this trend however. Four subjects declined less, or even improved, by at least one MMSE point/year according to the trilinear model compared to one or more of the other three methods. In two cases, subjects experienced a sharp decline at the beginning or end of a trajectory that was otherwise characterized by gradual deterioration. The trilinear ARC estimate for these subjects was based on the entire follow-up period (i.e., the slope of a linear regression of all data points on time). Thus, the trilinear method estimated a slower annual rate of decline than the two-point method which was based on only the first and last measurements. Scores for the other two subjects showed no clear trend over time but fluctuated somewhat over at least part of their follow-up. The trilinear estimate, in these cases, was based on a portion of follow-up where subjects improved over their previous performance.

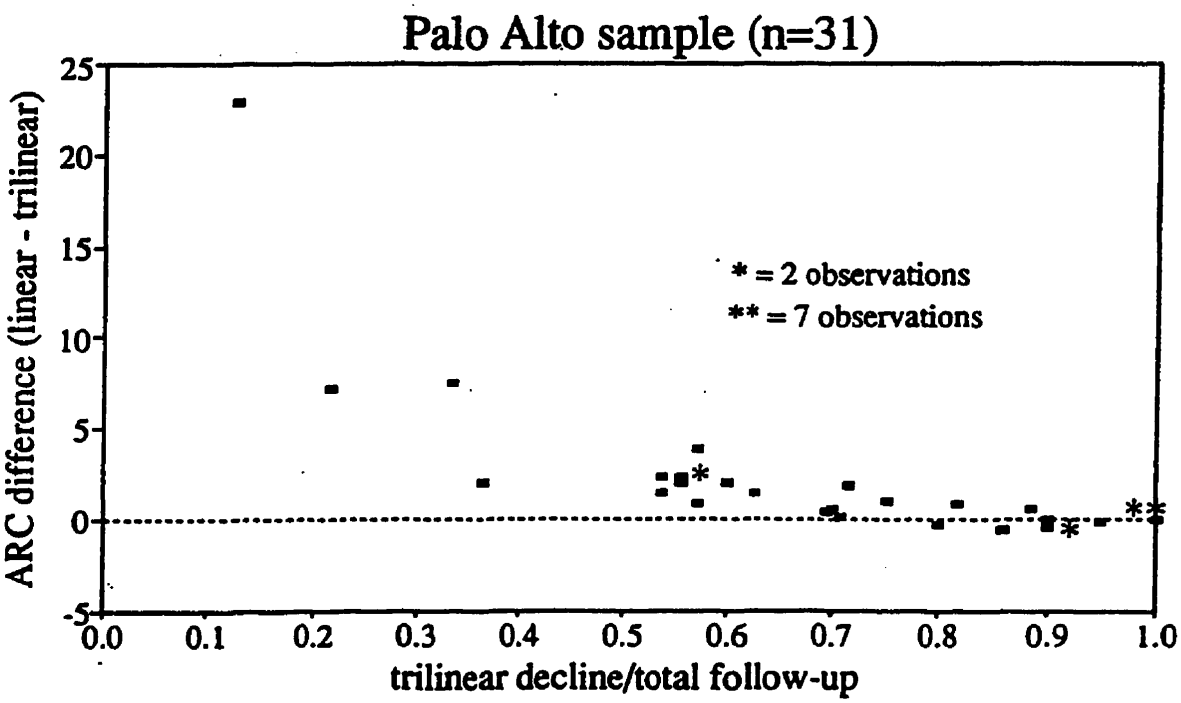
5.4.3.6 *Difference scores as a function of trilinear decline*

Plots of the difference between trilinear and linear ARC estimates as a function of the trilinear decline portion of total follow-up (i.e., the proportion of total follow-up upon which the trilinear estimate was based) are shown in Figure 12. These plots were motivated by the consistent finding of meaningful differences in the ARC values generated by the trilinear method compared to those of the other three methods, as well as by a priori expectations. By excluding relatively stable periods, the trilinear model was expected to estimate greater decline than those methods based on the entire follow-up. It was hypothesized that the difference would increase as the ratio of trilinear

Figure 12

Differences in linear and trilinear annual rate of change (ARC) estimates as a function of the trilinear decline portion of total follow-up





decline-phase to total follow-up period decreased.

A strong, negative linear relationship was observed for both samples, where larger discrepancies in the ARC values between the two methods were associated with trilinear ARC values based on relatively small portions of total follow-up. The correlations corresponding to these plots were r=-0.79 for the Minneapolis sample and r=-0.73 for Palo Alto subjects. The latter increased to r=-0.82 when calculated excluding the outlier. Negative correlations of the same magnitude were found when this comparison was repeated for the trilinear method with each of the two-point and adjusted two-point methods.

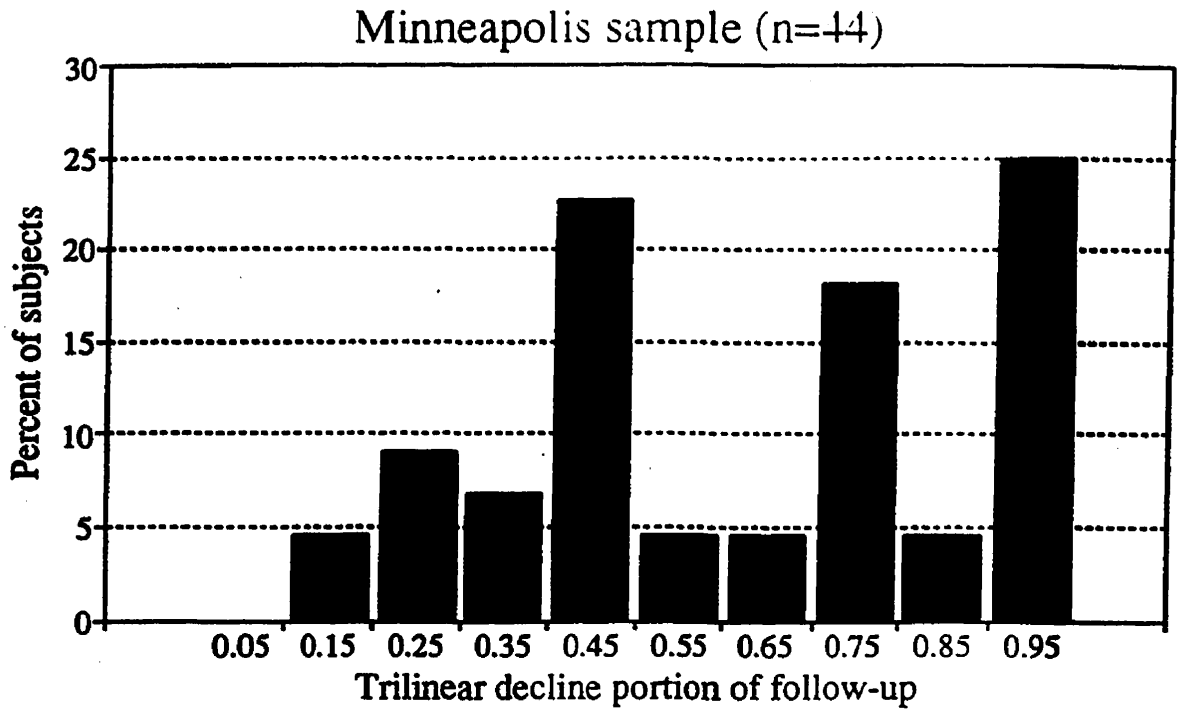
Of the 12 Minneapolis subjects with linear-trilinear ARC differences of four or more MMSE points/year, all but one had a trilinear decline portion of 50% or less (Figure 12). The same was true for all three of the corresponding subjects in the Palo Alto sample (Figure 12). Despite a similar trend, only 13% of the Palo Alto sample (n=4) had trilinear decline portions of 50% or less compared to 43% of Minneapolis subjects (n=19) (Figure 13). This difference in the distribution of trilinear decline portions between the two samples may explain the earlier observation that relatively few Palo Alto subjects had large linear-trilinear decline differences compared to the Minneapolis sample.

5.4.3.7 *Linear and trilinear models fitted to progression curves*

Plots of the linear and trilinear models fitted to individual progression curves are shown for a sample of subjects with linear-trilinear ARC differences of four or more MMSE points/year (Figure 14). These graphs illustrate the potential for dramatic differences in how these models fit individual data, the result of which is divergence in their estimates of ARC. This potential is greatly enhanced when a subject experiences a proportionately long period of little or no decline, a period which the trilinear model may deem to be a plateau and consequently exclude from the assessment of rate of change. In extreme cases, all but a small portion of follow-up is considered a plateau, as was the case for subjects A and C in Figure 14. In fact, all subjects in both data sets whose trilinear ARC values were based on just two data points six months apart had linear-trilinear ARC differences of greater than four MMSE points/year (n=6 for

Figure 13

Distribution of the trilinear decline portion of total follow-up



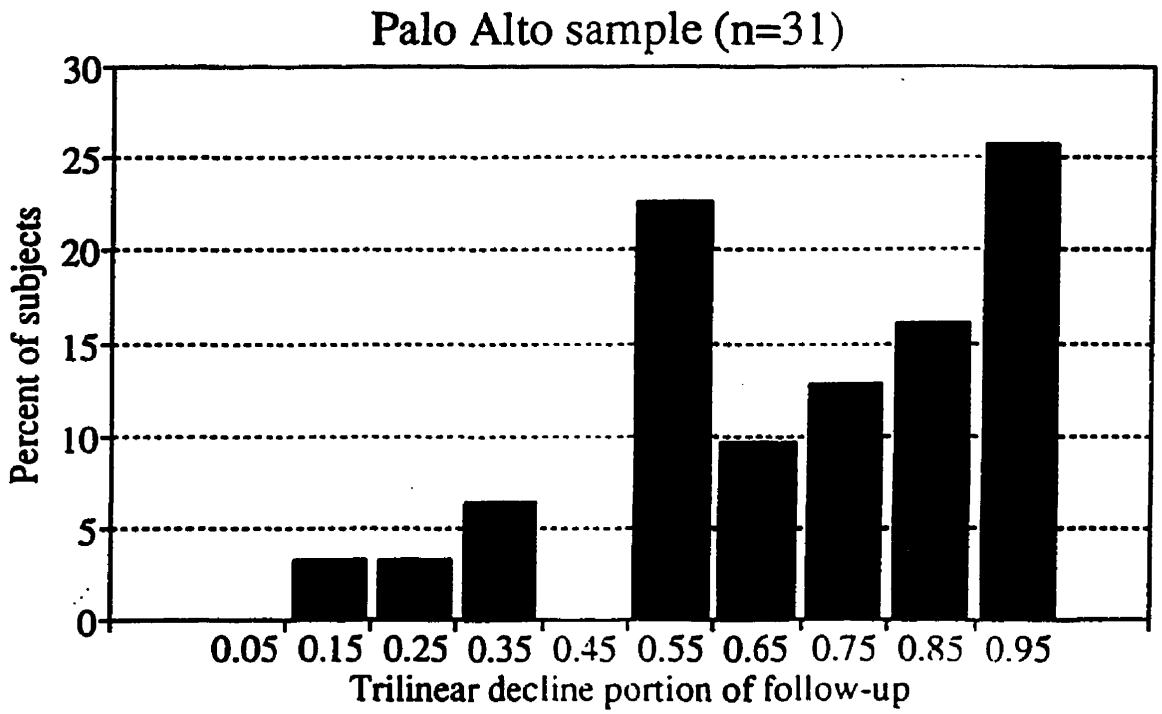
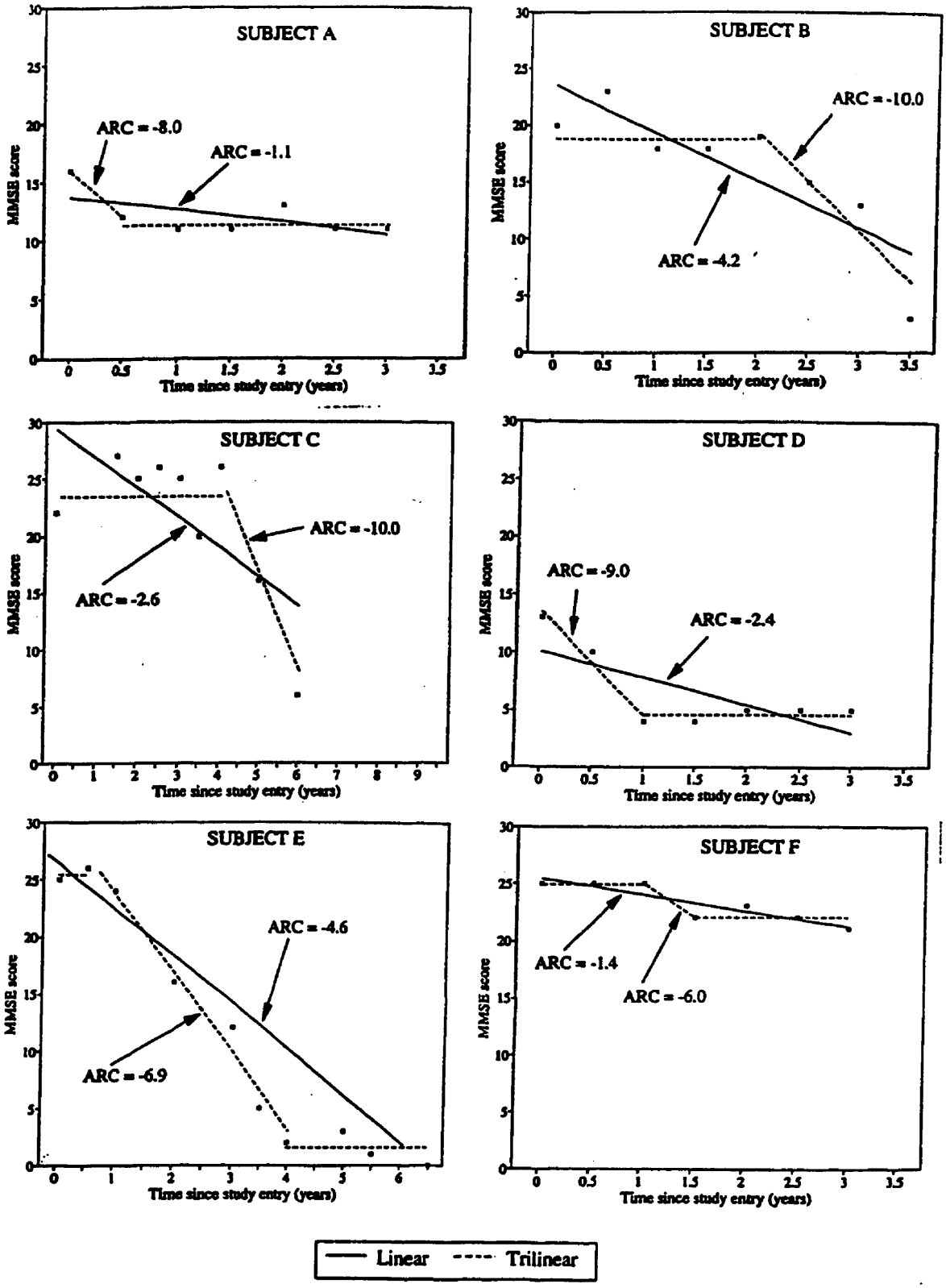


