NEUROLOGY

Dementia diagnoses from clinical and neuropsychological data compared: The Cache County study

J. T. Tschanz, K. A. Welsh-Bohmer, I. Skoog, N. West, M. C. Norton, B. W. Wyse, R. Nickles and J. C. S. Breitner

Neurology 2000;54;1290-1296

This information is current as of July 16, 2007

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://www.neurology.org/cgi/content/full/54/6/1290

Neurology is the official journal of AAN Enterprises, Inc. A bi-monthly publication, it has been published continuously since 1951. Copyright © 2000 by AAN Enterprises, Inc. All rights reserved. Print ISSN: 0028-3878. Online ISSN: 1526-632X.

AMERICAN ACADEMY OF NEUROLOGY

Downloaded from www.neurology.org by Renzo Rozzini on July 16, 2007

Dementia diagnoses from clinical and neuropsychological data compared

The Cache County study

J.T. Tschanz, PhD; K.A. Welsh-Bohmer, PhD; I. Skoog, MD, PhD; N. West, MS; M.C. Norton, MS; B.W. Wyse, PhD; R. Nickles, BS; and J.C.S. Breitner, MD, MPH

Article abstract—*Objective:* To validate a neuropsychological algorithm for dementia diagnosis. *Methods:* We developed a neuropsychological algorithm in a sample of 1,023 elderly residents of Cache County, UT. We compared algorithmic and clinical dementia diagnoses both based on DSM-III-R criteria. The algorithm diagnosed dementia when there was impairment in memory and at least one other cognitive domain. We also tested a variant of the algorithm that incorporated functional measures that were based on structured informant reports. *Results:* Of 1,023 participants, 87% could be classified by the basic algorithm, 94% when functional measures were considered. There was good concordance between basic psychometric and clinical diagnoses (79% agreement, kappa = 0.57). This improved after incorporating functional measures (90% agreement, kappa = 0.76). *Conclusions:* Neuropsychological algorithms may reasonably classify individuals on dementia status across a range of severity levels and ages and may provide a useful adjunct to clinical diagnoses in population studies. **Key words:** Dementia—Neuropsychology—Psychometric classification.

NEUROLOGY 2000;54:1290–1296

Epidemiological studies of dementia often rely on diagnoses made by cognitive screening that is followed-up by clinical, laboratory, neuroimaging, and neuropsychological studies. This method simulates the diagnostic practice followed in most university clinics, but it may be vulnerable to changes or differences in the interpretation of diagnostic criteria over time (e.g., *Diagnostic and Statistical Manual of Mental Disorders,* 3rd edition, revised [DSM-III-R],[1] DSM-IV,[2] International Classification of Diseases–10[3]) and by different diagnosticians. To address these issues, prior researchers have developed an objective algorithm to diagnose dementia. This algorithm, based only on individuals' neuropsychological test performances, exhibited moderate-to-good agreement with physician diagnoses,[4,5] and very good agreement when considering the presence or absence of dementia.[5] In the well-known Nun Study of dementia,[6] a neuropsychological algorithm supplemented with functional measures was the sole method used to identify cases of dementia.

An algorithmic approach to dementia classification avoids biases in self-reported or informant-based information from clinical interviews, circumvents vagaries of clinical judgment, and is less vulnerable to biases in decision making.[7] But it ignores the clinical history, which may be more sensitive to mild dementia,[8] and it overlooks functional abilities that are an important component of most criteria for dementia diagnosis. In addition, it is not yet clear that neuropsychological tests, with or without measures of functional impairment, can accurately diagnose dementia.

We explored the latter question in a sample of 1,023 elderly participants in an epidemiologic study of dementia,[9] comparing diagnoses derived from two neuropsychological algorithms with clinical diagnoses that followed DSM-III-R criteria.

