Original Research Article

Dementia and Geriatric Cognitive Disorders

Dement Geriatr Cogn Disord 1998;9:299–308

Accepted: March 8, 1998

The Stages of Alzheimer's Disease: A Reappraisal

Helena C. Kraemer
Joy L. Taylor
Jared R. Tinklenberg
Jerome A. Yesavage

Veterans Affairs Health Care System and Stanford University, Palo Alto, Calif., USA

Key Words
Alzheimer's disease
Drug therapy
Psychology
Dementia severity

Abstract
'Stages', as used in clinical practice and research, are defined, their value described, and criteria are proposed for their evaluation. The specific interest is in staging Alzheimer's disease (AD). Two staging systems, one based on the Global Deterioration Scale (GDS) and one based on the Mini-Mental State Exam (MMSE), are compared in terms of these criteria, as an illustration of the process involved. We propose that there is not one unique staging system, that different staging criteria might be appropriate to different research or clinical needs, depending on which part of the temporal course of the disease is of primary interest, and on whether the focus is on cognitive, functional, neurological, behavioral, economic, or other issues. GDS staging seems a better choice for the later stages of AD when the focus is on functional change. MMSE staging seems a better choice for tracking the earlier stages of AD when the focus is on cognitive change.

Introduction

In clinical practice and research, a 'stage' is a clinically distinct period or phase of a progressive disease or process characterized by a certain pattern of signs, symptoms, responses or reactions. Salient examples are the stages of cancer, the stages of labor, the stages of childhood, and the stages of adolescence. At issue here are the stages of Alzheimer's disease (AD).

Generally, stages are clinically meaningful and recognizable, but are arbitrarily defined divisions of a process that is both complex and continuous. The divisions are arbitrary in that how many stages are defined, on what information classification into stages is based, what aspects of change are emphasized, and where the thresholds are placed defining a transition from one stage to another, are the choice of the designer of the staging system. Thus stages are not intrinsic to the disease or process. Because of this arbitrariness, there can be different valid ways of staging a particular disease or process that will produce conclusions, not necessarily identical, but complementary to each other. Different ways of staging a particular disease or process may be more or less useful in different population types (e.g., community versus clinical, mildly ill versus severely ill) or for different types of research questions (e.g., neurological versus behavioral versus social).

Stages may be especially important scientifically when there is heterogeneity among patients in the timing and course of the disease or process. By matching patients on stage, one can compare patients in a more meaningful fashion both for clinical decision making and for research purposes. Because patients within a stage are more homogeneous than patients at different stages, there is greater power to detect patterns of response in research using stage-matched patients. For example, the assessment of

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

Accessible online at:
http://BioMedNet.com/karger

Jerome A. Yesavage, MD
Department of Psychiatry and Behavioral Sciences (C-301)
Stanford University School of Medicine
Stanford, CA 94305-5548 (USA)
Tel. +1 650 852 3987, Fax +1 650 852 3297, E-Mail yesavage@leland.stanford.edu

the efficacy or effectiveness of treatments is facilitated when patients assigned to treatment and control groups are comparable in where they are in the course of the disorder. Not only is power to detect efficacy or effectiveness increased, but the effect size is maximized.

When the process or disease extends over many years, staging permits efficient segmentation of research efforts, allowing researchers to focus on one stage, or one stage-to-stage transition at a time. For example, it may be that those AD patients who have a specific genotype, etiology, gender, racial/ethnic background, or socioeconomic status will tend to have a more rapid progression, or a more virulent course than others. There may be subtypes of AD with different etiologies, different responses to treatments, and/or different courses of illness that could be identified by the pattern of stage durations. Moreover, it may well be that certain treatments are efficacious or effective only when initiated at certain stages of the disorder, or that the efficacy or effectiveness is signaled by a prolongation of the duration of an early stage of the disorder. Finally, the success of screening programs is often assessed by comparing the distribution of stage at time of initial diagnosis for those who undergo screening versus those who do not [1, pp. 162–166].

Stages also have colloquial utility [2] in that they provide the means for quick and concise communication. Between clinicians and researchers familiar with the staging system, stages locate a patient in what is often a long and complex process. In clinical practice, stages aid in explaining to patients and families what to expect as the disease progresses, how long each stage can be expected to last, and what level of care will likely be needed at each stage.

To accomplish the above objectives, a scientifically valid staging system must satisfy certain criteria:

(1) Stages must be operationally defined.

(2) Staging must have excellent inter-observer and adequate test-retest reliability (day-to-day inconsistency of the patients' responses should be the major source of unreliability) as well as convergent validity.

(3) Stages must be exhaustive and exclusive. After onset of the disease, every patient is in some stage, and only one stage, at each point of time.

(4) Stages must be progressive. Unreliability aside, after onset, the patient moves from one stage to the next in succession. Diagnosis of the disorder can be done at the earliest in the first stage, and, if death is related to the disorder and not to competing risks, death occurs in the last stage.