Figure 14

Linear and trilinear models fitted to the progression of six subjects on the MMSE



Minneapolis group, n=1 for Palo Alto group).

The premise of the trilinear model is that a subject's ARC may be more accurately estimated by excluding periods of stability in mild and severe disease (e.g., subject E, Figure 14) (259,260). Of those to whom the trilinear model could be applied, 33 Minneapolis subjects (75%) and 24 Palo Alto subjects (77%) were estimated to have "top" and/or "bottom" plateau periods. It is noteworthy, however, that seven Minneapolis and two Palo Alto subjects had estimated "tops" below MMSE=17 while eight and two patients from each sample respectively had "bottoms" greater than MMSE=10 (e.g., subject A, Figure 14). Furthermore, five Minneapolis and two Palo Alto subjects were estimated to have had all three trilinear phases within a score range of less than 10 MMSE points (e.g., subject F, Figure 14). These findings indicate that Brooks et al.'s (259) trilinear model software does not restrict the range of scores over which "tops" and "bottoms" are estimated. The estimation of plateaus in moderate disease severity, however, is not consistent with the rationale for this method. Also, there is little mention in the literature regarding the existence of periods of stability in moderate stages of AD. These observations should be considered in future applications of this model.

CHAPTER 6 - DISCUSSION

6.1 Methodological issues in prognostic studies of AD

Fifty-nine studies evaluating potential predictors of the clinical course of Alzheimer's disease (AD) satisfied eligibility criteria for inclusion in the review. In contrast to expectation, the results for several of the clinical and demographic predictors investigated were reasonably consistent across studies. Most studies assessing the prognostic value of age at symptom onset, age at study entry, gender, symptom duration, education, family history of dementia, and depression reported negative results. Language impairment and agitation, on the other hand, were generally found to be predictive of more rapid deterioration in AD. These two factors appear to be the most promising among the potential predictors investigated to date and therefore merit further research.

The findings for other potential prognostic factors were inconsistent. These include psychotic symptoms, extrapyramidal signs, previous rate of decline, and initial disease severity. Further research on the predictive value of these factors, using rigorously designed studies, will hopefully allow regularities in the results to emerge. In the meantime, the wide variability in the rate of disease progression among individuals with AD remains largely unexplained (5).

Diversity of study methods

Several features of study design and analysis were found to be highly varied among the reports including the outcome measure used, the method of assessing potential prognostic factors, the method of estimating disease progression, and the analytic approach used to identify predictors. This methodological diversity makes comparisons of the study results difficult. Moreover, the use of different methods may have contributed to discrepancies in the results. This possibility is supported by the observation that different approaches used within the same study led to discordant conclusions regarding a factor's predictive value. In particular, discrepancies were noted in conjunction with the use of different cognitive scales (178,186), different approaches to estimating disease progression (207), and different prognostic analyses within

individual studies (192,193). In one study, for example, agitation was significantly associated with faster cognitive decline as measured by the Mini-Mental State Examination but not by the Dementia Rating Scale (186).

Among the various classes of outcome measures identified, cognitive rating scales were used most frequently, followed by activities of daily living (functional) scales, global staging instruments, and neuropsychologic test batteries. Trends in the results for individual potential predictors were comparable across the different types of outcome measures. In some studies, however, the use of both cognitive and functional scales was associated with divergent conclusions (6,7,208). Even within classes of outcome measures investigators used a variety of scales and these varied with respect to the number and types of abilities measured.

The operational definition of several potential predictors, including language impairment, psychosis, extrapyramidal signs and depression, differed considerably among the reports. Studies also varied with respect to the time frame of the assessment of clinical predictors. While most determined the presence of clinical features at study entry, others used various time windows prior to enrolment. Still others assessed the presence of symptoms over the same period of follow-up upon which the estimate of disease progression was based. The latter approach is inappropriate in prognostic research where the goal is to identify predictors of *future* disease course.

Three different approaches were used to assess disease progression. These consisted of estimating the annual rate of change (ARC) of tests scores, the time to reach a given score or clinical endpoint, and determining whether or not subjects progressed according to some criteria for change. The method of computing ARC also varied across studies. It is unclear whether ARC estimates obtained from the different approaches are comparable. The statistical analyses performed to identify predictors followed from the three methods of estimating disease progression (e.g., *t*-tests or multiple linear regression of ARC estimates). In all three cases, most studies performed bivariate analyses only. More sophisticated approaches involving the direct analysis of repeated measures (repeated measures ANOVA and growth curve modelling) were only used in a few studies.

It is not clear which of these analytic approaches, if any, is more appropriate for the study of prognosis in AD. There are several issues to consider, however, in choosing an approach. Practical considerations include the minimum number of data points required per subject and whether the time interval between assessments must be uniform, as is the case for repeated measures ANOVA. There are also conceptual issues inherent to each strategy. The rate of change method is based on the assumption of linear decline on outcome measures, longitudinal analysis requires making assumptions about the covariance structure of errors, and the survival approach raises issues of stability and clinical significance in the choice of endpoint.

Validity of study methods

The prognostic studies reviewed were plagued by several potential sources of bias including nonrepresentative study samples, selective attrition, diagnostic inaccuracy, inadequate sample sizes, confounding by baseline severity, mixed zero times, and poor reliability and sensitivity to change (floor effects) of outcome scales. The extent to which individual studies were influenced by these sources of bias appears to have been varied. This, in turn, might also account for inconsistent findings in the literature.

Approximately 75% of prognostic studies were based solely or largely on convenience samples of clinic patients. The representativeness of such samples is doubtful and reasons for referral may be related to prognosis. For example, individuals with more rapid progression, earlier symptom onset, or associated behavioral or psychiatric symptoms may be preferentially referred to specialized clinics. This not only limits the generalizability of the results but also suggests the potential for selection bias. Insight into the magnitude of selection bias in these studies was precluded, however, by a lack of information regarding the referral process.

An evaluation of potential selection bias due to nonrandom attrition was also hindered by poor reporting. Fifty-eight percent of the studies made no mention of losses to follow-up. Most of these, however, required a minimum number of data points and/or follow-up for inclusion in the report. Thus, any subjects who died, dropped-out, or were otherwise lost to follow-up before satisfying these criteria would have been excluded. Of

those studies which reported losses, 56% experienced an attrition rate of greater than 20% of the original cohort. Furthermore, many failed to provide adequate, if any, information related to the reasons for attrition and the distribution of prognostic factors among those patients lost.

Only those studies using contemporary diagnostic guidelines were included in the review. Even adherence to these guidelines, however, may have resulted in as many as 30% of the study samples not having had AD (83,86). This creates the potential for bias due to contamination of the samples by subjects with other disorders and consequently different prognoses. In addition to excluding patients with other causes of dementia, some studies also excluded those with other health problems that might affect brain function (e.g., diabetes, cancer). It is not clear if, and to what extent, these exclusions improve diagnostic accuracy. Their use, therefore, may have unnecessarily limited the representativeness of the samples and, hence, the generalizability of the results. The exclusion criteria used will also affect the number of subjects available for study which, in turn, determines the statistical power to identify important prognostic factors. The number of subjects analyzed in the 59 reports ranged from 16 to 430. Two thirds of the studies likely had insufficient power to detect prognostic factors present among 15% of patients, and half of these were likely unable to identify predictors with 50% prevalence.

It has been suggested that initial disease severity may confound the association between other potential predictors and disease progression (7,8). A review of studies of the association between initial severity and disease progression found that severity was a significant predictor in the few studies that used time-to-endpoint estimates of decline. Studies using rate of change estimates of progression were inconsistent however. While most of these found some relationship, the nature of the reported associations varied. The evidence suggests, therefore, that baseline severity may be a potential confounder but that the nature of its association with the rate of progression on outcome scales is unresolved. Prognostic studies were varied with respect to the extent to which they controlled for potential confounding by baseline severity. Forty-one percent of studies failed to control for severity, while another 20% were considered to have only achieved partial control. The validity of the findings of these studies, therefore, may be questionable.

Most reports assembled subjects who were at various stages in their disease course (mixed zero times). Whether a given subject has a symptom of prognostic interest at enrolment, however, may well depend on how far they have progressed into the disease at that time. For example, a patient enroled at a mild stage of their disease may not yet have language impairment while the same subject, recruited later in their disease, would likely have this potential prognostic factor. Examining the prognostic importance of the presence of clinical features at study entry among subjects who are at a variety of severity levels is potentially problematic since trends may be obscured. In the presence of effect modification, a prognostic study's findings would be driven by the sample's particular mix of zero times. Furthermore, differences between study samples in their distribution of zero times may contribute to inconsistencies in their results.

Measurement instruments used to monitor disease progression must be reliable and sensitive measures of change. Few of the scales used in the studies reviewed have been assessed for these properties. Of those that have been evaluated, estimates of change appear to be reasonably reliable when based on follow-up periods of more than one year (160,163). In 22% of studies, however, the minimum length of follow-up was less than one year. The assessment of sensitivity to change has been hindered by the lack of a gold standard measure of the clinical progression of AD. The results of several studies suggest that some cognitive and functional scales may suffer from floor effects (e.g., 7,151,152,251). The presence of floor effects has generally been interpreted as reflecting poor sensitivity to change in severe disease. The findings of some studies also suggest that disease progression, as described by some scales, may be nonlinear (152,160,170). The possibility that decline on scales used in these studies is nonlinear is important since many of the methods used to estimate progression assume linearity. If this assumption does not hold for particular scales, estimates of the rate of deterioration on these measures may be biased.

6.2 Estimation of disease progression

One of the main methodological challenges in prognostic research of AD is the accurate estimation of disease progression. The approaches used in the studies were

diverse both in terms of the measurement instruments used and the method of estimating progression from outcome data. Several issues related to the comparability of the different computational methods and the measurement properties of the scales were identified in the literature review. Two of these issues were examined in greater depth through illustrative data analyses: whether the assumption of linear decline on outcome scales is valid and whether different methods of computing the annual rate of change (ARC) of test scores produce comparable estimates.

The re-analyses were performed on Mini-Mental State Examination (MMSE) data from two longitudinal cohorts of probable AD patients (Minneapolis and Palo Alto). Subjects in both cohorts were recruited from AD referral clinics and were similar with respect to age, gender, and disease duration at study entry. Palo Alto subjects were less cognitively impaired at intake and had longer follow-up with more data points.

