

## Reducing case ascertainment costs in U.S. population studies of Alzheimer's disease, dementia, and cognitive impairment—Part 1

David R. Weir<sup>a,\*</sup>, Robert B. Wallace<sup>b,c</sup>, Kenneth M. Langa<sup>a,d,e</sup>, Brenda L. Plassman<sup>f</sup>,  
Robert S. Wilson<sup>g,h,i</sup>, David A. Bennett<sup>g,h</sup>, Ranjan Duara<sup>j,k,l,m</sup>, David Loewenstein<sup>j,m</sup>,  
Mary Ganguli<sup>n,o</sup>, Mary Sano<sup>p,q,r</sup>

<sup>a</sup>Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

<sup>b</sup>Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA, USA

<sup>c</sup>Department of Internal Medicine, College of Medicine, University of Iowa, Iowa City, IA, USA

<sup>d</sup>Division of General Medicine, Department of Medicine, School of Medicine, University of Michigan, Ann Arbor, MI, USA

<sup>e</sup>Veterans Affairs Center for Practice Management and Outcomes Research, Ann Arbor, MI, USA

<sup>f</sup>Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA

<sup>g</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

<sup>h</sup>Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA

<sup>i</sup>Department of Behavioral Sciences, Rush University Medical Center, Chicago, IL, USA

<sup>j</sup>Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami, FL, USA

<sup>k</sup>Department of Medicine, Miller School of Medicine, University of Miami, Miami, FL, USA

<sup>l</sup>Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA

<sup>m</sup>Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, FL, USA

<sup>n</sup>Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

<sup>o</sup>Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

<sup>p</sup>Department of Psychiatry, Mount Sinai Medical Center, New York, NY, USA

<sup>q</sup>Alzheimer's Disease Center, Mount Sinai Medical Center, New York, NY, USA

<sup>r</sup>Bronx Veterans Administration Hospital, Bronx, NY, USA

### Abstract

Establishing methods for ascertainment of dementia and cognitive impairment that are accurate and also cost-effective is a challenging enterprise. Large population-based studies often using administrative data sets offer relatively inexpensive and reliable estimates of severe conditions including moderate to advanced dementia that are useful for public health planning, but they can miss less severe cognitive impairment which may be the most effective point for intervention. Clinical and epidemiological cohorts, intensively assessed, provide more sensitive detection of less severe cognitive impairment but are often costly. In this article, several approaches to ascertainment are evaluated for validity, reliability, and cost. In particular, the methods of ascertainment from the Health and Retirement Study are described briefly, along with those of the Aging, Demographics, and Memory Study (ADAMS). ADAMS, a resource-intensive sub-study of the Health and Retirement Study, was designed to provide diagnostic accuracy among persons with more advanced dementia. A proposal to streamline future ADAMS assessments is offered. Also considered are algorithmic and Web-based approaches to diagnosis that can reduce the expense of clinical expertise and, in some contexts, can reduce the extent of data collection. These approaches are intended for intensively assessed epidemiological cohorts where goal is valid and reliable case detection with efficient and cost-effective tools.

© 2011 The Alzheimer's Association. All rights reserved.

### Keywords:

Alzheimer's disease; Dementia; Mild cognitive impairment; Cognitive impairment not dementia; Diagnostic algorithms; Cognition; Epidemiology; Screening; Technology; Education; Race; Ethnicity

All of the authors contributed equally to this article.

\*Corresponding author. Tel.: 734-615-4694; Fax: 734-647-1186.

E-mail address: [dweir@isr.umich.edu](mailto:dweir@isr.umich.edu)

## 1. Introduction

Alzheimer's disease (AD) is an enormous public health problem that is expected to markedly increase in the coming decades with the aging of the post-World War II generation in the United States and many other countries. Developing strategies to delay the onset of the signs and symptoms of AD is critical for disease prevention [1,2]. Although longitudinal cohort studies and other population surveys have contributed much to the current knowledge, very large clinical trials will likely be needed in the future. In fact, there will be upward cost pressures on the budgets of the longitudinal cohort studies, the population surveys, and the clinical trials, particularly those emphasizing preclinical and mild impairment, as more biomarker technology is used to assist in the diagnosis of AD and other age-related causes of cognitive impairment, and as sample sizes increase to accommodate the need to study subjects aged <65 years in addition to those aged  $\geq 65$  years. Another major contributor to the costs of cohort studies and clinical trials is the costs related to clinical diagnoses. Several groups across the country have investigated a variety of approaches to rein in such costs without compromising reliability or validity.