Methods. *Participants.* We studied all individuals who completed the stages of dementia evaluation in the Cache County (UT) study. Briefly, we screened 5,092 out of 5,677 eligible participants (90%) for cognitive impairment using the Modified Mini-Mental State Examination (3MS),[10] slightly revised for this study,[9] or interview of a knowledgeable respondent with the Informant Questionnaire for Cognitive Decline (IQCODE).[11] Participants were genotyped at *APOE*.[12] The subsequent screening and assessment of participants have been described elsewhere.[9]

We also asked a weighted stratified subsample of 960 county residents to complete all stages of dementia assess-

From the Department of Psychology (Dr. Tschanz and M.C. Norton), College of Family Life (Drs. Tschanz and Wyse, N. West, M.C. Norton, and R. Nickles), Utah State University; the Department of Psychiatry and Behavioral Sciences and the Bryan Alzheimer's Disease Research Center (Drs. Welsh-Bohmer and Breitner), Duke University; the Institute of Clinical Neuroscience (Dr. Skoog), Section of Psychiatry, Göteborg University; and the Department of Mental Hygiene (Dr. Breitner), The Johns Hopkins University, Baltimore, MD.

Supported by National Institutes of Health grant AG-11380.

Presented in part at the 26th annual meeting of the International Neuropsychological Society; February, 1998; Honolulu, HI.

Received April 6, 1999. Accepted in final form December 3, 1999.

Address correspondence and reprint requests to Dr. J.T. Tschanz, Utah State University, Cache County Memory Study, UMC 4440, Logan, UT 84322-4440; e-mail: joannt@cc.usu.edu

 Copyright © 2000 by the American Academy of Neurology

Downloaded from www.neurology.org by Renzo Rozzini on July 16, 2007

ment. We asked these individuals and 1,029 other participants with positive screening results (1,989 total) to undergo a second stage of dementia screening with the Dementia Questionnaire (DQ).[13,14] The DQ informant interview inventories cognitive symptoms, relevant medical history, and difficulties in daily functioning. A neuropsychologist or a geropsychiatrist rated the DQ interviews for the presence of cognitive impairment. When necessary, as an alternative, a research nurse conducted a brief "doorstep" evaluation to assess the need for additional testing. We asked participants whose DQs were rated as having "substantial cognitive impairment" or "dementia" and all members of the stratified subsample (total n = 1,196), to undergo the final stage of dementia evaluation, a clinical assessment (CA).

Clinical assessment. At the CA, each participant underwent a brief physical assessment (height, weight, review of systems), a standardized neurologic examination and blood pressure measurement, a videotaped brief mental status/neurologic examination, and a 1-hour battery of neuropsychological tests that included the battery developed by the Consortium to Establish a Registry for AD (CERAD)[15] and additional measures (see below). An informant provided a clinical, medical and family history and ratings of impairment in daily functioning using the Dementia Severity Rating Scale (DSRS).[16] The DSRS reviews severity of impairment across a range of cognitive abilities, community and home activities, self-care, and mobility. In some cases, we interviewed multiple informants and ordered medical records to obtain the most complete information. During neuropsychological testing, the examiner sometimes deviated from standardized administration protocol because of a participant's sensory or motor impairments. The present analysis considers only participants who required little or no deviation from the standard protocol (e.g., altering the mode of administration from a visual to an auditory modality).

Clinical classification of dementia. A geropsychiatrist, neuropsychologist, and members of the clinical team reviewed all CA data at an initial diagnostic conference. Specifically reviewed were the clinical history (detailing the presence of cognitive symptoms and any decline in functioning), medical history, family history (reviewed for the presence of neurodegenerative dementia in first-degree relatives), responses on the DSRS, physical and neurologic examination, and neuropsychological test results as interpreted by a neuropsychologist. Infrequently, these diagnosticians questioned the validity of an informant report or noted discrepancies between such reports and neuropsychological test performances. In these circumstances, the team sought explanations so that one or both could be discounted. The diagnostic team did not review formalized ratings of cognitive impairment, but they were otherwise free to consider all information. The team assigned working clinical diagnoses of dementia that followed DSM-III-R criteria,[1] except that they did not require impairment in both short- and long-term memory.[9] They assigned differential diagnoses according to criteria listed elsewhere,[17] and rated dementia severity using the Clinical Dementia Rating scale (CDR).[18] Participants with suspected dementia were asked to undergo examination by a Board-certified geropsychiatrist and to complete an MRI scan and standard laboratory tests.[9]

Physician examinations and neuroimaging were obtained when possible. A panel of experts consisting of two geropsychiatrists, a Board-certified neurologist, two neuropsychologists, and a senior cognitive neuroscientist then reviewed all data from participants with suspected dementia. Their consensus diagnoses, and the initial working diagnoses for other nondemented individuals, served as the final "gold standard" diagnoses for the present analyses.