Linearity of decline

Statistical analyses evaluating the assumption of linear decline confirmed the presence of significant group and individual linear trends in MMSE scores over time (Minneapolis sample). The common rate of decline across Minneapolis subjects was estimated to be 3.8 MMSE points/year. This value is within the range of mean MMSE ARC estimates reported in the literature (7,166). The data failed to provide convincing evidence of a common quadratic trend or of effect modification of the linear or quadratic trends by baseline severity. The power to detect a significant departure from the linear trend, particularly in later visits, may have been limited since only a few subjects attended all follow-up visits. Even with the limited amount of data, however, no evidence for a quadratic effect was found either in early follow-up, where more data was available, or across the entire of follow-up (3.5 years) of those subjects with all data points. Thus, the presence of significant linear trends and the absence of a common quadratic effect support the assumption of linear decline on the MMSE within the range of severity represented in the Minneapolis sample. This does not preclude, however, the possible existence of individual quadratic trends or of other nonlinear, higher-order trends which could not be assessed due to the limited number of measurements per subject.

Some studies of the pattern of decline on the MMSE have reported nonlinear trends. Haxby et al. (167) found that a bilinear model, consisting of an initial plateau period followed by a decline phase, provided a better fit to MMSE data than the linear model but only for those subjects with isolated memory impairment at study entry. Their analysis, unlike the one performed in this thesis, did not appear to account for the interdependence of repeated scores within subjects. Another approach to the assessment of linearity of decline has been to determine whether the rate of decline on a given scale differs according to baseline severity. Investigations using this approach have yielded inconsistent findings for the MMSE (7,191,197,211,212,217) and for other cognitive and functional scales in general (6,151,152,169,170,188,193,196,217).

Clearly there is a need for additional research to establish - for individual scales - whether the pattern of decline is nonlinear, the nature of the nonlinear trend, and the portion of the scale over which it occurs. It is unclear which, if either, of the two analytic approaches used to date is more appropriate: examining the pattern of change in scores over time through longitudinal within-subject analyses or the annual rate of change of scores as a function of initial severity in cross-sectional between-subject comparisons. Potential limitations of these approaches include regression to the mean effects and, for the latter, the exaggeration of floor effects. The latter problem occurs when several subjects, including those less impaired at entry, reach a scale's floor and where the use of floor data attenuates rate of change estimates. Regardless of the analysis chosen, future studies of the pattern of decline on scales used to monitor progression in AD should ensure that adequate amounts of data are collected over the scale's full range.

Concordance among different methods of estimating annual rate of change

The comparability of ARC estimates of disease progression generated by five different computational methods was examined through separate re-analyses of MMSE data from both cohorts. The within-subject variability of the different ARC estimates was substantial for approximately half of the subjects in both samples. Part of this variability was due to the considerable disparity among estimates produced by the multiple interval method (MIM) alone. This disparity reflects the year-to-year variation in the progression

of individual subjects: it may be an artifact of poor reliability of MMSE estimates of change based on one-year time intervals or the result of real fluctuations in disease course. The fact that the MIM yields more than one ARC estimate per subject is also problematic since the treatment of such estimates as independent observations in conventional statistical analyses violates the assumption of independence. The use of such analyses with MIM ARC estimates, as was the case in three studies using the MIM approach (152,169,170), renders the results uninterpretable.

Pairwise comparisons of the other four methods revealed that ARC estimates generated by the two-point, adjusted two-point, and linear regression methods were comparable for all subjects in both samples. Stern et al. (169) similarly reported no significant difference between mean ARC estimates on the modified Blessed Information-Memory-Concentration scale (168) derived from the two-point and linear regression approaches. These findings do not support the possibility that disparities among these three methods may have contributed meaningfully to inconsistencies in the results of those prognostic studies that used them. The comparability observed between estimates based on only two data points and those based on three or more measurements also fails to support concerns that the former may be less accurate (194). However, the possibility that there may be systematic differences between ARC estimates obtained from these methods in other studies cannot be ruled out.

The similarity between the two-point and adjusted two-point estimates reflects the lack of association between the former and baseline scores in the two samples. This finding should not be interpreted as supporting the general equivalence of these two methods, particularly when the unadjusted estimate is a simple difference score as was the case in the prognostic study that used this adjustment approach (6). Though concern regarding a regression to the mean effect between difference scores and their baseline values is warranted, this adjustment procedure may be extreme since it is based on the assumption that the association between the two is entirely due to this artifact. A more "balanced" approach may be to include baseline severity as a covariate, together with other prognostic factors, in multivariable analyses.

Annual rate of change estimates derived from each of the two-point, adjusted two-point, and linear regression methods showed poorer concordance with those of the trilinear model in both samples. As expected, the latter method tended to estimate faster rates of decline. Substantial differences between individual subjects' trilinear ARC estimate and those of the other methods occurred more frequently in the Minneapolis sample and were associated, in both samples, with trilinear estimates based on relatively small portions of total follow-up. This finding suggests the potential for important differences in individuals' ARC estimates depending on the assumptions made regarding the pattern of measured decline in AD.

The two-point, adjusted two-point, and linear regression methods are based on the assumption that measured decline is linear over the period of follow-up upon which the ARC estimate is based. For most prognostic studies using these methods, this period consisted of a subject's entire follow-up which, for many subjects, began at a mild level of dementia and/or continued to maximum impairment on the measurement scale. In contrast, the trilinear model posits three phases of progression where initial and final phases of relative stability are factored out of the ARC estimation (middle or decline phase). Although the trilinear model was not used in any of the prognostic studies reviewed, it represents a formalization of a strategy used in a number of studies - that of excluding subjects' outcome data in mild and/or severe ranges of a scale when estimating their ARC. The poor concordance observed between estimates from the trilinear model and those of the other three methods suggests, therefore, that excluding ceiling and/or floor data may have a considerable impact on ARC estimates. Furthermore, inconsistencies among prognostic studies in the use of this strategy and in the range of scores excluded may have contributed to discrepancies in their results.

Although the trilinear model may be more accurate than the linear model in describing the pattern of disease progression, its application to the study of prognosis raises some practical and theoretical issues. In prognosis, one would like to predict a patient's progression throughout their disease course. For some patients, however, the trilinear decline period and its corresponding ARC estimate represent only a small portion of their total follow-up. Although the trilinear model requires at least five data points,

its ARC estimate may be based on just two measurements. Research on some scales suggests that ARC estimates are reasonably reliable only for follow-up intervals of greater than one year (160,163). Satisfying both these criteria would limit the application of the trilinear model to those subjects with a minimum of five data points over four or more years of follow-up. Such a requirement may limit both the number and type of patients to whom this model may be applied.

Lastly, the application of the trilinear model produced, for several patients, estimates of "top" and/or "bottom" plateau periods corresponding to moderate severity on the MMSE. The model's fit, in these cases, was not consistent with the rationale for this method - that rate of change may be more accurately estimated by excluding periods of stability in mild and severe disease (floor and ceiling effects). This observation should be considered in future applications of this model.

6.3 Conclusions and recommendations

Existing research on the prognosis of AD is characterized by several methodological limitations many of which can be addressed through more rigorous study design. With respect to overall design, longitudinal studies that prospectively monitor changes in disease status over time are essential to providing an accurate picture of the disease course of individual patients. Even in longitudinal research, however, the accurate estimation of clinical progression in AD remains a major challenge. The difficulty stems from the diversity of signs and symptoms, and from the broad spectrum of decline ranging from subtle cognitive or behavioral deficits to profound brain failure.

Most measurement instruments developed to date assess only one domain of AD symptomatology and none appears to be a sensitive measure of change over the entire disease course. One possible solution is to use a comprehensive battery of different types of measures that collectively capture the complete course of AD, and to combine their scores into a single composite measure (5,262). Galasko et al. have also proposed using clinical milestones as measures of progression including those defined according to performance on outcome instruments (220). Clearly, further research is needed to evaluate the ability of existing and future scales to monitor disease progression over time.

Such research should seek, in particular, to establish their reliability as measures of change and to elucidate the pattern of decline on these instruments. Knowledge of these measurement properties is necessary to ensure the accurate estimation of decline in AD. Further research examining new and existing methods of estimating progression from outcome data is also vital. Such investigations are currently underway (P. Bélisle, L. Joseph, D. Wolfson, X. Zhou, submitted to the Journal of the American Statistical Association).

Other steps that can be taken to improve the quality of research include assembling representative samples of AD patients, following study subjects for periods of greater than one year, and using multivariable statistical techniques that account for varying follow-up intervals, level of impairment and other important covariates. When potential sources of bias, such as selective attrition, cannot be avoided or adjusted for, it is essential that investigators adequately report and analyze the potential impact on study results. Future prognostic studies should also assess the possibility that clinical features may be predictive of decline only at certain stages of the disease by performing subgroup analyses or by modelling the interaction between the presence of symptoms and disease severity or stage at entry. Lastly, efforts should be made to assemble sufficient numbers of subjects so that important prognostic factors are not missed due to inadequate statistical power.

Alzheimer's disease is a devastating disorder afflicting increasing numbers of elderly people. Given the debilitating nature of this disease, the marked variability of its clinical progression, and the lack of a definitive treatment, the elucidation of predictors of decline remains a vital area of research in AD. Despite considerable research, the prognostic significance of most potential predictors investigated to date is inconclusive. Inconsistencies in the literature may be attributable, in part, to the methodological diversity and limitations of the studies. Though many challenges remain in this complex area of research, several measures can be taken to enhance the methodological rigour of future prognostic studies.

REFERENCES

1. Katzman R. The prevalence and malignancy of Alzheimer disease: A major killer. *Arch.Neurol.* 1976; 33:217-218.

2. Ostbye T and Crosse E. Net economic costs of dementia in Canada. *Can.Med.Assoc.J.* 1994; 151:1457-1464.

3. Ernst RL and Hay JW. The U.S. economic and social costs of Alzheimer's disease revisited. *Am.J.Public Health* 1994; 84:1261-1264.

4. Katzman R and Jackson JE. Alzheimer disease: basic and clinical advances. *J.Am.Geriatr.Soc.* 1991; 39:516-525.

5. Galasko D, Corey-Bloom J, and Thal LJ. Monitoring progression in Alzheimer's disease. *J.Am.Geriatr.Soc.* 1991; 39:932-941.

6. Lucca U, Comelli M, Tettamanti M, Tiraboschi P, and Spagnoli A. Rate of progression and prognostic factors in Alzheimer's disease: A prospective study. *J.Am.Geriatr.Soc.* 1993; 41:45-49.

7. Mortimer JA, Ebbitt B, Jun S-P, and Finch MD. Predictors of cognitive and functional progression in patients with probable Alzheimer's disease. *Neurology* 1992; 42:1689-1696.

8. Drachman DA, O'Donnell BF, Lew RA, and Swearer JM. The prognosis in Alzheimer's disease: 'How far' rather than 'how fast' best predicts the course. *Arch.Neurol.* 1990; 47:851-856.

9. McKhann G, Drachman DA, Folstein M, Katzman R, Price D, and Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology* 1984; 34:939-944.

10. Larson EB, Kukull WA, and Katzman R. Cognitive impairment: dementia and Alzheimer's disease. *Annu.Rev.Public Health* 1992; 13:431-449.

11. Tomlinson BE, Blessed G, and Roth M. Observations on the brains of demented old people. *J.Neurol.Sci.* 1970; 11:205-242.

12. Canadian Study of Health and Aging Working Group . Canadian Study of Health and Aging: study methods and prevalence of dementia. *Can.Med.Assoc.J.* 1994; 150(6):899-913.

13. Fratiglioni L, Grut M, Forsell Y, Viitanen M, Grafström M, Holmén K, Ericsson K, Bäckman L, Ahlbom A, and Winblad B. Prevalence of Alzheimer's disease and other dementias in an elderly population: relationship with age, sex, and education. *Neurology* 1991; 41:1886-1892.

14. Kokmen E, Beard CM, O'Brien PC, Offord KP, and Kurland LT. Is the incidence of dementing illness changing? A 25-year time trend study in Rochester, Minnesota (1960-1984). *Neurology* 1993; 43:1887-1892.

15. Bachman DL, Wolf PA, Linn RT, Knoefel JE, Cobb JL, Belanger AJ, D'Agostino RB, and White LR. Prevalence of dementia and probable senile dementia of the Alzheimer type in the Framingham study. *Neurology* 1992; 42:115-119.

16. Rocca WA. Frequency, distribution, and risk factors for Alzheimer's disease. *Nurs.Clin.North Am.* 1994; 29:101-111.

17. Macera CA, Sun RKP, Yeager KK, and Brandes DA. Sensitivity and specificity of death certificate diagnoses for dementing illnesses, 1988-1990. *J.Am.Geriatr.Soc.* 1992; 40:479-481.

18. Raiford K, Anton-Johnson S, Haycox Z, Nolan K, Schaffer A, Caimano C, Fillenbaum G, and Heyman A. CERAD part VII: accuracy of reporting dementia on death certificates of patients with Alzheimer's disease. *Neurology* 1994; 44:2208-2209.

19. Jorm AF, Korten AE, and Henderson AS. The prevalence of dementia: A quantitative integration of the literature. *Acta Psychiatr.Scand.* 1987; 76:465-479.

20. Rockwood K and Stadnyk K. The prevalence of dementia in the elderly: a review. *CanJ.Psychiatry* 1994; 39:253-257.

21. Breteler MMB, Claus JJ, van Duijn CM, Launer LJ, and Hofman A. Epidemiology of Alzheimer's disease. *Epidemiol.Rev.* 1992; 14:59-82.