(5) There must be enough stages to absorb a major part of the heterogeneity of the process, but not so many that one could not study each stage, one stage at a time, nor communicate effectively in clinical practice. Three stages (e.g., mild, moderate, severe) are workable, but 5 or 7 stages are more typical. For examples, 5 stages of cancer are typically defined, and the GDS staging to be discussed below has 7 stages. In the case of AD, the median time from diagnosis to death is of the order of 5–10 years [3], and significantly longer for female than for male AD patients. Thus 5 stages of AD, with a median duration of about 1–2 years in each, would seem to be ideal for research and clinical purposes.

(6) Stages must have clinical validity such that the stages should be sensitive not only to the key clinical symptoms addressed, but also to the major clinical symptoms not explicitly considered in defining the stages. For example, the stages of cancer are defined by clinical signs, but are useful because they also relate strongly to treatment responsiveness and to probable survival time.

This paper examines two proposals for staging patients after onset of AD. One is based on the Global Deterioration Scale (GDS) [4], which was specifically designed as a staging system. The second is based on the Mini-Mental State Examination (MMSE) [5], which is an interval scale not specifically designed as a staging system, but which can be adapted to perform that function. These are selected to illustrate the process by which staging systems might be evaluated for use in clinical research and practice.

Materials and Methods

Staging Instruments

GDS Stages. The GDS [6] is a 7-point rating instrument designed for the staging of cognitive and functional capacity in normal aging, age-associated cognitive impairment, and dementia. The GDS is one of the most widely used instruments for clinical assessment of the overall magnitude of severity of age-associated memory impairment (AAMI) and primary degenerative dementia (PDD), particularly PDD of the Alzheimer type [6]. The GDS assigns a label from GDS-1 to GDS-7 to each patient, with GDS-7 indicating the greatest impairment.

GDS-1 indicates absence of any subjective or objective cognitive impairment. As this rating would necessarily predate the clinical diagnosis of AD, GDS-1 is not a clinical stage of AD per se. GDS-2 indicates subjective complaints but no objective evidence of impairment on clinical interview, nor in employment and social situations. Diagnosis of AAMI or age-related cognitive decline by DSM-IV [7] may be made during GDS-2 and diagnosis of possible AD is occasionally but rarely made during GDS-2. In effect, once clinical diagnosis of AD has been made there are 5 possible GDS stages: GDS-2 and 3 (combined), GDS-4, GDS-5, GDS-6, GDS-7.

Operational guidelines are presented for assigning GDS stages based on information about subjective complaints of memory deficit, objective observation of deficit on careful clinical interview, and an assessment of the functional ability of the patient, in part based on information provided by a knowledgeable informant [4]. While the GDS staging system provides for stages that are exhaustive and exclusive, users sometimes signify their inability to assign an AD patient to one and only one stage by assigning half-stages, i.e. GDS-3.5, to signify a mixture of GDS-3 and 4. These half-stages have no operational definition, and compromise the criterion of exhaustive and exclusive stages for GDS staging. Here, those assigned to a half-stage are reassigned to the stage below, i.e. GDS-3.5 is treated as GDS-3, because the implication is that the operational criteria for GDS-4 are not yet fully satisfied.

The GDS stages so defined have clinical validity, and have been demonstrated to have both interobserver and test-retest reliability [8].

MMSE Stages. MMSE [5] is a widely used clinical scale, designed for preliminary screening, diagnosis and serial assessment of psychogeriatric patients, providing a very brief but formal and relatively thorough measure of cognition, including orientation, memory, attention, language, and design copying [9]. Scores range from 0 to 30 on an interval scale, with 30 the best possible score. Instructions for administering and scoring the examination, including consideration of special circumstances, are available [10]. The MMSE also has clinical validity, and has been demonstrated to have both interobserver and test-retest reliability [5].

We here propose the following staging system for the MMSE: MMSE-1: 24–30; MMSE-2: 15–23; MMSE-3: 8-14; MMSE-4: 4–7; MMSE-5: 0–3. The cut-point of MMSE=23 is often used clinically [11] to define onset of dementia. The cut-point of MMSE=15 was selected as the lower bound of 'early' AD as a selection criterion into our studies. The cut-points under 15 are set at intervals in a 2:1:1 ratio because clinical changes are more rapid as MMSE approaches zero. Nevertheless, while the choice of cut-points may be informed by expected clinical changes, by expected durations, or by test-retest unreliability, it should again be noted, that they remain, until validation, arbitrary choices.