Psychometric classification of dementia. Before we classified participants using psychometric algorithms, we developed normative standards for each test. We applied reverse sample weights to 838 of the 960 members of the weighted stratified subsample who completed testing. This procedure generated a hypothetical sample of 4,903 that was matched to the Cache County elderly population on age, sex, and genotype at *APOE* (the deviation from 5,092 resulting from refusal of *APOE* genotyping by 160 participants and from rounding error). We next categorized the hypothetical population into two groups with "high" ($\geq$12 years) and "low" ($<$12 years) levels of education. The smaller number with less education was then stratified by age alone (66 to 85; 86 to 106), whereas others were stratified by both age (66 to 75; 76 to 85; 86 to 102) and sex. We imputed a score of 0 (or maximum time for timed tests) for all participants who could not attempt a test because of substantial cognitive impairment. Because some of the test distributions deviated from normality, we used percentiles (rather than standard deviations of a specified distribution) to identify cut-points designating impairment. Scores at or below the 7th percentile were considered to reflect borderline impairment.

We assessed six cognitive domains: immediate memory (Trial 3 of CERAD Word List,[15] Logical Memory I of the Wechsler Memory Scale-Revised,[19] Benton Visual Retention Test [BVRT])[20]; delayed memory (delayed recall trials of CERAD Word List,[15] Constructional Praxis,[15] Logical Memory)[19]; language (CERAD Naming and Animal Fluency,[15] Controlled Oral Word Association Test [COWAT])[21]; visuospatial ability (CERAD constructional praxis)[15]; executive functions (Trail Making Tests A and B,[22] Symbol Digits Modalities Test [SDMT])[23]; and intelligence (Shipley Vocabulary Test).[24] Impairment in daily functioning was coded if there was impairment in either the DQ interview or the DSRS. For this purpose, we adapted the DQ interview responses (from Stage 2 of dementia screening) into a quantitative measure. We recoded item scores on cognitive or functional abilities into dichotomous ratings of "impaired" (1 point) or "not impaired" (0 points) and summed across all items. After plotting a receiver operating characteristic (ROC) curve of cut-off scores against the clinical dementia diagnoses, we selected a cut-point of 7/8 on this "Quantitative DQ." We used a cut-point of 21/22 on the DSRS informant ratings, following published guidelines for the DSRS discriminating between questionable/mild from moderate dementia.[16]

Two algorithms. We applied modified DSM-III-R criteria in two algorithms. First, we used a "pure" psychometric model using only neuropsychological test data to show impairment in memory (either immediate or delayed) and one other cognitive domain. The second algorithm added the criterion of functional impairment, as indicated by either the Quantitative DQ or the DSRS. Both algorithms required only one completed test to code impairment in

Downloaded from www.neurology.org by Renzo Rozzini on July 16, 2007

Table 1 *Seventh percentile scores and sample sizes for normative groups on neuropsychological tests*

Age	Al Flu	BNT	WL Im	WL D	C P C	C P D	TM A	TM B	LM Im	LM D	BVRT Corr	BVRT Err	COWA	SDMT	Shipley
<12 Years of education															
66–85	9	12	3	1	8	1	93	274	6	3	1	18	11	15	14
n	75	74	75	75	72	72	60	47	70	68	64	64	73	54	66
86–106	5	8	2	0	8	0	257	300	3	0	0	23	8	6	12
n	89	67	83	83	66	66	54	26	79	77	53	53	77	38	63
≥12 Years of education—(M/F)*															
66–75	11/14*	13	6	4/5*	9	6	68/54*	164/204*	12/13*	5/8*	3/4*	12	16/20*	25/27*	25/23*
Men, n	83	82	83	83	83	83	80	77	83	83	83	83	83	82	83
Women, n	88	88	89	89	89	89	85	83	86	85	87	87	87	87	88
76–85	10	11/12*	5	1/2*	8	2	98/76*	296/258*	8/10*	1/4*	2/3*	17/14*	17/16*	16/20*	18/19*
Men, n	118	117	116	116	112	112	100	85	109	106	108	108	112	93	114
Women, n	140	136	137	137	133	133	121	101	130	127	124	124	135	110	132
86–96	6/4*	11/8*	1/2*	0	7/6*	0	141/222*	300	5	3/0*	0	23	12/8*	12/14*	20/19*
Men, n	58	51	56	55	46	46	39	29	49	49	42	42	53	31	49
Women, n	132	100	125	125	101	101	77	48	109	103	83	83	120	60	96

The table displays cut-scores for designating impairment (7th percentile) for individuals according to age, education, and gender.