22. Rocca WA, Hofman A, Brayne C, Breteler MMB, Clarke M, Copeland JRM, Dartigues JF, Engedal K, Hagnell O, Heeren TJ, Jonker C, Lindesay J, Lobo A, Mann AH, Molsa PK, Morgan K, O'Connor DW, da Silva Droux A, Sulkava R, Kay DWK, and Amaducci LA. Frequency and distribution of Alzheimer's disease in Europe: a collaborative study of 1980-1990 prevalence findings. *Ann.Neurol.* 1991; 30:381-390.

23. Kokmen E, Beard CM, Offord KP, and Kurland LT. Prevalence of medically diagnosed dementia in a defined United States population: Rochester, Minnesota, January 1, 1975. *Neurology* 1989; 39:773-776.

24. Pfeffer RI, Afifi AA, and Chance JM. Prevalence of Alzheimer's disease in a retirement community. *Am.J.Epidemiol.* 1987; 125(3):420-436.

25. Evans DA, Funkenstein HH, Albert MS, Scherr PA, Cook NR, Chown MJ, Hebert LE, Hennekens CH, and Taylor JO. Prevalence of Alzheimer's disease in a community population of older persons. *JAMA* 1989; 262:2551-2556.

26. Larson EB. Alzheimer's disease in the community. (Editorial). *JAMA* 1989; 262:2591-2592.

27. Kokmen E, Beard CM, and Kurland LT. The presence of Alzheimer's disease in a community population. (Letter). *JAMA* 1990; 263:2447-2448.

28. Sapir M. Re: "Prevalence of Alzheimer's disease in a retirement community." (Letter). *Am.J.Epidemiol.* 1988; 127:1093-1094.

29. Ebly EM, Parhad IM, Hogan DB, and Fung TS. Prevalence and types of dementia in the very old: results from the Canadian Study of Health and Aging. *Neurology* 1994; 44:1593-1600.

30. Jorm AF. Cross-national comparisons of the occurrence of Alzheimer's and vascular dementias. *Eur.Arch.Psychiatry.Clin.Neurosci.* 1991; 240:218-222.

31. Schoenberg BS, Anderson DW, and Haerer AF. Severe dementia. Prevalence and clinical features in a biracial U.S. population. *Arch.Neurol.* 1985; 42:740-743.

32. Katzman R. Education and the prevalence of dementia and Alzheimer's disease. *Neurology* 1993; 43:13-20.

33. Mortimer JA and Graves AB. Education and other socioeconomic determinants of dementia and Alzheimer's disease. *Neurology* 1993; 43(suppl 4):S39-S44.

34. Rorsman B, Hagnell O, and Lanke J. Prevalence and incidence of senile and multi-infarct dementia in the Lundby Study: a comparison between the time periods 1947-1957 and 1957-1972. *Neuropsychobiology* 1986; 15:122-129.

35. Beard CM, Kokmen E, Offord KP, and Kurland LT. Is the prevalence of dementia changing? *Neurology* 1991; 41:1911-1914.

36. Jorm AF, Korten AE, and Jacomb PA. Projected increases in the number of dementia cases for 29 developed countries: application of a new method for making projections. *Acta Psychiatr.Scand.* 1988; 78:493-500.

37. Kramer MS. *Clinical epidemiology and biostatistics.* Berlin: Springer-Verlag, 1988.

38. Launer LJ and Hofman A. Studies on the incidence of dementia: the European perspective. *Neuroepidemiology* 1992; 11:127-134.

39. Jorm AF. *The epidemiology of Alzheimer's disease and related disorders.* London: Chapman and Hall, 1990.

40. Bachman DL, Wolf PA, Linn RT, Knoefel JE, Cobb JL, Belanger AJ, White LR, and D'Agostino RB. Incidence of dementia and probable Alzheimer's disease in a general population: the Framingham study. *Neurology* 1993; 43:515-519.

41. Drachman DA. If we live long enough, will we all be demented? (Editorial). *Neurology* 1994; 44:1563-1565.

42. Hagnell O, Öjesjö L, and Rorsman B. Incidence of dementia in the Lundby study. *Neuroepidemiology* 1992; 11(suppl 1):61-66.

43. Kokmen E, Chandra V, and Schoenberg BS. Trends in incidence of dementing illness in Rochester, Minnesota, in three quinquennial periods, 1960-1974. *Neurology* 1988; 38:975-980.

44. Schoenberg BS, Kokmen E, and Okazaki H. Alzheimer's disease and other dementing illnesses in a defined United States population: Incidence rates and clinical features. *Ann.Neurol.* 1987; 22:724-729.

45. Launer LJ. Overview of incidence studies of dementia conducted in Europe. *Neuroepidemiology* 1992; 11(suppl 1):2-13.

46. Amaducci LA, Falcini M, and Lippi A. Descriptive epidemiology and risk factors for Alzheimer's disease. *Acta Neurol.Scand.* 1992; Suppl 139:21-25.

47. Fratiglioni L. Epidemiology of Alzheimer's disease: Issues of etiology and validity. *Acta Neurol.Scand.* 1993; 87 (145) Suppl:1-70.

48. van Duijn CM, Stijnen T, and Hofman A. Risk factors for Alzheimer's disease: overview of the EURODEM collaborative re-analysis of case-control studies. *Int.J.Epidemiol.* 1991; 20(suppl 2):S4-S12.

49. Heston LL, Mastri AR, Anderson VE, and White J. Dementia of the Alzheimer type: Clinical genetics, natural history, and associated conditions. *Arch.Gen.Psychiatry* 1981; 38:1085-1090.

50. Breitner JC and Folstein M. Familial Alzheimer dementia: a prevalent disorder with specific clinical features. *Psychol.Med.* 1984; 14:63-80.

51. Mendez MF, Underwood KL, Zander BA, Mastri AR, Sung JH, and Frey WH. Risk factors in Alzheimer's disease: a clinicopathologic study. *Neurology* 1992; 42:770-775.

52. Canadian Study of Health and Aging Working Group . The Canadian Study of Health and Aging: risk factors for Alzheimer's disease in Canada. *Neurology* 1994; 44:2073-2080.

53. Fratiglioni L, Ahlbom A, Viitanen M, and Winblad B. Risk factors for late-onset Alzheimer's disease: a population-based, case-control study. *Ann.Neurol.* 1993; 33:258-266.

54. van Duijn CM, Clayton D, Chandra V, Fratiglioni L, Graves AB, Heyman A, Jorm AF, Kokmen E, Kondo K, Mortimer JA, Rocca WA, Shalat SL, Soininen H, Hofman A, and for the EURODEM Risk Factors Research Group . Familial aggregation of Alzheimer's disease and related disorders: a collaborative re-analysis of case-control studies. *Int.J.Epidemiol.* 1991; 20(suppl 2):S13-S20.

55. Heyman A, Wilkinson WE, Stafford JA, Helms MJ, Sigmon AH, and Weinberg T. Alzheimer's disease: A study of epidemiological aspects. *Ann.Neurol.* 1984; 15:335-341.

56. Amaducci LA, Fratiglioni L, Rocca WA, Fieschi C, Livrea P, Pedone D, Bracco L, Lippi A, Gandolfo C, Bino G, Prencipe M, Bonatti ML, Girotti F, Carella F, Tavolato B, Ferla S, Lenzi GL, Carolei A, Gambi A, Grigoletto F, and Schoenberg BS. Risk factors for clinically diagnosed Alzheimer's disease: a case-control study of an Italian population. *Neurology* 1986; 36:922-931.

57. Clark RF and Goate AM. Molecular genetics of Alzheimer's disease. *Arch.Neurol.* 1993; 50:1164-1172.

58. Saunders AM, Strittmatter WJ, Schmechel D, St.George-Hyslop PH, Pericak-Vance MA, Joo SH, Rosi BL, Gusella JF, Crapper MacLachlan DR, Alberts MJ, Hulette C, Crain B, Goldgaber D, and Roses AD. Association of apolipoprotein E allele e4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 1993; 43:1467-1472.

59. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, and 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.

60. Rocca WA, van Duijn CM, Clayton D, Chandra V, Fratiglioni L, Graves AB, Heyman A, Jorm AF, Kokmen E, Kondo K, Mortimer JA, Shalat SL, Soininen H, Hofman A, and for the EURODEM Risk Factors Research Group . Maternal age and Alzheimer's disease: a collaborative re-analysis of case-control studies. *Int.J.Epidemiol.* 1991; 20(suppl 2):S21-S27.

61. Mortimer JA. Epidemiology of dementia: cross-cultural comparisons. *Adv.Neurol.* 1990; 51:27-33.

62. Mortimer JA, van Duijn CM, Chandra V, Fratiglioni L, Graves AB, Heyman A, Jorm AF, Kokmen E, Kondo K, Rocca WA, Shalat SL, Soininen H, Hofman A, and for the EURODEM Risk Factors Research Group . Head trauma as a risk factor for Alzheimer's disease: a collaborative re-analysis of case-control studies. *Int.J.Epidemiol.* 1991; 20(suppl 2):S28-S35.

63. Graves AB, van Duijn CM, Chandra V, Fratiglioni L, Heyman A, Jorm AF, Kokmen E, Kondo K, Mortimer JA, Rocca WA, Shalat SL, Soininen H, Hofman A, and for the EURODEM Risk Factors Research Group . Alcohol and tobacco consumption as risk factors for Alzheimer's disease: a collaborative re-analysis of case-control studies. *Int.J.Epidemiol.* 1991; 20(suppl 2):S48-S57.

64. Stern Y, Gurland B, Tatemichi TK, Tang MX, Wilder D, and Mayeux R. Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA* 1994; 271:1004-1010.

65. Beard CM, Kokmen E, Offord KP, and Kurland LT. Lack of association between Alzheimer's disease and education, occupation, marital status, or living arrangement. *Neurology* 1992; 42:2063-2068.

66. Jorm AF, van Duijn CM, Chandra V, Fratiglioni L, Graves AB, Heyman A, Kokmen E, Kondo K, Mortimer JA, Rocca WA, Shalat SL, Soininen H, Hofman A, and for the EURODEM Risk Factors Research Group . Psychiatric history and related exposures as risk factors for Alzheimer's disease: a collaborative re-analysis of case-control studies. *Int.J.Epidemiol.* 1991; 20(suppl 2):S43-S47.

67. Breteler MMB, van Duijn CM, Chandra V, Fratiglioni L, Graves AB, Heyman A, Jorm AF, Kokmen E, Kondo K, Mortimer JA, Rocca WA, Shalat SL, Soininen H, Hofman A, and for the EURODEM Risk Factors Research Group . Medical history and the risk of Alzheimer's disease: a collaborative re-analysis of case-control studies. *Int.J.Epidemiol.* 1991; 20(suppl 2):S36-S42.

68. Gatz M, Lowe B, Berg S, Mortimer JA, and Pedersen N. Dementia: not just a search for the gene. *Gerontologist* 1994; 34:251-255.

69. Cotton P. Constellation of risks and processes seen in search for Alzheimer's clues. *JAMA* 1994; 271:89-91.

70. Hollander E, Mohs RC, and Davis KL. Antemortem markers of Alzheimer's disease. *Neurobiol.Aging* 1986; 7:367-387.

71. Katzman R. Alzheimer's disease. *N.Engl.J.Med.* 1986; 314(15):964-973.

72. Khachaturian ZS. Diagnosis of Alzheimer's disease. *Arch.Neurol.* 1985; 42:1097-1105.

73. Blessed G, Tomlinson BE, and Roth M. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *Br.J.Psychiatry* 1968; 114:797-811.

74. Crystal H, Dickson D, Fuld P, Masur D, Scott R, Mehler H, Masdeu J, Kawas C, Aronson H, and Wolfson L. Clinico-pathological studies in dementia: nondemented subjects with pathologically confirmed Alzheimer's disease. *Neurology* 1988; 38:1682-1687.

75. American Psychiatric Association . *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R), 3rd ed., revised.* Washington, DC: American Psychiatric Association, 1987.

76. Canadian Consensus Conference on the Assessment of Dementia. Conference report. *Can.Med.Assoc.J.* 1991; 144(Suppl):

77. American Psychiatric Association . *Diagnostic and Statistical Manual of Mental Disorders (DSM-III), 3rd ed.* Washington, DC: American Psychiatric Association, 1980.

78. Lopez OL, Swihart AA, Becker JT, Reinmuth OM, Reynolds CF, Rezek DL, and Daly FL. Reliability of NINCDS-ADRDA clinical criteria for the diagnosis of Alzheimer's disease. *Neurology* 1990; 40:1517-1522.

79. Baldereschi M, Amato MP, Nencini P, Pracucci G, Lippi A, Amaducci LA, Gauthier S, Beatty L, Quiroga P, Klassen G, Galea A, Muscat P, Osuntokun B, Ogunniyi A, Potera-Sanchez A, Bermejo F, Hendrie H, Burdine V, Brashear A, Farlow M, Maggi S, and Katzman R. Cross-national interrater agreement on the clinical diagnostic criteria for dementia. *Neurology* 1994; 44:239-242.