Sample

All 206 patients in this study were recruited into the Stanford-Palo Alto VA Aging Clinical Research Center (ACRC) between January, 1982 and January, 1997 and had a clinical consensus diagnosis of probable Alzheimer's [12]. To determine the diagnosis of probable AD, all patients had a complete medical, psychiatric, neurologic, neuroimaging, and neuropsychological assessment. Based on these evaluations, a consensus diagnosis was reached by an interdisciplinary team including 1–3 physicians and at least 2 other experienced clinicans. Only AD patients with active major medical problems (e.g., congestive heart failure or recent life-threatening cancer) that would have made participation in an intensive longitudinal study difficult, were excluded. Of the patients recruited on whom we have neuropathology results (n = 46), 89% have been confirmed as definite AD with or without other diagnoses (e.g. Lewy body variant, ischemic vascular changes). Patients were recruited as early in the disease process as possible, and were recruited to have an MMSE score of 15 or above at entry (i.e., MMSE-1 or 2). At entry all were home-dwelling or living in board-and-care homes, i.e., none were institutionalized in skilled nursing facilities at entry. Written informed consent was obtained from patients or from their caregivers at entry. Each patient was to be followed longitudinally at approximately 6-month intervals. Because the emphasis of the ACRC is on the early and middle stages of AD, routine clinical follow-up ceased after two successive cognitive assessments (1 year or more) at which MMSE = 0. Effort was made to solicit donation of brains at time of death from all patients for neuropathological studies.

The population from which this sample is drawn tends to be of high socioeconomic status (42% college graduates) and predominantly nonminority (93% Caucasian). Of the sample, 62% were men. The data used in this report were taken from the resulting longitudinal database.

Clinical Measures

The Alzheimer's Disease Assessment Scale (ADAS Cog, Non-cog) [13] contains 21 items divided into two subscales: (1) the 11-item cognitive subscale for assessment of receptive and expressive language ability, orientation, constructional and ideational praxis, and word-list recall and recognition, and (2) the 10-item noncognitive subscale for assessment of mood, psychotic symptoms, motor activity, tremors, and appetite change. Possible cognitive scores range from 0 to 70, and noncognitive scores from 0 to 50, with higher scores on both subscales reflecting greater dysfunction. The cognitive subscale was administered by a clinician (orientation items) and a trained psychology technician (remaining items). The noncognitive subscale ratings were made by the clinician, based on information obtained through separate interviews with the patient and caregiver.

From the ADAS cognitive subscale, we also obtained measures of aphasia and apraxia, which are salient indicators of rapidity of cognitive decline [14]. Following recent work by the ADAS developers [15], the aphasia score was the mean of ratings for 'difficulty making self understood', 'comprehension of spoken language', 'word-finding difficulty', 'following commands', and 'confrontation naming'. The apraxia score was the mean of the two praxis items. The aphasia and apraxia scores range from 0 to 5, indicating 'no' to 'severe impairment' respectively.

Time-Based Behavioral Disturbance Questionnaire (TBDQ). The TBDQ is designed to measure diurnal variations in behavioral disturbance [16, 17]. The TBDQ was filled out by consenting caregivers if they lived with the patient at the time of the assessment. The TBDQ requests caregivers to report whether, over the previous month, the patient exhibited any of the following seven behaviors: combativeness, agitation, wandering, incoherent speech, hallucinations, confusion, and disorientation (definitions were provided for each behavior). Respondents were asked to check off the time periods during which the behavior took place: morning (Wake until 12 p.m.); early afternoon (12 noon to 4 p.m.); late afternoon/evening (4 p.m. to 10 p.m.); night (10 p.m. to Wake). The TBDQ score used here is the overall score, computed for each patient as the percentage of the seven behaviors checked across the four time periods.

Incontinence. Incontinence was assessed through clinical interview with the caregiver. In each interview, one or two recording methods were used: either the incontinence item of the Blessed-Roth Dementia Rating Scale [18], or the incontinence item of the California AD Program's Minimum Uniform Data Set which is the data collection instrument of the California AD Diagnostic and Treatment Centers. In either case, no differentiation is made between incontinence due to cognitive or physical impairment.

In all cases, the MMSE and GDS assessments associate with a particular measure that was nearest to the time of measurement within 90 days.

Table 1. Correspondence between GDS and MMSE stages

GDS/MMSE	MMSE-1	MMSE-2	MMSE-3	MMSE-4	MMSE-5	Total
MMSE score	24–30	15–23	8–14	4–7	0–3	
GDS-2,3	**106**	78	1	1	0	186 (17)
GDS-4	33	**247**	17	0	0	297 (28)
GDS-5	0	110	**157**	23	4	294 (27)
GDS-6	0	7	65	**49**	**135**	256 (24)
GDS-7	0	0	0	0	**37**	37 (3)
Total	139 (13)	442 (41)	240 (22)	73 (7)	176 (16)	1,070 (100)

Numbers in bold indicate the modal response in a row and/or colum. Values in parentheses are percentages.

Table 2. Transition frequencies for MMSE stages (total transitions = 863), to assess whether MMSE stages are progressive

From/To	MMSE-1	MMSE-2	MMSE-3	MMSE-4	MMSE-5	Total
MMSE-1	63	58	1	0	2	124
MMSE-2	**21**	227	121	10	12	391
MMSE-3	0	**13**	105	51	32	201
MMSE-4	0	0	**7**	10	45	62
MMSE-5	0	0	0	**2**	83	85

Numbers in bold indicate the number of backward transitions.

Table 3. Transition frequencies for GDS stages to assess whether GDS stages are progressive

From/To	GDS-2,3	GDS-4	GDS-5	GDS-6	GDS-7	Total
GDS-2,3	89	68	8	2	1	168
GDS-4	**8**	134	105	10	4	261
GDS-5	**2**	**7**	148	93	3	253
GDS-6	0	0	4	148	20	172
GDS-7	0	0	0	**1**	8	9

Numbers in bold indicate the number of backward transitions.