* Where cut-points differ between men and women in the high educational group. n refers to the original sample sizes before reverse sample weighting in each normative group for each test completed in the neuropsychological battery. The variation in the sample size reflects differences in the numbers of participants who completed the various test procedures.

Al Flu = Animal Fluency; BNT = Boston Naming Test; WL = Word List Memory; Im = Immediate; D = Delay; CP = Constructional Praxis; C = Copy; TM = Trail Making; LM = Logical Memory; BVRT = Benton Visual Retention Test; Corr = Correct; Err = Error; COWA = Controlled Oral Word Association; SDMT = Symbol Digits Modalities Test.

any given cognitive domain. Because no inference can be made from missing data, however, a rating of "not impaired" required that all tests in a given cognitive domain be completed.

We examined the agreement between the two algorithms and the "gold standard" clinical diagnoses by calculating percent agreement and Cohen's kappa statistic. We used the following guidelines provided by Fleiss[25] to interpret kappa values: <0.40 = poor agreement, 0.40 to 0.75 = fair to good agreement, >0.75 = excellent agreement. To evaluate other performance characteristics of the algorithms, we calculated sensitivity (percent detection of true cases of dementia), specificity (percent detection of true noncases of dementia), positive predictive value (probability that an individual classified algorithmically "with dementia" actually has the condition), and negative predictive value (probability that an individual classified "without dementia" actually does not have the condition). We also explored the above statistics in relation to age and dementia severity as defined by CDR score ("mild" = CDR 0.5 or 1, "moderate" = CDR 2, "severe" = CDR 3, 4, 5). We applied the two algorithms to determine how well each discriminated noncases from cases of dementia with specified severity. Finally, because of the problem of missing test data, we explored the algorithms' accuracy in dementia classification when only one completed test per cognitive domain was required for a rating of "not demented."

Results. There were 1,033 individuals (435 men and 598 women) who completed a CA. Participant ages ranged from 66 to 106 (mean = 81.63, SD = 7.64). Educational attainment ranged from no formal schooling to 20 years (mean = 12.93, SD = 2.97). We identified 353 participants with clinical diagnoses of dementia and 680 without. A majority of the demented participants (246 of 353) had been examined by a geropsychiatrist.

Psychometric classification of dementia. Ten of the 1,033 participants refused all testing. Table 1 illustrates the derived cut-points for each neuropsychological test.

Pure psychometric algorithm. The algorithm relying only on psychometric test results classified 892 of the 1,023 participants (87%), missing 44 cases with and 87 cases without dementia. Table 2 shows that 447 individuals (50%) were classified as having dementia. This proportion deviated appreciably from the "gold standard" diagnoses that indicated only 296 (33%) with dementia. Overall agreement was 79%, with a kappa of 0.57 (95% CI: 0.52 to 0.62), suggesting reasonable agreement between the algorithm and the gold standard. There was a clear tendency toward overdiagnosis of dementia, however, as this algorithm produced 171 false-positives out of 191 disagreements. Of these 171, 95 (56%) had a cognitive disorder, but no dementia. Of the 20 dementia cases overlooked by the algorithm, 15 were identified neuropsychologically as having impairment in one cognitive domain. As the

Downloaded from www.neurology.org by Renzo Rozzini on July 16, 2007

Table 2 *Dementia classification*

NP algorithm	Clinical classification: Dementia	Clinical classification: No dementia	Total
Algorithm without functional measures			
Dementia	276	171	447
No dementia	20	425	445
Total	296	596	892
Algorithm with functional measures			
Dementia	243	35	278
No dementia	61	623	684
Total	304	658	962

table shows, the algorithm had good sensitivity (93%) but only a modest positive predictive value: only 62% of participants classified as "demented" actually had the condition. Specificity was also modest (71%), but negative predictive value was excellent (96%).