This article and also another [see Ref 3] pertain to reducing case ascertainment costs for large cohort studies, and are applicable to primary and secondary prevention trials. Sections 2 and 3 pertain to the Health and Retirement Study (HRS), a nationally representative cohort study of health, retirement, and aging. In section 2, Weir draws on his experience with the HRS to emphasize the need for cost-effective methods of case ascertainment in population-based studies of prevalence and burden of disease. In section 3, Wallace, Langa, and Plassman present a possible alternative approach to the current methods of dementia assessment within the HRS, by emphasizing more complete data collection by telephone and other electronic means, and more intensive acquisition of available clinical records. Sections 4, 5, and 6 are more general and focus on aspects of the diagnostic process itself. In section 4, Wilson and Bennett illustrate utility of a cost-efficient algorithmic approach that combines computer-based summaries of neuropsychological performance tests with expert clinical judgment to generate diagnoses of cognitive impairment, dementia, and AD, using data from the Rush Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP). In section 5, Duara and Loewenstein demonstrate the method by which a valid consensus diagnosis (ConsDx) of dementia and predementia states can be achieved with greater reliability and considerably reduced effort and cost by avoiding the traditional consensus deliberation and using instead a simple algorithm that combines an independent neuropsychological diagnosis (NPDx) and a functional assessment by a clinician. In section 6, Ganguli offers a novel, Web-based approach to diagnosis and diagnostic consensus. This approach avoids the logistic costs of, and saves time over, the typical live “con-

sensus conference” of experts; it adds standardization, allows reliability to be monitored, and provides opportunities for finer-grained analyses of the components of expert diagnosis. Finally, in section 7, the discussion, Sano provides a synthesis of the material laid out in the earlier sections, and offers some perspective of her own.

## 2. Need for cost-effective, case-ascertainment methods in surveys of dementia: Experience of the HRS

Population surveys play an important role in improving scientific understanding of the causes, consequences, and morbidity levels of AD, dementia, and cognitive impairment. These surveys are particularly valuable for understanding the context in which the disease occurs and progresses, including its interaction with other conditions and its effect on the families of those affected. However, survey planners must be concerned about the participation of the cognitively impaired if they aim to represent adequately that part of the older population. More than any other condition or characteristic, cognitive decline directly affects the ability of a potential subject to participate in a survey because—by their nature—surveys are cognitively demanding conversations.

An important issue in finding cost-effective, case-ascertainment methods is deciding what should be measured. AD, dementia, and cognitive impairment have multiple causes and are complex diagnostic categories involving social functioning as well as cognitive ability. Identifying persons with outright dementia and sub-categorizing those attributable to AD would lead to one sort of study design. A different design is required for potential identification of milder cognitive impairments and/or all cases of AD (mild to severe). Because of the much larger number of persons with milder forms of cognitive impairment, designs to capture this group are necessarily more expensive unless the assessment itself can be made much less costly.

In many surveys, clinical diagnosis involves a combination of cognitive testing, informant reporting, and clinical judgment. Survey planners must decide the amount of weight to place on each of those. Clinical judgment is the most difficult and costly element to incorporate into surveys. Informant reports raise concerns about inter-rater reliability (i.e., some informants may report differently or have different knowledge of the subject). Cognitive testing at a single point in time also faces issues of reliability, especially if variability in cognitive performance increases with cognitive decline. Moreover, some subjects may not wish to be tested precisely because they know their abilities have declined, resulting in a biased sample.

The HRS started in 1992 with a cohort of individuals aged 51 to 61 years and was supplemented in 1993 with a cohort of individuals aged  $\geq 70$  years, then referred to as the Asset and Health Dynamics among the Oldest Old Study. Both studies included a cognitive battery [4]. New cohorts were added in 1998 to make the combined study representative

of the U.S. population born before 1948, and in 2004 another new cohort made it representative of the U.S. population born before 1954. In this section, the HRS will be used to comment on three critical case-ascertainment issues faced in surveys of AD, dementia, and cognitive impairment in the United States. The issues include nursing home coverage, selective nonparticipation because of cognitive difficulties, and determination of diagnostic status. Each issue relates to potential bias of survey results. For the third issue, cost implications are also briefly considered.