80. Kukull WA, Larson EB, Reifler BV, Lampe TH, Yerby MS, and Hughes JP. Interrater reliability of Alzheimer's disease diagnosis. *Neurology* 1990; 40:257-260.

81. Kukull WA, Larson EB, Reifler BV, Lampe TH, Yerby MS, and Hughes JP. The validity of 3 clinical diagnostic criteria for Alzheimer's disease. *Neurology* 1990; 40:1364-1369.

82. Boller F, Lopez OL, and Moossy J. Diagnosis of dementia: Clinicopathologic correlations. *Neurology* 1989; 39:76-79.

83. Tierney MC, Fisher RH, Lewis AJ, Zorzitto ML, Snow WG, Reid DW, and Nieuwstraten P. The NINCDS-ADRDA Work Group criteria for the clinical diagnosis of probable Alzheimer's disease: A clinicopathologic study of 57 cases. *Neurology* 1988; 38:359-364.

84. Martin EM, Wilson RS, Penn RD, Fox JH, Clasen RA, and Savoy SM. Cortical biopsy results in Alzheimer's disease: Correlation with cognitive deficits. *Neurology* 1987; 37:1201-1204.

85. Morris JC, McKeel DW, Fulling K, Torack RM, and Berg L. Validation of clinical diagnostic criteria for Alzheimer's disease. *Ann.Neurol.* 1988; 24:17-22.

86. Risse SC, Raskind MA, Nochlin D, Sumi SM, Lampe TH, Bird TD, Cubberley L, and Peskind ER. Neuropathological findings in patients with clinical diagnoses of probable Alzheimer's disease. *Am.J.Psychiatry* 1990; 147:168-172.

87. Burns A, Luthert P, Levy R, Jacoby R, and Lantos P. Accuracy of clinical diagnosis of Alzheimer's disease. *Br.Med.J.* 1990; 301:1026.

88. Gottlieb GL and Kumar A. Conventional pharmacologic treatment for patients with Alzheimer's disease. *Neurology* 1993; 43(suppl 4):S56-S63.

89. Schneider LS and Tariot PN. Emerging drugs for Alzheimer's disease. *Med.Clin.North Am.* 1994; 78:911-934.

90. Davis RE, Emmerling MR, Jaen JC, Moos WH, and Spiegel K. Therapeutic intervention in dementia. *Crit.Rev.Neurobiol.* 1993; 7(1):41-83.

91. Forette F, Panisset M, and Boller F. Clinical trials in cognitive impairment in the elderly. *Aging Clin.Exp.Res.* 1992; 4:239-250.

92. Ganesan R, Standish T, Molloy DW, Darzins P, and Orange JB. New treatments for dementia. Myth, magic and science. *Can.Fam.Physician* 1994; 40:1149-1157.

93. Hodes RJ. Alzheimer's disease: treatment research finds new targets. *J.Am.Geriatr.Soc.* 1994; 42:679-681.

94. Harrell LE. Alzheimer's disease. *South.Med.J.* 1991; 84(Suppl 1):32-40.

95. Bartus RT, Dean RL,III, Beer B, and Lippa AS. The cholinergic hypothesis of geriatric memory dysfunction. *Science* 1982; 217:408-417.

96. Rossor M. Alzheimer's disease. *Br.Med.J.* 1993; 307:779-782.

97. Davis KL, Thal LJ, Gamzu ER, Davis CS, Woolson RF, Gracon SI, Drachman DA, Schneider LS, Whitehouse PJ, Hoover TM, Morris JC, Kawas CH, Knopman DS, Earl NL, Kumar V, and Doody RS. A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease. *N.Engl.J.Med.* 1992; 327(18):1253-1259.

98. Farlow M, Gracon SI, Hershey LA, Lewis KW, Sadowsky CH, Dolan-Ureno J, and for the Tacrine Study Group . A controlled trial of Tacrine in Alzheimer's disease. *JAMA* 1992; 268:2523-2529.

99. Davis KL and Haroutunian V. Strategies for the treatment of Alzheimer's disease. *Neurology* 1993; 43(suppl 4):S52-S55.

100. Roth M. Advances in Alzheimer's disease. A review for the family physician. *J.Fam.Pract.* 1993; 37:593-607.

101. Petry S, Cummings JL, Hill MA, and Shapira J. Personality alterations in dementia of the Alzheimer type: a three-year follow-up study. *J.Geriatr.Psychiatry Neurol.* 1989; 4:203-207.

102. Persson G and Skoog I. Subclinical dementia: relevance of cognitive symptoms and signs. *J.Geriatr.Psychiatry Neurol.* 1992; 5:172-178.

103. Reisberg B, Ferris SH, Torossian C, Kluger A, and Monteiro I. Pharmacologic treatment of Alzheimer's disease: a methodologic critique based upon current knowledge of symptomatology and relevance for drug trials. *Int.Psychogeriatr.* 1992; 4(Suppl 1):9-42.

104. Terry RD, Masliah E, Salmon DP, Butters N, DeTeresa R, Hill R, Hansen LA, and Katzman R. Physical basis of cognitive alterations in Alzheimer's disease: Synapse loss is the major correlate of cognitive impairment. *Ann.Neurol.* 1991; 30:572-580.

105. Folstein M and Whitehouse PJ. Cognitive impairment of Alzheimer disease. *Neurobehav.Toxicol.Teratol.* 1983; 5:631-634.

106. Rubin EH. Psychopathology of senile dementia of the Alzheimer type. *Adv.Neurol.* 1990; 51:53-59.

107. Reisberg B, Borenstein J, Salob SP, Ferris SH, Franssen E, and Georgotas A. Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. *J.Clin.Psychiatry* 1987; 48(suppl 5):9-15.

108. Baker FM, Kokmen E, Chandra V, and Schoenberg BS. Psychiatric symptoms in cases of clinically diagnosed Alzheimer's disease. *J.Geriatr.Psychiatry Neurol.* 1991; 4:71-78.

109. Zubenko GS and Moossy J. Major depression in primary dementia: Clinical and neuropathologic correlates. *Arch.Neurol.* 1988; 45:1182-1186.

110. Forstl H, Burns A, Levy R, Cairns N, Luthert P, and Lantos P. Neurologic signs in Alzheimer's disease. Results of a prospective clinical and neuropathologic study. *Arch.Neurol.* 1992; 49:1038-1042.

111. Deutsch LH and Rovner B. Agitation and other noncognitive abnormalities in Alzheimer's disease. *Psychiatr.Clin.North Am.* 1991; 14(2):341-351.

112. Swearer JM, Drachman DA, O'Donnell BF, and Mitchell AL. Troublesome and disruptive behaviors in dementia. Relationships to diagnosis and disease severity. *J.Am.Geriatr.Soc.* 1988; 36:784-790.

113. Merriam AE, Aronson M, Gaston P, Wey SL, and Katz I. The psychiatric symptoms of Alzheimer's disease. *J.Am.Geriatr.Soc.* 1988; 36:7-12.

114. Cummings JL, Miller B, Hill MA, and Neshkes R. Neuropsychiatric aspects of multi-infarct dementia and dementia of the Alzheimer type. *Arch.Neurol.* 1987; 44:389-393.

115. Miller TP, Tinklenberg J, Brooks JO,III, Fenn HH, and Yesavage JA. Selected psychiatric symptoms associated with rate of cognitive decline in patients with Alzheimer's disease. *J.Geriatr.Psychiatry Neurol.* 1993; 6:235-238.

116. Corey-Bloom J, Galasko D, Hofstetter R, Jackson JE, and Thal LJ. Clinical features distinguishing large cohorts with possible AD, probable AD, and mixed dementia. *J.Am.Geriatr.Soc.* 1993; 41:31-37.

117. Richards M, Bell K, Dooneief G, Marder K, Sano M, Mayeux R, and Stern Y. Patterns of neuropsychological performance in Alzheimer's disease patients with and without extrapyramidal signs. *Neurology* 1993; 43:1708-1711.

118. Reisberg B, Ferris SH, de Leon MJ, and Crook T. The global deterioration scale for assessment of primary degenerative dementia. *Am.J.Psychiatry* 1982; 139:1136-1139.

119. Reisberg B, Ferris SH, de Leon MJ, and Crook T. Global Deterioration Scale (GDS). *Psychopharmacol.Bull.* 1988; 24(4):661-663.

120. Hughes CP, Berg L, Danzinger WL, Coben LA, and Martin RL. A new clinical scale for the staging of dementia. *Br.J.Psychiatry* 1982; 140:566-572.

121. Berg L. Clinical Dementia Rating (CDR). *Psychopharmacol.Bull.* 1988; 24(4):637-639.

122. Overall JE, Scott J, Rhoades HM, and Lesser J. Empirical scaling of the stages of cognitive decline in senile dementia. *J.Geriatr.Psychiatry Neurol.* 1990; 3:212-220.

123. Fabiszewski KJ. Alzheimer's disease: overview and progression. *J.Adv.Med.Surg.Nurs.* 1989; 1(2):1-17.

124. Spooner MA. Is it really Alzheimer's? *Can.Fam.Physician* 1994; 40:1141-1145.

125. Lee VK. Language changes and Alzheimer's disease: a literature review. *J.Gerontol.Nurs.* 1991; 17(1):16-20.

126. Chobor KL and Brown JW. Semantic deterioration in Alzheimer's: the patterns to expect. *Geriatrics* 1990; 45(10):68-75.

127. Burns A, Jacoby R, and Levy R. Behavioral abnormalities and psychiatric symptoms in Alzheimer's disease: preliminary findings. *Int.Psychogeriatr.* 1990; 2(1):25-36.

128. Molsa PK, Marttila RJ, and Rinne UK. Survival and cause of death in Alzheimer's disease and multi-infarct dementia. *Acta Neurol.Scand.* 1986; 74:103-107.

129. Treves TA, Korczyn AD, Zilber N, Kahana E, Liebowitz Y, Alter M, and Schoenberg BS. Presenile dementia in Israel. *Arch.Neurol.* 1986; 43:26-29.

130. Kukull WA, Brenner DE, Speck CE, Nochlin D, Bowen J, McCormick W, Teri L, Pfanschmidt ML, and Larson EB. Causes of death associated with Alzheimer disease: variation by level of cognitive impairment before death. *J.Am.Geriatr.Soc.* 1994; 42:723-726.

131. Walsh JS, Welch HG, and Larson EB. Survival of outpatients with Alzheimer-type dementia. *Ann.Intern.Med.* 1990; 113:429-434.

132. Beard CM, Kokmen E, O'Brien PC, and Kurland LT. Are patients with Alzheimer's disease surviving longer in recent years? *Neurology* 1994; 44:1869-1871.

133. van Dijk PTM, Dippel DWJ, and Habbema JDF. Survival of patients with dementia. *J.Am.Geriatr.Soc.* 1991; 39:603-610.

134. Heyman A, Wilkinson WE, Hurwitz BJ, Helms MJ, Haynes CS, Utley CM, and Gwyther LP. Early-onset Alzheimer's disease: Clinical predictors of institutionalization and death. *Neurology* 1987; 37:980-984.

135. Barclay LL, Zemcov A, Blass JP, and Sansone J. Survival in Alzheimer's disease and vascular dementia. *Neurology* 1985; 35:834-840.

136. Stern Y, Folstein M, Albert M, Richards M, Miller L, Bylsma F, Lafleche G, Marder K, Bell K, Sano M, Devanand D, Loreck D, Wootten J, and Bello J. Multicenter study of predictors of disease course in Alzheimer disease (the "predictors study"). I. Study design, cohort description, and intersite comparisons. *Alzheimer Dis.Assoc.Disord.* 1993; 7(1):3-21.

137. Burns A, Jacoby R, and Levy R. Neurological signs in Alzheimer's disease. *Age Ageing* 1991; 20:45-51.

138. Bakchine S, Lacomblez L, Palisson E, Laurent M, and Derouesne C. Relationship between primitive reflexes, extra-pyramidal signs, reflective apraxia and severity of cognitive impairment in dementia of the Alzheimer type. *Acta Neurol.Scand.* 1989; 79:38-46.

139. Huff FJ and Growdon JH. Neurological abnormalities associated with severity of dementia in Alzheimer's disease. *Can.J.Neurol.Sci.* 1986; 13:403-405.

140. Chui HC, Teng EL, Henderson VW, and Moy AC. Clinical subtypes of dementia of the Alzheimer type. *Neurology* 1985; 35:1544-1550.

141. Cooper JK, Mungas D, and Weiler PG. Relation of cognitive status and abnormal behaviors in Alzheimer's disease. *J.Am.Geriatr.Soc.* 1990; 38:867-870.

142. Teri L, Larson EB, and Reifler BV. Behavioral disturbance in dementia of the Alzheimer's type. *J.Am.Geriatr.Soc.* 1988; 36:1-6.

143. Fitz AG and Teri L. Depression, cognition, and functional ability in patients with Alzheimer's disease. *J.Am.Geriatr.Soc.* 1994; 42:186-191.