Psychometric tests plus functional impairment. Adding consideration of functional impairment to the algorithm resulted in improvements in all measures of association, with only a slight reduction in sensitivity. Because some participants with missing test data could now be classified as nondemented (providing they had negative scores on the DSRS or Quantitative DQ), this algorithm classified 962 participants (94% of the sample), missing 36 cases with and 25 cases without dementia. As Table 2 shows, 278 individuals (29%) were classified as having dementia, more closely matching the "gold standard" proportion and providing 90% overall agreement. Sensitivity decreased from 93% (using the pure psychometric approach) to 80%. But specificity (95%), positive predictive value (87%), and negative predictive value (91%) were excellent with this algorithm. The kappa value also suggested good to excellent agreement ($\kappa = 0.76$, 95% CI: 0.72 to 0.81). Disagreements between this algorithm and clinical diagnoses now showed that false-negatives (61) outnumbered false-positives (35). Again, the majority of the false-positives (33 out of 35) had a cognitive disorder, but no dementia. Of the 61 dementia cases overlooked by the algorithm, 55 were again identified as having impairment in one or more cognitive domain.

Performance of the algorithms across levels of dementia severity. Table 3 shows the performance characteristics of the two neuropsychological algorithms related to various levels of dementia severity. The modest overall agreement (74% to 75%) with the pure psychometric approach did not change appreciably with dementia severity. Although the sensitivity of this algorithm increased with dementia severity (85%→100%), the proportion of diagnostic false-positive participants changed little. Specificity was modest at 71% and remained constant, reflecting the fact that the same nondemented cohort was considered at each level of dementia severity.

The algorithm requiring functional impairment performed more predictably, achieving the best overall agreement with the "gold standard" in severe dementia (95%, $\kappa = 0.82$, 95% CI: 0.76 to 0.88). Agreement decreased to the satisfactory range for questionable or mild dementia (89%, $\kappa = 0.57$, 95% CI: 0.49 to 0.65). Generally, the algorithm accurately identified participants without dementia (specificity = 95%, negative predictive value = 92 to 100%) but was less apt in detecting participants with mild dementia (sensitivity = 59%, positive predictive value = 69%).

Performance of the algorithms across participant age. With increasing age, both algorithms were able to classify fewer participants. This was particularly true in the oldest age group, owing mostly to sensory or motor impairments (present in 65% of those not classified). Table 4 shows that the pure psychometric model produced moderate agreement with the gold standard across the age ranges. Surprisingly, agreement was best in participants who were 90 years of age and above (88% agreement, $\kappa = 0.75$, 95% CI: 0.64 to 0.87). Sensitivity, specificity, and positive and negative predictive values were also good for this group, with

Table 3 *Agreement, sensitivity, specificity, positive predictive value, and negative predictive value by CDR severity level*

Clinical Dementia Rating (CDR) level sample size*	Method	Number classified	Percent classified	Percent agreement	Kappa	Sensitivity, %	Specificity, %	PPV, %	NPV, %
Questionable–mild (0.5–1.0) (n = 138)	Algorithm without functional measures	720	88	74	0.38	85	71	38	96
	Algorithm with functional measures	789	96	89	0.57	59	95	69	92
Moderate (2.0) (n = 92)	Algorithm without functional measures	670	86	74	0.35	99	71	30	100
	Algorithm with functional measures	734	95	94	0.74	92	95	67	99
Severe–terminal (3.0–5.0) (n = 116)	Algorithm without functional measures	694	87	75	0.41	100	71	36	100
	Algorithm with functional measures	755	94	95	0.82	99	95	73	100

* Sample sizes reflect numbers of demented participants (identified by the clinical method) at specified CDR levels only. To calculate the above measures of association, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), 683 nondemented participants (also identified by the clinical method) were included in the above samples.