In some community-based studies, especially cross-sectional studies, the exclusion of nursing home residents is done for practical reasons, but at the risk of bias, because that group may include a substantial part of the cognitively impaired population. The HRS at baseline did exclude nursing home residents. However, currently the HRS is a longitudinal study, following all participants through the end of life. Sample members are followed even if they subsequently enter nursing homes. As a result, the HRS now fully represents nursing home residents.

A second major contributor to bias is selective nonparticipation by cognitively impaired elderly people. The HRS interview is cognitively demanding both for its length ( $\geq 75$  minutes) and the complexity of its multi-disciplinary content. To maintain coverage of the cognitively impaired, the HRS makes use of proxy interviews. If a sample member is unable or unwilling to participate in the HRS interview, a proxy is sought who can answer on his or her behalf. Most often this is a spouse or, in the absence of a spouse, another family member. The importance of this can be seen in Fig. 1 which maps response rate in the first follow-up wave against cognitive score in the starting (baseline) wave. The cognitive score is a measure of total recall (sum of words recalled at immediate plus delayed recall from a list of 10). Including the proxy interviews in HRS, there is little relationship between cognitive score and continued participation. The English Longitudinal Study of Aging (ELSA) is, by design, quite similar to the HRS in its content and

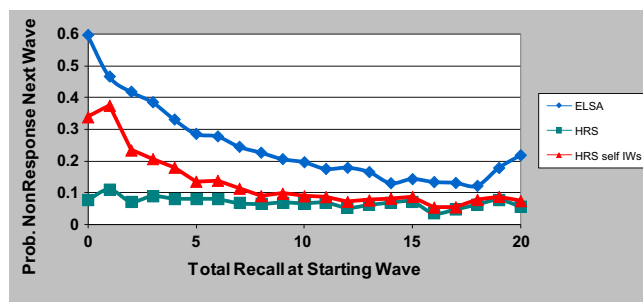


Fig. 1. The role of proxy interviewing in maintaining participation of the cognitively impaired. The Health and Retirement Study versus the English Longitudinal Study of Aging. “Total Recall at Starting Wave” refers to the sum of words recalled at immediate plus delayed recall from a list of 10. “Self-IWs” means only interviews taken with the respondents are included (i.e., no proxy respondents). For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.

age coverage. However, proxies are generally not used. In ELSA, the rates of nonresponse are very much related to previous cognitive measures. Finally, if the proxy interviews from HRS were excluded (as if there had been no response), the HRS and ELSA curves look quite similar.

The third case-ascertainment issue is the difficulty of determining diagnostic status. Many chronic conditions are commonly diagnosed in the healthcare system, and survey participants can and do report them accurately. However, cognitive impairment is not systematically diagnosed, and many people with diagnosable impairment either do not know or do not report it. The advantage of proxy interviews is that persons with cognitive impairment do not disappear from the study. Many aspects of their lives, including their other health conditions, use of medical services and informal care, and economic circumstances, can be reported by proxies. What is lost however, is the direct measurement of their cognitive abilities. The HRS uses several proxy-reported measures, including the Jorm Informant Questionnaire on Cognitive Decline in the Elderly, to assess the cognitive status of a sample member through a proxy interview [5].

Both the HRS cognitive score for self-respondents and the Informant Questionnaire on Cognitive Decline in the Elderly for proxied cases provide a useful ranking of participants by cognitive status that can be used for important questions like the costs to the family of formal and informal care for persons with cognitive impairment [6,7]. The survey measures are, however, far short of a medical diagnosis of AD or dementia. To bridge that gap, an HRS supplemental study known as the Aging, Demographics, and Memory Study (ADAMS) was undertaken. In ADAMS, the HRS cognitive and proxy measures were used to stratify the sample of subjects aged  $>70$  years by cognitive status and then participants were selected within each stratum to receive a comprehensive in-home evaluation leading to a research diagnosis of dementia, cognitive impairment not dementia (CIND), or no cognitive impairment (NCI) [8]. ADAMS found a prevalence of 13.7% demented and 22.2% CIND in the population age  $\geq 71$  years [9,10]. AD was identified as a subtype of dementia at a prevalence of 9.7%.