144. Pearson JL, Teri L, Reifler BV, and Raskind MA. Functional status and cognitive impairment in Alzheimer's patients with and without depression. *J.Am.Geriatr.Soc.* 1989; 37:1117-1121.

145. Jeste DV, Wragg RE, Salmon DP, Harris MJ, and Thal LJ. Cognitive deficits of patients with Alzheimer's disease with and without delusions. *Am.J.Psychiatry* 1992; 149:184-189.

146. Binetti G, Bianchetti A, Padovani A, Lenzi G, De Leo D, and Trabucchi M. Delusions in Alzheimer's disease and multi-infarct dementia. *Acta Neurol.Scand.* 1993; 88:5-9.

147. Teri L, Borson S, Kiyak A, and Yamagishi M. Behavioral disturbance, cognitive dysfunction, and functional skill: Prevalence and relationship in Alzheimer's disease. *J.Am.Geriatr.Soc.* 1989; 37:109-116.

148. Bozzola FG, Gorelick PB, and Freels S. Personality changes in Alzheimer's disease. *Arch.Neurol.* 1992; 49:297-300.

149. Kurita A, Blass JP, Nolan KA, Black RS, and Thaler HT. Relationship between cognitive status and behavioral symptoms in Alzheimer's disease and mixed dementia. *J.Am.Geriatr.Soc.* 1993; 41:732-736.

150. Mielke R, Herholz K, Grond M, Kessler J, and Heiss WD. Clinical deterioration in probable Alzheimer's disease correlates with progressive metabolic impairment of association areas. *Dementia* 1994; 5:36-41.

151. Katzman R, Brown T, Thal LJ, Fuld PA, Aronson M, Butters N, Klauber MR, Wiederholt W, Pay M, Renbing X, Ooi WL, Hofstetter R, and Terry RD. Comparison of rate of annual change of mental status score in four independent studies of patients with Alzheimer's disease. *Ann.Neurol.* 1988; 24:384-389.

152. Green CR, Mohs RC, Schmeidler J, Aryan M, and Davis KL. Functional decline in Alzheimer's disease: A longitudinal study. *J.Am.Geriatr.Soc.* 1993; 41:654-661.

153. Botwinick J, Storandt M, and Berg L. A longitudinal, behavioral study of senile dementia of the Alzheimer type. *Arch.Neurol.* 1986; 43:1124-1127.

154. Jorm AF. Subtypes of Alzheimer's dementia: a conceptual analysis and critical review. *Psychol.Med.* 1985; 15:543-553.

155. Friedland RP, Koss E, Haxby JV, Grady CL, Luxenberg J, and Schapiro MB. NIH conference. Alzheimer disease: clinical and biological heterogeneity. *Ann.Intern.Med.* 1988; 109:298-311.

156. Ritchie K and Touchon J. Heterogeneity in senile dementia of the Alzheimer type: individual differences, progressive deterioration or clinical sub-types? *J.Clin.Epidemiol.* 1992; 45(12):1391-1398.

157. Chui HC. The significance of clinically defined subgroups of Alzheimer's disease. *J.Neural.Transm.* 1987; 24(Suppl):57-68.

158. Berg L, Smith DS, Morris JC, Miller JP, Rubin EH, Storandt M, and Coben LA. Mild senile dementia of the Alzheimer type: 3. Longitudinal and cross-sectional assessment. *Ann.Neurol.* 1990; 28:648-652.

159. Farrer LA, Cupples LA, van Duijn CM, Connor-Lacke L, Kiely DK, and Growdon JH. Rate of progression of Alzheimer's disease is associated with genetic risk. *Arch.Neurol.* 1995; 52:918-923.

160. Morris JC, Edland S, Clark C, Galasko D, Koss E, Mohs RC, van Belle G, Fillenbaum GG, and Heyman A. The consortium to establish a registry for Alzheimer's disease (CERAD). Part IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease. *Neurology* 1993; 43:2457-2465.

161. Department of Clinical Epidemiology and Biostatistics McMaster University Health Sciences Centre. How to read clinical journals: III. To learn the clinical course and prognosis of disease. *Can.Med.Assoc.J.* 1981; 124:869-872.

162. Berg L. Mild senile dementia of the Alzheimer type: diagnostic criteria and natural history. *Mt.Sinai J.Med.* 1988; 55(1):87-96.

163. van Belle G, Uhlmann RF, Hughes JP, and Larson EB. Reliability estimates of changes in mental status test performance in senile dementia of the Alzheimer type. *J.Clin.Epidemiol.* 1990; 43(6):589-595.

164. Folstein M, Folstein SE, and McHugh PR. "Mini-Mental State": A pratical method for grading the cognitive state of patients for the clinician. *J.Psychiatr.Res.* 1975; 12:189-198.

165. Schlesselman JJ. Planning a longitudinal study: II. Frequency of measurement and study duration. *J.Chronic Dis.* 1973; 26:561-570.

166. Becker JT, Huff FJ, Nebes RD, Holland A, and Boller F. Neuropsychological function in Alzheimer's disease. Pattern of impairment and rates of progression. *Arch.Neurol.* 1988; 45:263-268.

167. Haxby JV, Raffaele K, Gillette J, Schapiro MB, and Rapoport SI. Individual trajectories of cognitive decline in patients with dementia of the Alzheimer type. *J.Clin.Exp.Neuropsychol.* 1992; 14(4):575-592.

168. Fuld PA. Psychological testing in the differential diagnosis of the dementias. In: *Alzheimer's Disease: Senile Dementia and Related Disorders, Aging, vol 7.* Eds. Katzman R, Terry RD, and Bick KL. New York: Raven Press, 1978.

169. Stern RG, Mohs RC, Bierer LM, Silverman JM, Schmeidler J, Davidson M, and Davis KL. Deterioration on the Blessed Test in Alzheimer's disease: Longitudinal data and their implications for clinical trials and identification of subtypes. *Psychiatry Res.* 1992; 42:101-110.

170. Stern RG, Mohs RC, Davidson M, Schmeidler J, Silverman JM, Kramer-Ginsberg E, Searcey T, Bierer LM, and Davis KL. A longitudinal study of Alzheimer's disease: Measurement, rate, and predictors of cognitive deterioration. *Am.J.Psychiatry* 1994; 151:390-396.

171. Helmes E, Merskey H, Fox H, Fry RN, Bowler JV, and Hachinski VC. Patterns of deterioration in senile dementia of the Alzheimer type. *Arch.Neurol.* 1995; 52:306-310.

172. Huff FJ, Growdon JH, Corkin S, and Rosen TJ. Age at onset and rate of progression of Alzheimer's disease. *J.Am.Geriatr.Soc.* 1987; 35:27-30.

173. Soininen H, Helkala EL, Laulumaa V, Soikkeli R, Hartikainen P, and Riekkinen PJ. Cognitive profile of Alzheimer patients with extrapyramidal signs: a longitudinal study. *J.Neural.Transm.* 1992; 4:241-254.

174. Grady CL, Haxby JV, Horwitz B, Berg G, and Rapoport SI. Neuropsychological and cerebral metabolic function in early vs late onset dementia of the Alzheimer type. *Neuropsychologia* 1987; 25(5):807-816.

175. Miller TP, Tinklenberg J, Brooks JO,III, and Yesavage JA. Cognitive decline in patients with Alzheimer's disease: Differences in patients with and without extrapyramidal signs. *Alzheimer Dis.Assoc.Disord.* 1991; 5(4):251-256.

176. Uhlmann RF, Larson EB, and Koepsell TD. Hearing impairment and cognitive decline in senile dementia of the Alzheimer's type. *J.Am.Geriatr.Soc.* 1986; 34:207-210.

177. Reifler BV, Larson EB, Teri L, and Poulsen M. Dementia of the Alzheimer's type and depression. *J.Am.Geriatr.Soc.* 1986; 34:855-859.

178. Lopez OL, Becker JT, Brenner RP, Rosen J, Bajulaiye OI, and Reynolds CF. Alzheimer's disease with delusions and hallucinations: neuropsychological and electroencephalographic correlates. *Neurology* 1991; 41:906-912.

179. Lopez OL, Boller F, Becker JT, Miller M, and Reynolds CF,III. Alzheimer's disease and depression: Neuropsychological impairment and progression of the illness. *Am.J.Psychiatry* 1990; 147:855-860.

180. Morris JC, Drazner M, Fulling K, Grant EA, and Goldring J. Clinical and pathological aspects of parkinsonism in Alzheimer's disease. A role for extranigral factors? *Arch.Neurol.* 1989; 46:651-657.

181. Drevets WC and Rubin EH. Psychotic symptoms and the longitudinal course of senile dementia of the Alzheimer type. *Biol.Psychiatry* 1989; 25:39-48.

182. Faber-Langendoen K, Morris JC, Knesevich JW, LaBarge E, Miller JP, and Berg L. Aphasia in senile dementia of the Alzheimer type. *Ann.Neurol.* 1988; 23:365-370.

183. Knesevich JW, Toro FR, Morris JC, and LaBarge E. Aphasia, family history, and the longitudinal course of senile dementia of the Alzheimer type. *Psychiatry Res.* 1985; 14:255-263.

184. Basun H, Grut M, Winblad B, and Lannfelt L. Apolipoprotein $\in 4$ allele and disease progression in patients with late-onset Alzheimer's disease. *Neurosci.Lett.* 1995; 183:32-34.

185. Flicker C, Ferris SH, and Reisberg B. A two-year longitudinal study of cognitive function in normal aging and Alzheimer's disease. *J.Geriatr.Psychiatry Neurol.* 1993; 6:84-96.

186. Teri L, McCurry SM, Edland S, Kukull WA, and Larson EB. Cognitive decline in Alzheimer's disease: A longitudinal investigation of risk factors for accelerated decline. *J.Gerontol.* 1995; 50A:M49-M55.

187. Teri L, Hughes JP, and Larson EB. Cognitive deterioration in Alzheimer's disease: Behavioral and health factors. *J.Gerontol.* 1990; 45(2):P58-P63.

188. Huff FJ, Belle SH, Shim YK, Ganguli M, and Boller F. Prevalence and prognostic value of neurologic abnormalities in Alzheimer's disease. *Dementia* 1990; 1:32-40.

189. Seltzer B and Sherwin I. A comparison of clinical features in early- and late-onset primary degenerative dementia. One entity or two? *Arch.Neurol.* 1983; 40:143-146.

190. Kraemer HC, Tinklenberg J, and Yesavage JA. 'How far' vs 'how fast' in Alzheimer's disease: The question revisited. *Arch.Neurol.* 1994; 51:275-279.

191. Hogan DB, Thierer DE, Ebly EM, and Parhad IM. Progression and outcome of patients in a Canadian dementia clinic. *Can.J.Neurol.Sci.* 1994; 21:331-338.

192. Stern Y, Albert M, Brandt J, Jacobs DM, Tang MX, Marder K, Bell K, Sano M, Devanand D, Bylsma F, and Lafleche G. Utility of extrapyramidal signs and psychosis as predictors of cognitive and functional decline, nursing home admission, and death in Alzheimer's disease: Prospective analyses from the Predictors Study. *Neurology* 1994; 44:2300-2307.

193. Jacobs DM, Sano M, Marder K, Bell K, Bylsma F, Lafleche G, Albert M, Brandt J, and Stern Y. Age at onset of Alzheimer's disease: Relation to pattern of cognitive dysfunction and rate of decline. *Neurology* 1994; 44:1215-1220.

194. Ortof E and Crystal HA. Rate of progression of Alzheimer's disease. *J.Am.Geriatr.Soc.* 1989; 37:511-514.

195. Bracco L, Gallato R, Grigoletto F, Lippi A, Lepore V, Bino G, Lazzaro MP, Carella F, Piccolo T, Pozzilli C, Giometto B, and Amaducci LA. Factors affecting course and survival in Alzheimer's disease. A 9-year longitudinal study. *Arch.Neurol.* 1994; 51:1213-1219.

196. Thal LJ, Grundman M, and Klauber MR. Dementia: Characteristics of a referral population and factors associated with progression. *Neurology* 1988; 38:1083-1090.

197. Burns A, Jacoby R, and Levy R. Progression of cognitive impairment in Alzheimer's disease. *J.Am.Geriatr.Soc.* 1991; 39:39-45.

198. Barclay LL, Zemcov A, Blass JP, and McDowell FH. Factors associated with duration of survival in Alzheimer's disease. *Biol.Psychiatry* 1985; 20:86-93.

199. Burns A, Lewis G, Jacoby R, and Levy R. Factors affecting survival in Alzheimer's disease. *Psychol.Med.* 1991; 21:363-370.

200. Molsa PK, Marttila RJ, and Rinne UK. Long-term survival and predictors of mortality in Alzheimer's disease and multi-infarct dementia. *Acta Neurol.Scand.* 1995; 91:159-164.

201. Jagger C, Clarke M, and Stone A. Predictors of survival with Alzheimer's disease: a community-based study. *Psychol.Med.* 1995; 25:171-177.

202. Anthony JC, Le Resche L, Niaz U, Von Korff MR, and Folstein M. Limits of the 'Mini-Mental State' as a screening test for dementia and delirium among hospital patients. *Psychol.Med.* 1982; 12:397-408.

203. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J.Am.Geriatr.Soc.* 1975; 23:433-441.

204. Stern Y, Hesdorffer D, Sano M, and Mayeux R. Measurement and prediction of functional capacity in Alzheimer's disease. *Neurology* 1990; 40:8-14.

205. Silverman JM, Breitner JC, Mohs RC, and Davis KL. Reliability of the family history method in genetic studies of Alzheimer's disease and related dementias. *Am.J.Psychiatry* 1986; 143:1279-1282.

206. Chen JY, Stern Y, Sano M, and Mayeux R. Cumulative risks of developing extrapyramidal signs, psychosis, or myoclonus in the course of Alzheimer's disease. *Arch.Neurol.* 1991; 48:1141-1143.

207. Chui HC, Lyness SA, Sobel E, and Schneider LS. Extrapyramidal signs and psychiatric symptoms predict faster cognitive decline in Alzheimer's disease. *Arch.Neurol.* 1994; 51:676-681.

208. Stern Y, Mayeux R, Sano M, Hauser WA, and Bush T. Predictors of disease course in patients with probable Alzheimer's disease. *Neurology* 1987; 37:1649-1653.

209. Kertesz A, Appell J, and Fisman M. The dissolution of language in Alzheimer's disease. *Can.J.Neurol.Sci.* 1986; 13:415-418.

210. Berg L, Danzinger WL, Storandt M, Coben LA, Gado M, Hughes CP, Knesevich JW, and Botwinick J. Predictive features in mild senile dementia of the Alzheimer type. *Neurology* 1984; 34:563-569.

211. Boller F, Becker JT, Holland A, Forbes MM, Hood PC, and McGonigle-Gibson KL. Predictors of decline in Alzheimer's disease. *Cortex* 1991; 27:9-17.

212. Goldblum MC, Tzortzis C, Michot JL, Panisset M, and Boller F. Language impairement and rate of cognitive decline in Alzheimer's disease. *Dementia* 1994; 5:334-338.

213. Yesavage JA, Brooks JO,III, Taylor J, and Tinklenberg J. Development of aphasia, apraxia, and agnosia and decline in Alzheimer's disease. *Am.J.Psychiatry* 1993; 150:742-747.

214. Knesevich JW, LaBarge E, and Edwards DF. Predictive value of the Boston Naming Test in mild senile dementia of the Alzheimer type. *Psychiatry Res.* 1986; 19:155-161.

215. Rosen J and Zubenko GS. Emergence of psychosis and depression in the longitudinal evaluation of Alzheimer's disease. *Biol.Psychiatry* 1991; 29:224-232.

216. Forstl H, Besthorn C, Geiger-Kabisch C, Sattel H, and Schreiter-Gasser U. Psychotic features and the course of Alzheimer's disease: relationship to cognitive, electroencephalographic and computerized tomography findings. *Acta Psychiatr.Scand.* 1993; 87:395-399.

217. Salmon DP, Thal LJ, Butters N, and Heindel WC. Longitudinal evaluation of dementia of the Alzheimer type: A comparison of three standardized mental status examinations. *Neurology* 1990; 40:1225-1230.

218. Berg G, Edwards DF, Danzinger WL, and Berg L. Longitudinal change in three brief assessments of SDAT. *J.Am.Geriatr.Soc.* 1987; 35:205-212.

219. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993; 43(11):2412-2414.

220. Galasko D, Edland S, Morris JC, Clark C, Mohs RC, and Koss E. The consortium to establish a registry for Alzheimer's disease (CERAD). Part XI. Clinical milestones in patients with Alzheimer's disease followed over 3 years. *Neurology* 1995; 45:1451-1455.

221. Burns A, Jacoby R, and Levy R. Psychiatric phenomena in Alzheimer's disease. I: Disorders of thought content. *Br.J.Psychiatry* 1990; 157:72-76.

222. Burns A, Jacoby R, and Levy R. Psychiatric phenomena in Alzheimer's disease. II: Disorders of perception. *Br.J.Psychiatry* 1990; 157:76-81.

223. Burns A, Jacoby R, and Levy R. Psychiatric phenomena in Alzheimer's disease. III: Disorders of mood. *Br.J.Psychiatry* 1990; 157:81-86.

224. Burns A, Jacoby R, and Levy R. Psychiatric phenomena in Alzheimer's disease. IV: Disorders of behaviour. *Br.J.Psychiatry* 1990; 157:86-94.

225. La Rue A. Methodological concerns: longitudinal studies of dementia. *Alzheimer Dis.Assoc.Disord.* 1987; 1(3):180-192.

226. Rubin EH, Morris JC, and Berg L. The progression of personality changes in senile dementia of the Alzheimer's type. *J.Am.Geriatr.Soc.* 1987; 35:721-725.

227. Cockrell JR and Folstein M. Mini-Mental State Examination (MMSE). *Psychopharmacol.Bull.* 1988; 24(4):689-692.

228. Roth M, Tym E, Mountjoy CQ, Huppert FA, Hendrie H, Verma S, and Goddard R. CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *Br.J.Psychiatry* 1986; 149:698-709.

229. Overall JE. A guide to the main instruments. In: *Classification and diagnosis of Alzheimer's disease: an international perspective.* Eds. Hovaguimian T, Henderson S, Khachaturian Z, and Orley J. Lewiston: Hogrete & Huber Publishers, 1987. pp. 65-77.

230. Rosen WG, Mohs RC, and Davis KL. A new rating scale for Alzheimer's disease. *Am.J.Psychiatry* 1984; 141(11):1356-1364.

231. Mattis S. Mental status examination for organic mental syndrome in the elderly patient. In: *Geriatric psychiatry: A handbook for psychiatrists and primary care physicians.* Eds. Bellak L and Karasu TB. New York: Grune & Stratton, 1976. pp. 77-121.

232. Mohs RC and Cohen L. Alzheimer's Disease Assessment Scale (ADAS). *Psychopharmacol.Bull.* 1988; 24(4):627-628.

233. Kaplan EF, Goodglass H, and Weinstraub S. *The Boston Naming Test.* Boston: Boston Veterans Administration Medical Center, 1978.

234. Wechsler D. A standardized memory scale for clinical use. *J.Psychol.* 1945; 19:87-95.

235. Wechsler D. *Wechsler Adult Intelligence Scale - Revised.* Cleveland: Psychological Corporation, 1981.

236. Blessed G, Tomlinson BE, and Roth M. Blessed-Roth Dementia Scale (DS). *Psychopharmacol.Bull.* 1988; 24(4):705-708.

237. Davous P, Lamour Y, Debrand E, and Rondot P. A comparative evaluation of the short orientation memory concentration test of cognitive impairment. *J.Neurol.Neurosurg.Psychiatry* 1987; 50:1312-1317.

238. Klein LE, Roca RP, McArthur J, Vogelsang G, Klein GB, Kirby SM, and Folstein M. Diagnosing dementia: Univariate and multivariate analyses of the mental status examination. *J.Am.Geriatr.Soc.* 1985; 33:483-488.

239. Katzman R, Brown T, Fuld PA, Peck A, Schechter R, and Schimmel H. Validation of a short orientation-memory-concentration test of cognitive impairment. *Am.J.Psychiatry* 1983; 140(6):734-739.

240. Fillenbaum GG, Heyman A, Wilkinson WE, and Haynes CS. Comparison of two screening tests in Alzheimer's disease: The correlation and reliability of the Mini-Mental State Examination and the Modified Blessed Test. *Arch.Neurol.* 1987; 44:924-927.

241. Yesavage JA, Poulsen SL, Sheikh J, and Tanke E. Rates of change of common measures of impairment in senile dementia of the Alzheimer type. *Psychopharmacol.Bull.* 1988; 24(4):531-534.

242. Ihl R, Frölich L, Dierks T, Martin E-M, and Maurer K. Differential validity of psychometric tests in dementia of the Alzheimer type. *Psychiatry Res.* 1992; 44:93-106.

243. Thal LJ, Grundman M, and Golden R. Alzheimer's disease: A correlational analysis of the Blessed Information-Memory-Concentration Test and the Mini-Mental State Exam. *Neurology* 1986; 36:262-264.

244. Stern Y, Sano M, Paulson J, and Mayeux R. Modified Mini-Mental State Examination: validity and reliability. *Neurology* 1987; 37(Suppl 1):179.(Abstract)

245. Summers WK, DeBoynton V, Marsh GM, and Majovski LV. Comparison of seven psychometric instruments used for evaluation of treament effect in Alzheimer's dementia. *Neuroepidemiology* 1990; 9:193-207.

246. Mayeux R, Stern Y, and Spanton S. Heterogeneity in dementia of the Alzheimer type: evidence of subgroups. *Neurology* 1985; 35:453-461.

247. Cronbach LJ and Furby L. How we should measure "change" - or should we? *Psychol.Bull.* 1970; 74(1):68-80.

248. Fleiss JL. *The design and analysis of clinical experiments*. New York: John Wiley & Sons, 1986.

249. Laird NM and Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; 38:963-974.

250. Moritz DJ and Petitti DB. Association of education with reported age at onset and severity of Alzheimer's disease at presentation: implications for the use of clinical samples. *Am.J.Epidemiol.* 1993; 137:456-462.

251. Kramer-Ginsberg E, Mohs RC, Aryan M, Lobel D, Silverman JM, Davidson M, and Davis KL. Clinical predictors of course for Alzheimer patients in a longitudinal study: A preliminary report. *Psychopharmacol.Bull.* 1988; 24(3):458-462.

252. Hamilton M. A rating scale for depression. *J.Neurol.Neurosurg.Psychiatry* 1960; 23:56-62.

253. Selzer ML, Vinokur A, and vonRooijen L. A self-administered short Michigan Alcoholism Screening Test (SMAST). *J.Stud.Alcohol* 1975; 36:117-126.

254. Rosen WG, Terry RD, Fuld PA, Katzman R, and Peck A. Pathological verification of ischemic score in differentiation of dementias. *Ann.Neurol.* 1980; 7:486-488.

255. Hachinski VC, Iliff LD, Zilhka E, Du Boulay GH, McAllister VL, Marshall J, Russell RWR, and Symon L. Cerebral blood flow in dementia. *Arch.Neurol.* 1975; 32:632-637.

256. Kreft IGG, de Leeuw J, and van der Leeden R. Review of five multilevel analysis programs BMDP-5V GENMOD HLM ML3 VARCL. *Am.Statistician* 1994; 48:324-335.

257. Jennrich R and Schluchter M. Unbalanced repeated measures models with structured covariance matrices. *Biometrics* 1986; 42:805-820.

258. Akaike H. A new look at the statistical identification model. *IEEE Transactions on Automatic Control* 1974; 19:716-723.

259. Brooks JO,III, Kraemer HC, Tanke ED, and Yesavage JA. The methodology of studying decline in Alzheimer's disease. *J.Am.Geriatr.Soc.* 1993; 41:623-628.

260. Yesavage JA and Brooks JO,III. On the importance of longitudinal research in Alzheimer's disease. *J.Am.Geriatr.Soc.* 1991; 39:942-944.

261. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol.Rep.* 1966; 19:3-11.

262. Crapper McLachlan DR, Dalton AJ, Galin H, Schlotterer G, and Daicar E. Alzheimer's disease: clinical course and cognitive disturbances. *Acta Neurol.Scand.* 1984; Suppl 99:83-89.

263. Mayeux R, Stern Y, Rosen J, and Leventhal J. Depression, intellectual impairment, and Parkinson disease. *Neurology* 1981; 31:645-650.

264. Lawton MP and Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969; 9:179-186.

265. Lawton MP and Brody EM. Instrumental Activities of Daily Living (IADL) scale: Original observer-rated version. *Psychopharmacol.Bull.* 1988; 24:785-788.

266. Lawton MP and Brody EM. Physical Self-Maintenance Scale (PSMS): Original observer-rated version. *Psychopharmacol.Bull.* 1988; 24:793-794.