Results

Staging Criteria 1 and 2: Operational Definition, Interobserver and Test-Retest Reliability, and Convergent Validity

The MMSE and GDS stages described above are both based on scales that have been operationally defined and have documented interobserver and test-retest reliability [5, 8]. In the current data, 206 patients were observed on 1,070 occasions when MMSE and GDS scores were both obtained for the same patient within the same week. Staging based on these observations are presented in table 1.

There is a strong correspondence between the two staging systems, in that modal responses for a row or column tend to lie on the diagonal. But, as would be expected from the less than perfect test-retest reliability of each staging system, there are also discrepancies between the two staging systems. In general, MMSE-1 corresponds to GDS-2 and 3; MMSE-2 corresponds to GDS-4; MMSE-3 corresponds to GDS-5. Discrepancies occur primarily in the later stages of dementia, such that MMSE-4 and 5 are less distinctly related to GDS-6 and 7.

Staging Criteria 3 and 4: Exhaustive and Exclusive Categories That Are Progressive

For both the GDS and the MMSE stages, transition matrices were compiled, counting for each pair of stages, how many times on two successive observations of an AD patient (usually about 6 months apart), the AD patients were seen initially in one stage and subsequently in another (n = 573). If a staging system is both reasonably reliable and progressive, there may be a few 'one-step backward' transitions, e.g. from stage 5 to 4, due to unreliability, but very few, if any, 'two or more-step backward transitions', e.g. from stage 5 to stage 3.

In table 2 are presented the transition frequencies for the MMSE stages, and in table 3 those for the GDS stages. Only 3% of the GDS stage transitions and 5% of the MMSE stage transitions went backward by one step. None of the MMSE stage transitions were backward by more than one step, but 2 of the GDS stage transitions (from GDS-5 to GDS-2,3) went backwards two steps or more. Both staging systems appear to reasonably satisfy this criterion.

It should also be noted that observed progression through stages, however defined, is not necessarily smooth. For example, there are 3 patients who transition from GDS-2,3 to GDS-6 or 7, and 2 patients who transition from MMSE-1 to MMSE-5. Occasionally, a patient appears to skip a stage in an approximately 6-month period. This phenomenon may be due in part to interobserver and test-retest unreliability of the measurements underlying the staging, but may also reflect important individual differences among AD patients in the progression of the disease.

Staging Criterion 5: Adequate Stage Durations

In a sample of AD patients representative of a population, with each patient followed at regular intervals from diagnosis to death, the proportion of observations in each stage is an unbiased estimate of the percentage of total duration spent in each stage in that population. Similarly, the number of transitions that do not change stages (the numbers on the diagonal of tables 2, 3) are also indicators of relative average duration in each stage.

The present sample is selected to be a representative sample of available and eligible AD patients, but is not a representative sample across the course of the illness, as it undersamples time in MMSE-5. In this sample, the longest durations in the GDS system are in GDS-4, 5, 6 and in the MMSE system are in MMSE-2, 3, with very short duration in GDS-7. In the present sample of AD patients, there were only 37 (3%) observations in GDS-7 (table 1).

Table 4. Estimated median duration (years) in each of the intermediate stages

Stage	MMSE stages		Stage	GDS stages	
	n	duration years		n	duration years
MMSE-2	111	1.6	GDS-4	100	1.2
MMSE-3	45	1.4	GDS-5	90	1.5
MMSE-4	43	0.8	GDS-6	19	2.0

Moreover, all time spans from entry to one stage to entry to the next higher stage in this sample were located, and an estimate of the median in each such intermediate stage was computed. This is a crude estimate, for successive observations are taken at approximately 6-month intervals, and occasionally 9 months or more may lapse until the next measurement. Thus entry to a stage first observed at a time point could actually have occurred at any time since the last observation. However, the bias occurs at both ends of the intervals, and thus should generally cancel out. In table 4 are presented the estimates of the medium duration (in years) at each corresponding intermediate stage for GDS and MMSE staging in this sample. Again this confirms the difference in how the two systems operate, as AD patients tend to spend a longer time in MMSE-2 and 3, and shorter times in MMSE-4, whereas they tend to spend a shorter time in GDS-4 and 5, and a longer time in GDS-6.

Despite the sampling differences, the unbiased estimates of duration of GDS-5 and -6 reported for a different AD population where there is likely an oversampling of time in MMSE-5, which are also estimated quite differently [8] (page 302), are 1.4 and 2.4 years, corresponding well to 1.5 and 2.0 years estimated here.

Of the 43 AD patients in the present sample who are deceased and for whom we have an MMSE within a year of death, 42% (18/43) had an MMSE=0. In general, 56% (24/43) were in stage 5, 12% (5/43) were in stage 4, 19% (8/43) were in stage 3, 12% (5/43) were in stage 2, and 2% (1/43) were in stage 1. Clearly, many AD patients do not survive to reach stage 5, although that is the modal stage at death.