Downloaded from www.neurology.org by Renzo Rozzini on July 16, 2007

Table 4 *Agreement, sensitivity, specificity, positive predictive value, and negative predictive value by participant age*

Method	Age group and sample sizes			
	66–69 (n = 64)	70–79 (n = 333)	80–89 (n = 445)	90 and over (n = 188)
Algorithm without functional impairment				
Number classified	64	320	386	122
Percent classified	100	96	87	65
Percent agreement	83	74	79	88
Kappa	0.53	0.46	0.59	0.75
Sensitivity, %	100	97	92	92
Specificity, %	80	67	71	84
Positive predictive value, %	45	47	67	85
Negative predictive value, %	100	99	93	91
Algorithm with functional impairment				
Number classified	64	323	415	160
Percent classified	100	97	93	85
Percent agreement	97	93	88	88
Kappa	0.88	0.79	0.74	0.73
Sensitivity, %	100	88	78	73
Specificity, %	96	94	94	98
Positive predictive value, %	82	81	88	96
Negative predictive value, %	100	96	88	83

values between 84 and 92%. Performance was only fair for participants in the other age groups (percent agreement 74% to 83%, $\kappa = 0.46$ to 0.59).

The algorithm that included functional measures exhibited improved performance in almost all age groups. Agreement was highest in the youngest age group (97% agreement, $\kappa = 0.88$, 95% CI: 0.72 to 1.0) and decreased slightly with age. Even at ages 90 and above, however, agreement was still quite good (88%, $\kappa = 0.73$, 95% CI: 0.63 to 0.84). Sensitivity of this algorithm decreased modestly with age to 73% in those 90 years of age or older. All other measures of this algorithm's performance remained satisfactory at 78% or better across all age groups.

Algorithms with relaxed test completion criteria. In an attempt to classify more participants, we explored the effect of relaxing the algorithmic criteria for "not demented" by requiring only one completed test per cognitive domain. Doing so allowed us to classify an additional 88 participants with the pure psychometric algorithm (total = 980) and 29 with the functional algorithm (total = 991). Examination of the newly classified participants shows 83% agreement between the pure psychometric algorithm and the "gold standard" (*cf* 79% with stringent criteria) and 69% when considering functional impairments (*cf* 90% with stringent criteria). The modification did not appreciably affect the algorithms' overall performance where the psychometric algorithm still achieved 79% agreement ($\kappa = 0.57$ 95% CI: 0.52 to 0.62) with the clinical diagnoses, and the algorithm considering functional measures achieved 89% agreement ($\kappa = 0.75$ 95% CI: 0.70 to 0.79). Kappa values could not be calculated for the subset of participants now classified by the modified algorithms because this adaptation would detect only additional nondemented cases. Although uncorrected for chance, the level of agreement with the modified algorithms suggests that algorithms relying on a single test to represent a cognitive domain may be useful for those participants with sensory or motor impairments, albeit at some cost to diagnostic accuracy.

Discussion. Generating psychometric norms in a large population sample and comparing algorithmic neuropsychological diagnoses of "dementia" with clinical diagnoses, we demonstrated good-to-excellent agreement. Although the diagnoses generated by a purely psychometric approach showed fair agreement with gold standard clinical diagnoses, agreement and classification accuracy improved considerably when measures of functional ability were considered. This result corroborates the importance of clinical information and functional assessments in the diagnostic process.[8,26] The improvement in classification accuracy was also evident across levels of dementia severity and age.

The agreement between the algorithms and the clinical diagnoses may not be entirely unexpected because the latter were made with the aid of neuropsychological testing, and one measure of functional impairment (the "Quantitative DQ") was normed in the present sample against the clinical diagnoses. However, we would emphasize that in the clinical method, interpretations of neuropsychological test results were considered along with other information in formulating the diagnoses, which represented the "best achievable" using state-of-the-art procedures. No features other than neuropsychological testing

Downloaded from www.neurology.org by Renzo Rozzini on July 16, 2007

and structured measures of functional impairment were incorporated in the algorithmic diagnoses.

Our results are consistent with previous work showing good agreement of algorithmic and physician's diagnoses.[4,5] Discrepancies between diagnoses in the prior studies were attributable in part to physicians' greater reliance on mental status testing and measures of functional ability. There was also a tendency of the algorithms to classify participants with low educational attainment as being more impaired, necessitating an educational correction for test scores.[4] We attempted to address both issues by incorporating measures of functional impairment in our algorithm and by stratifying normative groups not only by age but also by education.