APPENDIX 1

Summary form used in the review of prognostic studies of AD

Study type _____

Authors _____ Journal _____ Year _____

Title _____

METHODS

Study design: _____

Subjects: N _____ Type _____

Diagnosis _____ Source _____

Inclusion _____

Exclusion _____

Baseline: Severity _____ Duration _____ Age _____

Follow-up: Scheme _____ Duration _____ # Visits _____

Losses _____ Type of losses _____

Outcome(s):

Type	Criteria/Scale	Calculation

Blind assessment _____

Potential predictors: (type and criteria/scales)

_____

_____

_____

Control variables: _____

Other: _____

_____

ANALYSIS

Power _____ Multiple comparisons _____ Correction _____

Restrictions _____

Tests _____

_____

_____

_____

RESULTS

Mean progression rate: _____

_____

_____

_____

_____

_____

_____

CONCLUSIONS

_____

_____

_____

APPENDIX 2

Sample characteristics and outcome measures of 59 prognostic studies

Study[§]	n*	Source of subjects	Percent female	Age at entry (years)**	Disease duration (years)**	Outcome measure[+]	Follow-up (years)**
Berg et al. 1984 [i-(210)]	42	physicians & media	53	71.4±5.0 [63-81]	3.4±1.7 [1-9]	CDR	1
Knesevich et al. 1985 [i-(183)]	43	physicians & media	53	71.4±5.0 [63-81]	3.4±1.7 [1-9]	CDR BDS	2.5
Knesevich et al. 1986 [i-(214)]	27	physicians & media	53	71.4±5.0 [63-81]	3.4±1.7 [1-9]	CDR	2.5
Botwinick et al. 1986 [i-(153)]	21	physicians & media	53	71.4±4.4 [64-80]	3.4±1.7 [1-9]	CDR	4
Uhlmann et al. 1986 [b-(176)]	156	internal medicine clinics	65	76.5	3.6	MMSE	1
Reifler et al. 1986 [b-(177)]	131	geriatric outpatient clinic	70	77	NR	MMSE	1.4
Huff et al. 1987 [(172)]	77	memory disorders clinic	62	65.8±8.7	4.0±2.8	BIMC+BDS	>= 0.25
Grady et al. 1987 [g-(174)]	21	NR	44	65.9 [45-81]	4.1	NTB	1.6±0.8 [0.6-2.8]
Rubin et al. 1987 [i-(226)]	34	physicians & media	52	71.4±5.0 [63-81]	3.4±1.7 [1-9]	CDR	4.2
Berg et al. 1987 [i-(218)]	26	physicians & media	53	71.3±4.7 [64-81]	3.4±1.7 [1-9]	CDR	2.5
Stern et al. 1987 [e-(208)]	65	clinical research center	55	NR	NR	mMMSE BDS	2.8±1.6
Becker et al. 1988 [a-(166)]	44	AD research center	52	67.0±9.4	3.1	NTB	1.0±0.2
Faber-Langendoen et al. 1988 [i-(182)]	35	physicians & media	52	71.4±5.0 [63-81]	3.4±1.7 [1-9]	CDR	4.2
Thal et al. 1988 [(196)]	40	private referral practice	63	69.1±9.7	2.3±1.2	mBIMC	2.4 [1.2-4.5]
Katzman et al. 1988 [(151)]	142	nursing home, private practice, volunteers & research center	66	NR	NR	mBIMC	2.2 [1-6]

Study[§]	n*	Source of subjects	Percent female	Age at entry (years)**	Disease duration (years)**	Outcome measure[+]	Follow-up (years)**
Kramer-Ginsberg et al. 1988 [f-(251)]	60	AD research centers	32	65.9±7.5	NR	ADAS	[1-2]
Drevets & Rubin 1989 [i-(181)]	25	physicians & media	53	71.4±5.0 [63-81]	3.4±1.7 [1-9]	CDR BDS SPMSQ	5.5
Morris et al. 1989 [i-(180)]	37	physicians & media	52	71.4±5.0 [63-81]	3.4±1.7 [1-9]	CDR	2.8
Ortof & Crystal 1989 [(194)]	54	college of medicine	63	NR	2.6	mBIMC	2.6 [1-8.1]
Huff et al. 1990 [a-(188)]	53	AD research center	60	67.3 [53-83]	2.8	BDS MMSE	1.0 [0.8-1.2]
Lopez et al. 1990 [a-(179)]	20	AD research center	80	67.5	3.2	NTB MMSE	1
Drachman et al. 1990 [(8)]	42	university dementia clinic	55	NR	3.4±2.0 [1-10]	dependence in basic ADLs	4.5±2.1 [0.8-10.8]
Stern et al. 1990 [e-(204)]	67	clinical research center	54	NR	2.9±1.6 [0-9]	4 factors of BDS	3.1±1.6 [0.5-6.6]
Salmon et al. 1990 [(217)]	55	AD research center	62	72.4±6.9 [59-89]	NR	mBIMC MMSE DRS	[1-2]
Teri et al. 1990 [b-(187)]	106	geriatric outpatient clinic	72	77±6.7 [60-94]	3.9±2.2	MMSE	1.1±0.7
Burns et al. 1990 [j-(221-224)]	79	psychiatric hospitals	79	80.4±6.6 [56-99]	5.2±3.5 [0.5-20]	CAMCOG MMSE	1
Burns et al. 1991 [j-(197)]	85	psychiatric hospitals	79	80.7 [67-99]	4.9 [0.5-20]	CAMCOG MMSE	1
Lopez et al. 1991 [a-(178)]	34	AD research center	74	69.6	3.5	NTB MMSE	1.1
Boller et al. 1991 [a-(211)]	33	AD research center	64	67.1	NR	MMSE	2
Rosen & Zubenko 1991 [a-(215)]	32	ambulatory care setting	47	70.3±7.9	NR	MMSE	[0-3+]
Miller et al. 1991 [c-(175)]	81	dementia research center	38	NR	NR	MMSE	[1-2.5]
Stern et al. 1992 [f-(169)]	111	AD research centers	41	NR	NR	mBIMC	2.5±1.9 [0.5-8]

Study[§]	n*	Source of subjects	Percent female	Age at entry (years)**	Disease duration (years)**	Outcome measure[+]	Follow-up (years)**
Mortimer et al. 1992 [(7)]	65	VA medical center & community	22	NR	3.4±2.5 [1-13]	MMSE IADL+PSMS	[1-4]
Soininen et al. 1992 [(173)]	30	NR	66	68.3 [53-80]	2 [1-5]	NTB	3
Haxby et al. 1992 [g-(167)]	16	NR	25	63 [45-77]	3.3 [1-6]	DRS MMSE WAIS	4.6 [2.7-6.8]
Corey-Bloom et al. 1993 [(116)]	302	AD research centers	71	74.4±7.8	5.1	MMSE	1
Miller et al. 1993 [c-(115)]	74	dementia research center	41	NR	NR	MMSE	[1-2.5]
Yesavage et al. 1993 [c-(213)]	70 / 57	dementia research centers	75 / 69	NR	NR	MMSE	2.6±0.8 / 2.9±1.3
Green et al. 1993 [f-(152)]	104	AD research centers	38	68±7.8 [52-86]	4.5±2.5 [1-13]	PSMS IADL	2.5±1.3 [1-5.5]
Morris et al. 1993 [h-(160)]	430	university medical centers	54	70.9±8.0	4.1±2.2	MMSE BOMC mBDS NTB	[1-4]
Lucca et al. 1993 [(6)]	56	geriatric institutions & clinical centers	71	74.5±7.3	2.8±2.0	BIMC BDS	1
Förstl et al. 1993 [(216)]	50	psychiatric hospital & volunteers	60	68.8 [49-92]	4.5 [0.5-15]	CDR	2
Flicker et al. 1993 [(185)]	84	dementia research centre & media	NR	70.9	NR	GDS NTB	2.2±0.1
Chui et al. 1994 [(207)]	135	university medical centers	70	72.9±8.1	3.9±2.3	MMSE	[0.1-5.6]
Goldblum et al. 1994 [(212)]	16	NR	81	78 [66-90]	3.3 [0.5-6]	MMSE	1
Mielke et al. 1994 [(150)]	25	NR	44	66±7.5 [54-79]	3.3±2.1 [1-10]	MMSE	1.1±0.5 [0.5-2.2]
Kraemer et al. 1994 [c-(190)]	81	dementia research center	67	NR	4.5±2.3	MMSE	4.5±2.3 [1-14.5]

Study§	n*	Source of subjects	Percent female	Age at entry (years)**	Disease duration (years)**	Outcome measure+	Follow-up (years)**
Jacobs et al. 1994 [d-(193)]	127	outpatient dementia clinics	62	73.1	4.3	mMMSE BDS	>=2
Stern et al. 1994 [d-(192)]	236	outpatient dementia clinics	59	73.1±8.9	6.9±9.2	mMMSE BDS	>=0.5
Stern et al. 1994 [f-(170)]	111	AD research centers	40	68.2±7.9 [50-86]	NR	ADAS (C)	2.9±1.6 [1-7.5]
Hogan et al. 1994 [(191)]	135	dementia research clinic	63	NR	4.4±2.9	MMSE	>=0.5
Bracco et al. 1994 [(195)]	119	neurology departments	60	64.7±4.1	3.1±1.8	BIMC BDS	5.1±2.5 [2-9]
Teri et al. 1995 [(186)]	156	AD patient registry	67	79 [54-91]	NR	MMSE DRS	[1-4]
Galasko et al. 1995 [h-(220)]	231	university medical centres	53	70.9±7.9	NR	mBDS	2
Basun et al. 1995 [(184)]	32	general population	83	83.9	5.4	MMSE	3
Farrer et al. 1995 [(159)]	186	AD research centre	61	NR	NR	mBIMC mADL	2.8

§ Reports prefixed by the same letter are based on the same longitudinal cohort study.

* Sample sizes are based on number of subjects included in prognostic analyses.

** mean ± SD [range]

NR Not reported.

\+ **Cognitive scales:** ADAS = Alzheimer Disease Assessment Scale (including cognitive & noncognitive subscales) (230,232); ADAS (C) = ADAS (cognitive subscale only); BIMC = Blessed Information Memory Concentration scale (73,236); mBIMC = modified Blessed Information Memory Concentration scale (168); BOMC = Blessed Orientation Memory Concentration scale (239); CAMCOG = cognitive portion of the CAMDEX (228); DRS = Mattis Dementia Rating Scale (231); MMSE = Mini-Mental State Examination (164,227); mMMSE = modified Mini-Mental State Examination (263); SPMSQ = Short Portable Mental Status Questionnaire (203); WAIS = Wechsler Adult Intelligence Scale (235).
Activities of daily living scales: mADL = modified Activities of Daily Living Scale (159); BDS = Blessed Dementia Scale (73,236); mBDS = modified Blessed Dementia Scale (160); IADL = Instrumental Activities of Daily Living scale (264,265); PSMS = Physical Self-Maintenance Scale (264,266).
Staging instruments: CDR = Clinical Dementia Rating scale (120,121,219); GDS = Global Deterioration Scale (118).
Neuropsychologic test battery: NTB (number and type of tests vary from study to study).

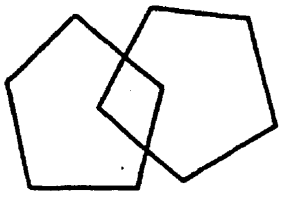
APPENDIX 3

Mini-Mental State Examination (MMSE)

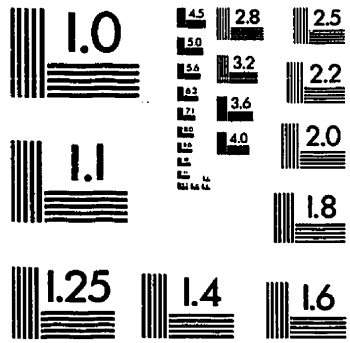
			Score	Points
Orientation				
1.	What is the	Year?	___	1
		Season?	___	1
		Date?	___	1
		Day?	___	1
		Month?	___	1
2.	Where are we?	State?	___	1
		County?	___	1
		Town or city?	___	1
		Hospital?	___	1
		Floor?	___	1
Registration				
3.	Name three objects, taking one second to say each. Then ask the patient all three after you have said them. Give one point for each correct answer. Repeat the answers until the patient learns all three.		___	3
Attention and calculation				
4.	Serial sevens. Give one point for each correct answer. Stop after five answers. *Alternate:* Spell WORLD backwards.		___	5
Recall				
5.	Ask for names of three objects learned in question 3. Give one point for each correct answer.		___	3
Language				
6.	Point to a pencil and a watch. Have the patient name them as you point.		___	2
7.	Have the patient repeat "No ifs, ands, or buts."		___	1
8.	Have the patient follow a three-stage command. "Take the paper in your right hand. Fold the paper in half. Put the paper on the floor."		___	3
9.	Have the patient read and obey the following: "CLOSE YOUR EYES." (Write it in large letters.)		___	1

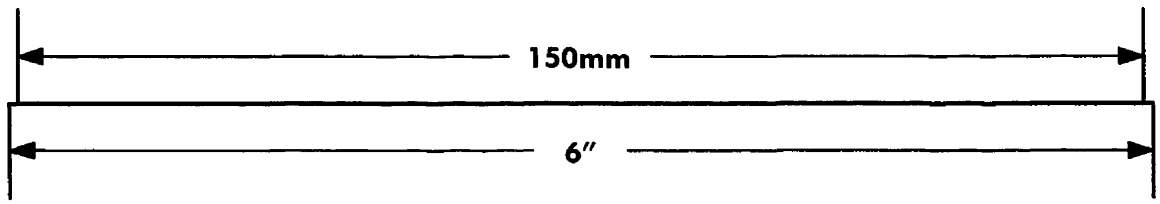
MMSE *(continued)*

10.	Have the patient write a sentence of his or her own choice. (The sentence should contain a subject and a verb and should make sense. Ignore spelling errors when scoring.)	___	1
11.	Have the patient copy the figure below. (Give one point if all sides and angles are preserved and if the intersecting sides form a quadrangle.)	___	1
		_____	= Total 30



TEST TARGET (QA-3)





APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989