Staging Criterion 6: Clinical Validity

A valid staging system must be sensitive to important clinical changes over time, including significant clinical symptoms not explicitly considered in defining the stages.

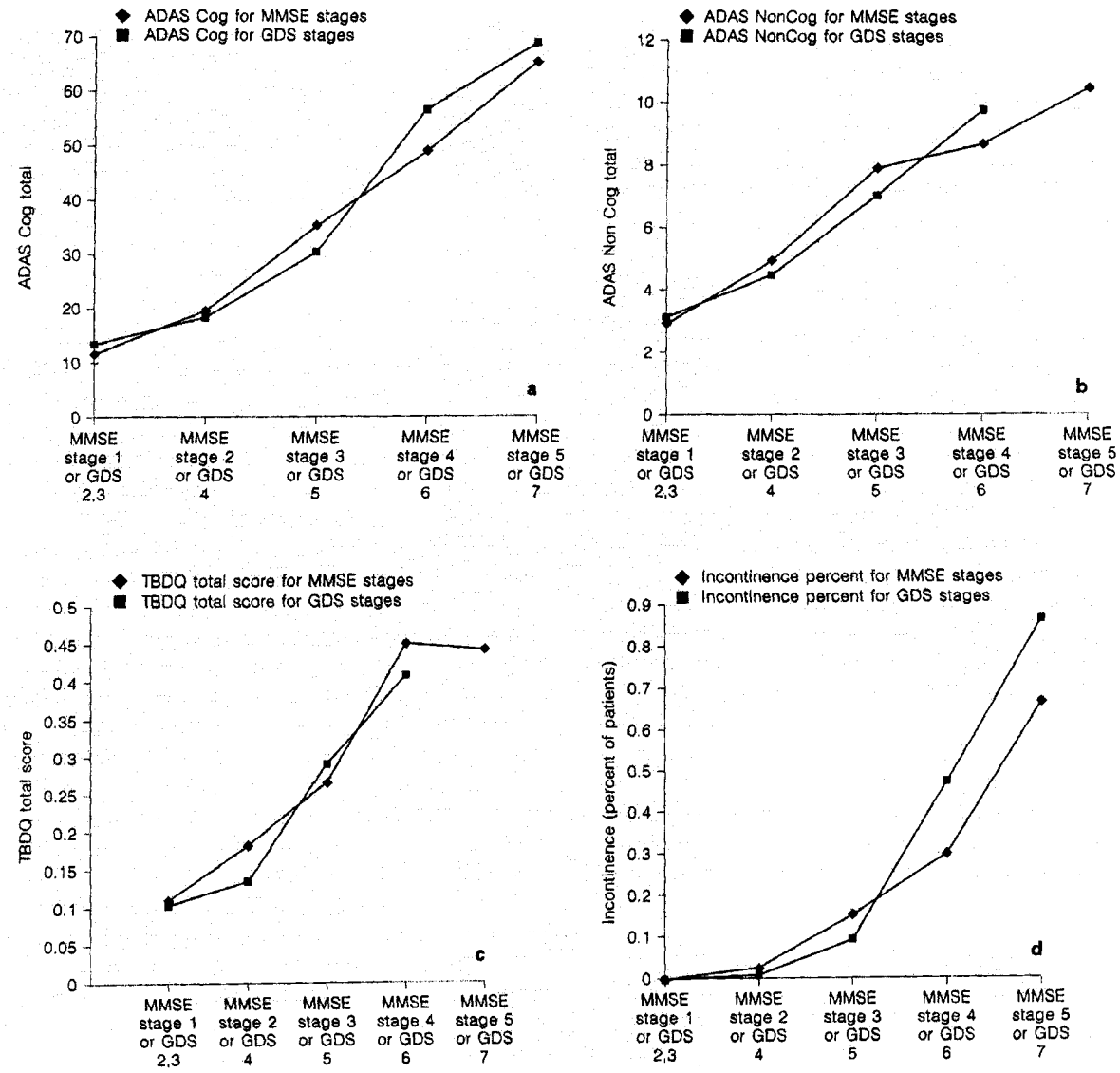

Fig. 1a–d. Means of response measures external to the staging systems for stages 1 (MMSE-1, GDS-2,3), 2 (MMSE-2, GDS-4), 3 (MMSE-3, GDS-5), 4 (MMSE-4, GDS-6), 5 (MMSE-5, GDS-7).

Figures 1a–d and table 5 present the means of observed characteristics clinically associated with AD progression: ADAS cognitive total scores, specific measures of aphasia and apraxia, ADAS noncognitive behavioral ratings, TBDQ disruptive behaviors, and incontinence. Note that often there were too few observations in GDS-7 to assess the results. These results show that in both staging systems all the clinical symptoms measured increased in severity in the more advanced stages. The pattern of changes shown in the table and figures suggests that it makes very little difference whether GDS or MMSE staging is used. Both appear to have clinical validity.