There are clear limitations in the use of neuropsychological algorithms. First, they require that participants complete a minimum number of tests. In this sample, 6% to 13% of participants could not be classified because of substantial sensory or motor impairments. Modifying the algorithm to require fewer tests helped to alleviate this problem, but at a cost in diagnostic accuracy. Second, the applicability of any algorithm is limited to the specific tests and the population on which it was validated. Some revalidation would probably be needed before the algorithm could be used with different tests or populations. Third, all current criteria for dementia require a decline from a previous level of functioning. Clinical diagnoses can rely (principally) on the history to document such decline but, in the absence of longitudinal assessment, neuropsychological measures can only infer a decline from the pattern of impairments present. Fourth, errors or misclassifications can be expected with algorithms. In this work, misclassification rates ranged from 10 to 21%. For the majority of these cases, both methods detected the presence of a cognitive disorder, and most disagreements with the clinical diagnoses occurred in the mild dementia group. Finally, although the rates of agreement between the algorithms and the "gold standard" appear promising, percent agreement and kappa values must be interpreted with consideration of the consequences of misclassification.

Notwithstanding these limitations, we suggest that algorithmic methods may be a useful adjunct in dementia classification. With its strengths of objectivity, stability, and reliability, the algorithmic approach can augment clinical evaluations by acting as an objective means to maintain a uniform, stable threshold for dementia across time and across diagnosticians. Maintenance of a stable threshold is important in epidemiologic studies so that prevalence and incidence rates are not distorted by "drift" in the application of diagnostic criteria. Conceivably, algorithms may also be useful when evaluating ambiguous or borderline cases where it is difficult for clinicians to agree whether criteria for dementia are met. Here, the objective nature of an algorithm may provide a standardized answer, or at a minimum, an additional piece of information to be considered in the diagnostic process. Another obvious use for algorithms is the standardization of the diagnostic threshold for dementia across sites, for instance in multicenter prevention trials that observe at-risk participants longitudinally for the development of dementia.

The neuropsychological algorithms developed for this study show promise for these applications. In conjunction with the results of previous investigators,[4,5] our findings suggest that neuropsychological algorithms may add valuably to the process of dementia classification. If the present results can be replicated in other samples, we would suggest that algorithmic approaches be considered as desirable (along with clinical diagnoses) in longitudinal epidemiologic studies of dementia. Incorporating longitudinal observations in algorithmic models will undoubtedly be the next step where a demonstrable decline in cognitive test scores will be included as one of several criteria for the diagnosis of a degenerative dementia. We plan to examine such longitudinal follow-up data of cognitively impaired individuals in the Cache County study. The results should speak to the utility of algorithmic methods in longitudinal work.

Acknowledgment

The authors thank Drs. James Anthony and Yaakov Stern for helpful discussions. Drs. Brenda Plassman, David Steffens, and James Burke contributed their expertise at the consensus diagnostic conferences. Dr. Steffens also conducted many of the geropsychiatrist's examinations. Andrea Hart, Joslin Werstak, Carol Leslie, Barbara Gau, Tiffany Newman, and Tony Calvert provided expert clinical and technical help. The authors also thank Dr. Albert Heyman and colleagues for permission to use the CERAD neuropsychological test battery. Western Psychological Services and the Psychological Corporation discounted their fees for use of the Vocabulary Test from the Shipley Institute of Living Scale and the Logical Memory Subtests I and II of the Wechsler Memory Scale, Revised, respectively. *APOE* genotypes were determined by the neurogenetics laboratory of the Bryan Alzheimer's Disease Center, Duke University.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd ed., rev. Washington, DC: American Psychiatric Association, 1987.
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th ed. Washington, DC: American Psychiatric Association, 1994.
3. World Health Organization. The ICD-10 classification of mental and behavioral disorders. Diagnostic criteria for research. Geneva, Switzerland: World Health Organization, 1993.
4. Stern YS, Andrews H, Pittman J, et al. Diagnosis of dementia in a heterogeneous population: development of a neuropsychological paradigm-based diagnosis of dementia and quantified correction for the effects of education. Arch Neurol 1992;49: 453–460.
5. Pittman J, Andrews H, Tatemichi T, et al. Diagnosis of dementia in a heterogeneous population. A comparison of paradigm-based diagnosis and physician's diagnosis. Arch Neurol 1992;49:461–467.
6. Snowden DA, Greiner LH, Mortimer JA, Riley KP, Greiner PA, Markesbery WR. Brain infarction and the clinical expression of Alzheimer disease: the nun study. JAMA 1997;277: 813–817.