Table 5. Mean ADAS aphasia and apraxia scores for each of the staging systems

Stage level	Aphasia score		Apraxia score	
	MMSE stages	GDS stages	MMSE stages	GDS stages
MMSE-1 or GDS-2,3	0.51	0.64	0.37	0.46
MMSE-2 or GDS-4	0.99	0.99	0.88	0.81
MMSE-3 or GDS-5	2.28	1.87	2.11	1.68
MMSE-4 or GDS-6	3.32	3.80	3.14	3.81
MMSE-5 or GDS-7	4.68	4.96	4.58	5.00

Discussion

Both the MMSE and GDS instruments were shown to be useful as staging systems. Each system has its own advantages and limitations in research and clinical practice. First, as a backdrop for discussing limitations, we discuss the sampling and measurement issues enmeshed in the development of any staging system.

Sampling and Measurement Issues

Sampling issues are crucial in estimating the total duration of any disease as well as the duration of separate stages. Ashford et al. [19] emphasize the necessity of testing the entire range of AD pathology to obtain a nonbiased measurement of duration in any stage of disease. Such a task can very likely only be accomplished by sampling patients at the time of diagnosis of AD with regular and frequent assessment up to the time of death.

Our Center, for example, focuses less on the end stage of AD, because we are particularly interested in individual differences and determinates of rate of cognitive decline and in predictors of nursing home placement. As a result, our duration indicators for GDS-7 – based on the proportion of observations in each stage and on the number of transitions that do not change stages – very likely underestimate the duration of the terminal stage of AD in the general AD population. On the other hand, a study giving an estimate of an average of a total of 7 years for GDS-7 is likely to be an overestimate [20]. As Reisberg [20] noted, each duration was computed conditional on reaching that stage. That is, patients who died before reaching that stage were excluded from the computation, not included as 0. In a later study [8], approximately 50% of patients over the age of 75 at entry (GDS 4–6) had died before the average follow-up interval of 4.6 years. The overall pattern of results, like those in the sample here reported, indicates that many AD patients, especially those over 75 at time of diagnosis, either do not survive to GDS-7 and/or have only a short duration at that stage.

The comorbidity that affects any aging population further complicates sampling. Moritz et al. [3], for example, found that at the time of diagnosis of probable or possible AD, 25.2% of males and 17.0% of females patients had a history of heart disease; 20.3% of males and 31.0% of females had a history of hypertension; 5.1% of males and 15.6% of females had chronic obstructive pulmonary disease. Comorbid conditions may exacerbate the course of AD, and AD may also exacerbate the course of the comorbid conditions. To impose stringent criteria in patient selection excluding most comorbidity in order to have a sample of 'pure AD', as some studies have, will produce a sample that is not representative of AD patients in general, and will tend to give an overestimation of the average total duration of AD. Indeed, Moritz et al. [3] report that for women (but not for men) who constitute the majority of AD patients, comorbid physical conditions such as a history of heart disease, stroke and diabetes, rather than signs and symptoms of AD, are the significant predictors of time from diagnosis to death.

Studies that exclude both certain stages of AD and certain patients with comorbid conditions are most likely to produce biased estimates of the duration of AD and its stages. For example, one study [8] excluded, at baseline, all individuals in GDS- 2 or 3, samples no one in GDS-7, and excluded all patients with comorbid conditions. Exclusion of comorbidities would tend to produce an overestimate of the total duration of AD, and the exclusion of early AD would tend to overestimate the proportion of time spent in later GDS stages. In a similar way, estimated durations of illness may be quite different in studies of institutionalized patients or in samples focusing on early age of onset of AD.

Intertwined with the issue of longevity is the issue of progression. More disease progression is conceptually possible in patients who live longer, and it is clinically obvious that when a patient reaches the end stage, however defined, the disease continues to progress until death. The MMSE staging system defined here will be more limited than is the GDS staging system in this respect, because of the differential emphases these two staging systems place on early versus late stages. Thus the GDS and its accompanying ordinal assessment of functional capacity, the functional assessment staging (FAST) [21], appear to be most applicable where the research emphasis is on late-stage AD.

Measurement and Analytic Issues Related to Staging

Crucial to the discussion of measurement and analysis of staging based upon rating scales is the determination of whether the scale is an ordinal scale or an interval scale. In mathematics [22], a nominal scale (e.g., ethnic group: Caucasian, African American, Asian American, Hispanic American, Other) applies different labels to different items; an ordinal scale (e.g., letter grades: A, B, C, D, E), in addition, imposes an ordering between the items; an interval scale (e.g., counts, length, weight, time), furthermore establishes a unit of measurement. One set of objects can be labeled in different ways, ordered in different ways, or scaled in different ways. For example, the MMSE score, like the Hamilton Rating Scale for Depression, the Brief Psychiatric Rating Scale and the ADAS, a count of signs and symptoms, is thus an interval scale. However, the items on the MMSE might also be given weights indicative of their different clinical severities, or might be assigned a weight equal to the average time after onset at which that sign or symptom appears, and other interval scores might be the average of the weights so assigned. The scores would likely be highly correlated, but not identical. In contrast, both GDS staging and MMSE staging, as defined above, are ordinal, allowing only the decision that one patient might be at a later stage than another.

The mathematical importance of the level of measurement lies in the fact that computational operations available to interval scales are different from those available to ordinal scales. It is often convenient to use numbers (1, 2, 3 ...) as the labels attached to stages. However, results based on adding (averaging) and subtracting numerical labels attached to the stages (e.g., in computing means, standard deviations, percentage of variance accounted for) are logically and mathematically questionable [23]. Linear models, such as those used in analyses of variance, regression, and correlation analyses are validly used only for interval scales. Yet, many of the results in the research literature are based on application of linear models to ordinal scales.

The issue of more meaningful quantification of rate of decline has been discussed in depth by Ashford et al. [19]. In this study they developed a 'time-index' model of severity of AD, thus also an interval scale. Of interest for this presentation is that they used both an expanded form of the MMSE and a series of clinical questions based in part on the GDS. Their time-index method set as a goal a measure quantified in terms of time of progression across the length of the average clinical course of AD.

Clinical Practice and Research Issues

Valid staging systems facilitate communication because a progressive disease or developmental process is split into a small number of clinically distinct periods. Staging also allows researchers to focus on more homogeneous patient groups to gain greater statistical power and clearer interpretation of results. However, the global ordinal nature of stages tends to make them suboptimal as measures for detecting intraindividual change if one's interest lies in short-term change or in tracking change in a specific ability or symptom. General cognitive scales, such as the ADAS Cognitive Subscale and the Dementia Rating Scale [24], and specific cognitive tests such as verbal fluency and clock drawing are preferable for detecting change because these measures are more finely scaled and can be performed by mild to moderately severe dementia patients for 1–2 years [25–28].

As staging systems for AD, both the MMSE and GDS were shown to be useful. Each system may have different advantages and disadvantages for specific AD subpopulations or for some particular focus of research interest. The acid test is whether a particular staging system leads to new insights into the course of AD that are important for clinical decision making and are replicated in future independent research. Any staging system based on different information, obtained from different sources, or emphasizing different aspects of the clinical changes that take place over the time course of AD is likely to be highly correlated but not identical to another. Yet each might be very useful for specific research and clinical purposes.

Other scales such as the ADAS or the Hierarchic Dementia Scale [29], like the GDS scale, correlate highly with the MMSE and with each other. Ashford et al. [19], for example, note that among the range of patients they tested, there was a close correlation between the MMSE and the time-index form of assessment, such that the MMSE could explain 90% of the variance of the latter. Each such scale could also be used to define stages that would perform comparably to both the GDS and the MMSE staging. Each could undergo the kind of evaluation demonstrated above for GDS and MMSE staging both to assess its merit and its performance relative to other proposed and validated staging systems. However stages are defined, within each stage, there are always finer grained clinical changes that might be of interest to those focusing on the study of one stage, or of one stage-to-stage transition. For example, delineation of the time of onset of symptoms to the time of diagnosis, both of which may occur in GDS-2, 3 or in MMSE-1, would require much finer resolution of the earliest observable

clinical changes that take place within those stages. In such cases, a strategy such as that used to define the FAST substages of the late stages of AD might be necessary to study the early stages.

With multiple valid and reliable staging systems potentially available, how would one choose among them? What issues are involved? To illustrate: For the purposes of our clinical research to date, we have elected to use MMSE staging, the definition, and documentation for which has been here described. MMSE staging is easier to use, takes less time, and has great clinical familiarity. In addition, the MMSE is based on clinical observation and evaluation of the patient rather than on informant report in an interview. This is especially advantageous when there is no informant available, as in screening frail elders who live alone, when there is no reliable or consistent informant, or where the quality of informant information may be inconsistent or questionable, as in the case of double-dementia couples.

Moreover, interval scales such as the MMSE scores can be used in a variety of powerful statistical analytic methods inappropriate for ordinal scales such as GDS or MMSE stages. For certain research investigations, having the MMSE interval scale available would yield greater power, and having the MMSE staging available for other studies where staging affords greater clarity is an advantage. Finally, because of our hopes of finding ways of delaying onset, prolonging the early stages of AD, and delaying the onset of the final stages and institutionalization, we have placed the emphasis of our study on the early stages. The increased emphasis that the MMSE places on the earlier stages is more consonant with our research goals than is GDS staging.

If our context had been different, the choice of staging system might also well have been different. A staging system more focused on specific neurological, physical, behavioral, social, or economic changes might be preferable to the GDS (functional change), or the MMSE (cognitive change) staging, for those researchers particularly interested in those specific facets of the multiple clinical changes that occur in the process of AD. In short, there need not be one standard staging system for the field. What is needed is that each research project use a staging system that is well-defined, documented to satisfy the criteria articulated above, justified in terms of the context and goals of the research projected, and consistently applied.