Downloaded from www.neurology.org by Renzo Rozzini on July 16, 2007

7. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. Science 1989;243:1668–1674.
8. Morris JC, McKell DW, Storandt M, et al. Very mild Alzheimer's disease: Informant-based clinical, psychometric, and pathologic distinction from normal aging. Neurology 1991;41: 469–478.
9. Breitner JCS, Wyse BW, Anthony JC, et al. APOE E4 count predicts age when prevalence of Alzheimer's disease increases, then declines. Neurology 1999;53:321–331.
10. Teng EL, Chui HC. The modified mini-mental state (3MS) examination. J Clin Psychiatry 1987;48:314–318.
11. Jorm AF, Jacomb PA. The informant questionnaire on cognitive decline in the elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms. Psychol Med 1989; 19:1015–1022.
12. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele E4 with late-onset familial and sporadic Alzheimer's disease. Neurology 1993;43:1467–1472.
13. Silverman JM, Breitner JCS, Mohs RC, et al. Reliability of the family history method in genetic studies of Alzheimer's disease and related dementias. Am J Psychiatry 1986;143:1279–1282.
14. Kawas C, Segal J, Stewart WF, Corrada M, Thal LJ. A validation study of the dementia questionnaire. Arch Neurol 1994; 51:901–906.
15. Welsh KA, Butters N, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V: a normative study of the neuropsychological battery. Neurology 1994;44:609–614.
16. Clark CM, Ewbank DC. Performance of the Dementia Severity Rating Scale: a caregiver questionnaire for rating severity in Alzheimer Disease. Alzheimer Dis Assoc Disord 1996;10: 31–39.
17. Breitner JCS, Welsh KA, Gau BA, et al. Alzheimer's disease in the National Academy of Sciences–National Research Council Registry of Aging Twin Veterans III: detection of cases, longitudinal results, and observations on twin concordance. Arch Neurol 1995;52:763–771.
18. Hughes CP, Berg L, Danziger WL, et al. A new clinical scale for the staging of dementia. Br J Psychiatry 1982;140:566–572.
19. Wechsler D. Wechsler Memory Scale-Revised. New York: Psychological Corp., 1987.
20. Benton AL. Benton Visual Retention Test. 5th ed. New York: Psychological Corp., 1992.
21. Benton AL, Sivan A, de Hamsher KS. Multilingual Aphasia Examination. Iowa City, IA: AJA Associates, 1994.
22. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 1958;8:271–276.
23. Smith A. Symbol Digit Modalities Test. Los Angeles: Western Psychological Services, 1973.
24. Shipley WS. Shipley Institute of Living Scale. Los Angeles: Western Psychological Services, 1967.
25. Fleiss JL. Statistical methods for rates and proportions. 2d ed. New York: John Wiley & Sons, 1981.
26. Steffens DC, Welsh KA, Burke JR, et al. Diagnosis of Alzheimer's disease in epidemiologic studies by staged review of clinical data. Neuropsychiatry Neuropsychol Behav Neurol 1996;9:107–111.

Downloaded from www.neurology.org by Renzo Rozzini on July 16, 2007

Dementia diagnoses from clinical and neuropsychological data compared: The Cache County study
J. T. Tschanz, K. A. Welsh-Bohmer, I. Skoog, N. West, M. C. Norton, B. W. Wyse, R. Nickles and J. C. S. Breitner
Neurology 2000;54;1290-1296

This information is current as of July 16, 2007

Updated Information & Services	including high-resolution figures, can be found at: http://www.neurology.org/cgi/content/full/54/6/1290
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://www.neurology.org/misc/Permissions.shtml
Reprints	Information about ordering reprints can be found online: http://www.neurology.org/misc/reprints.shtml

Downloaded from www.neurology.org by Renzo Rozzini on July 16, 2007