Acknowledgements

We express appreciation to Leah Friedman, PhD, for helpful critiques of earlier drafts and for collaborative collection of time-based behavioral data; to Art Noda, MS, for expert data management and analysis, and to our clinical research staff: Terry Miller, MD, Helen Davies, RNMS, Mary Meyer, RN, Quinn Kennedy, Stephanie Rogerson and Deryl Wicks, for clinical testing and assessments. This work was supported in part by the NIH grant MH-40041, the State of California Alzheimer's Disease Program, and the Medical Research Service of the Department of Veterans Affairs.

References

1 Morrison AS: Screening in Chronic Disease. New York, Oxford University Press, 1985.

2 Horowitz FD: Developmental theory, prediction, and the developmental equation in follow-up research; in Friedman SL, Haywood HC (eds): Developmental Follow-Up: Concepts, Domains and Methods. San Diego, Academic Press, 1994, pp 27–44.

3 Moritz DJ, Fox PJ, Luscombe SA, Kraemer HC: Neurological and psychiatric predictors of mortality in patients with Alzheimer's disease in California. Arch Neurol, in press.

4 Reisberg B, Ferris SH: The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry 1982;139: 1136–1139.

5 Folstein MF, Folstein SE, McHugh PH: Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–198.

6 Reisberg B, Ferris SH, de Leon MJ, Crook T: The Global Deterioration Scale (GDS). Psychopharmacol Bull 1988;24:661–663.

7 American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. Washington, American Psychiatric Association, 1994.

8 Reisberg B, Ferris SH, Franssen EH, Shulman E, Monteiro I, Sclan SG, Steinberg G, Kluger A, Torossian C, de Leon MJ, Laska E: Mortality and temporal course of probable Alzheimer's disease: A 5-year prospective study. Int Psychogeriatr 1996;8:291–311.

9 Cockrell JR, Folstein MF: Mini-Mental State Examination (MMSE). Psychopharmacol Bull 1988;24:689–690.

10 Spencer MP, Folstein MF: The Mini-Mental State Examination; in Keller PA, Ritt LG (eds): Innovations in Clinical Practice: A Source Book. Sarasota, Professional Resource Exchange, 1985, pp 305–310.

11 Anthony JC, LeResche L, Niaz U, Korff MRV, Folstein MF: Limits of the 'Mini-Mental State' as a screening test for dementia and delirium among hospital patients. Psychol Med 1982;12: 397–408.

12 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM: Clinical diagnosis of Alzheimer's disease: A report of the NINCDS-ADRDA work-group under the auspices of the Deparmtent of Health and Human Services Task Force on Alzheimer's disease. Neurology 1984;34:939–944.

13 Rosen WG, Mohs RC, Davis KL: A new rating scale for Alzheimer's disease. Am J Psychiatry 1984;141:1356–1364.

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

15 Lawlor BA, Ryan TM, Schmeidler J, Mohs RC, Davis KL: Clinical symptoms associated with age at onset in Alzheimer's disease. Am J Psychiatry 1994;151:1646–1649.
16 Bliwise DL, Yesavage JA, Tinklenberg JR: Sundowning and rate of decline in mental function in Alzheimer's disease. Dementia 1992;3:335–341.
17 Friedman L, Kraemer HC, Zarcone V, Sage S, Wicks D, Bliwise DL, Sheikh J, Tinklenberg J, Yesavage JA: Disruptive behavior and actigraphic measures in home-dwelling patients with Alzheimer's disease: Preliminary report. J Geriatr Psychiatry Neurol 1997;10:58–62.
18 Blessed G, Tomlinson BE, 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.
19 Ashford JW, Shan M, Butler S, Rajasekar A, Schmitt FA: Temporal quantification of Alzheimer's disease severity: 'Time index' model. Dementia 1995;6:269–280.
20 Reisberg B: Dementia: A systematic approach to identifying reversible causes. Geriatrics 1986;41:30–46.
21 Reisberg B: Functional assessment staging (FAST). Psychopharmacol Bull 1988;24:653–659.
22 Senders VL: Measurement and Statistics, New York, Oxford University Press, 1958.
23 Lyons J, Welkowitz J: Round table discussion: The statistician's view on outcome methodologies for pharmacologic trials in mild, moderate, and severe Alzheimer's disease. Int Psychogeriatr 1996;8:317–322.
24 Mattis S: Dementia Rating Scale. Odessa, Psychological Assessment Resources, 1989.
25 Salmon DP, Thal LJ, Butters N, Heindel WC: Longitudinal evaluation of dementia of the Alzheimer type: A comparison of 3 standardized mental status examinations. Neurology 1990;40:1225–1230.
26 Small BJ, Backman L: Predictors of longitudinal changes in memory, visuospatial, and verbal functioning in very old demented adults. Dementia, in press.
27 Mohs RC: The Alzheimer's Disease Assessment Scale. Int Psychogeriatr 1996;8:195–203.
28 Rogers SL, Friedhoff LT: The efficacy and safety of donezepil in patients with Alzheimer's disease: Results of a US multicentre, randomized, double-blind, placebo-controlled trial. Dementia 1996;7:293–303.
29 Cole MG, Dastoor DP, Koszycki D: The hierarchic dementia scale. J Clin Exp Gerontol 1983;5:219–234.