The ADAMS diagnostic categories correspond to distinct levels of morbidity and this can be noted in outcomes following the ADAMS assessment. At the next HRS wave in 2004, daily care hours were 0.7 for normal cognition, 2.8 for CIND, and 7.3 for the demented. Mortality risk was 2.6 times higher for CIND and 5.7 times higher for demented versus the normals. The CIND group, which is roughly twice as large as the demented group, is thus very different from both the normal cognition group and those with diagnosed dementia.

ADAMS provides useful insight into the strengths and limitations of self-report in the HRS. Of those diagnosed with dementia in ADAMS, only 40% had previously reported in HRS that they had been diagnosed with a memory-related disease (and only 4% of those with CIND).

During the ADAMS assessment, only 47% of individuals with dementia and 7% of individuals with CIND reported (through proxy respondents) that they had seen a doctor for memory problems. This may reflect reporting bias at least as much as it does under-diagnosis or under-treatment. Another study found 85% of ADAMS cases diagnosed as demented had a Medicare claim with a dementia-related diagnosis [11]. The percentage was much lower for individuals with CIND.

At the present time, with limited medical treatment options available for cognitive impairment, the use of self-report survey measures is clearly inadequate. By contrast, survey-based cognitive measurements can be valuable. The HRS-based cognitive status strata used to draw the ADAMS sample had a strong correlation with the eventual ADAMS diagnosis (ADAMS assessments were blinded to the sample strata), accounting for half of the variance in diagnosis. There is an enormous difference in cost between adding a few cognitive measures to a survey like HRS, and conducting a rigorous diagnostic protocol like ADAMS. The challenge ahead is to find more cost-effective designs (see section 3).

It seems likely that the way forward for case ascertainment in surveys will involve a multi-stage screening process, or what might also be termed an adaptive testing process. Many persons can be determined to be unimpaired with relatively minimal testing. Many definite dementia cases may be identifiable through Medicare records or relatively inexpensive follow-up interviews with informants. The difficult cases are those at the borderlines between normal and mild impairment and between mild impairment and dementia. Those are the cases for which the most extensive testing will be needed for definitive ascertainment.

### 3. An alternative protocol to study dementia occurrence in the HRS

Over the past decade, the HRS has included a population sub-study of AD, dementia, and cognitive impairment among persons aged  $\geq 71$  years [12]. The sub-study, ADAMS (see section 2), has produced important findings on the prevalence and incidence of AD, vascular dementia, dementia, and cognitive impairment in the U.S. population [9,10]. In addition, data are emerging from ADAMS on the antecedents and outcomes of the dementia syndrome [13–16]. ADAMS, with its national coverage and comprehensive case-finding approach, is unique in the United States. Other U.S. population-based studies with comparable aims and credible case-finding strategies are confined to local communities.

The ADAMS approach to case-finding includes home visits by a nurse and neuropsychological technician, who perform, among other things, a clinical and cognitive history from both participants and proxy respondents, and extensive neuropsychological and neurological testing. The methods have been previously described [8], and the data from

ADAMS are available on the HRS Web site ([www.hrsonline.isr.umich.edu](http://www.hrsonline.isr.umich.edu)). The data collected at the home visits allow an in-depth assessment of cognitive function and illnesses, although without neuro imaging or a standard battery of laboratory tests.

Because ADAMS is a subsample of the HRS, it has the great advantage of having a nationally-referent study population, as well as possessing a great deal of earlier participant information, including dementia screening instrument findings. However, despite these important strengths, the methodology of ADAMS is resource-intensive, requiring a substantial amount of time to arrange the home visits in advance. There are usually incumbent airfares and hotel and staff costs to conduct the home-based study evaluations. In addition, perhaps not unexpected because of advanced age and frequent cognitive impairment of participants, the ADAMS home visit participation rate of 56% [9] was less than hoped. Thus, despite the detailed information collected in ADAMS, more efficient methods for widely distributed populations would allow a lower cost alternative when resources are scarce.

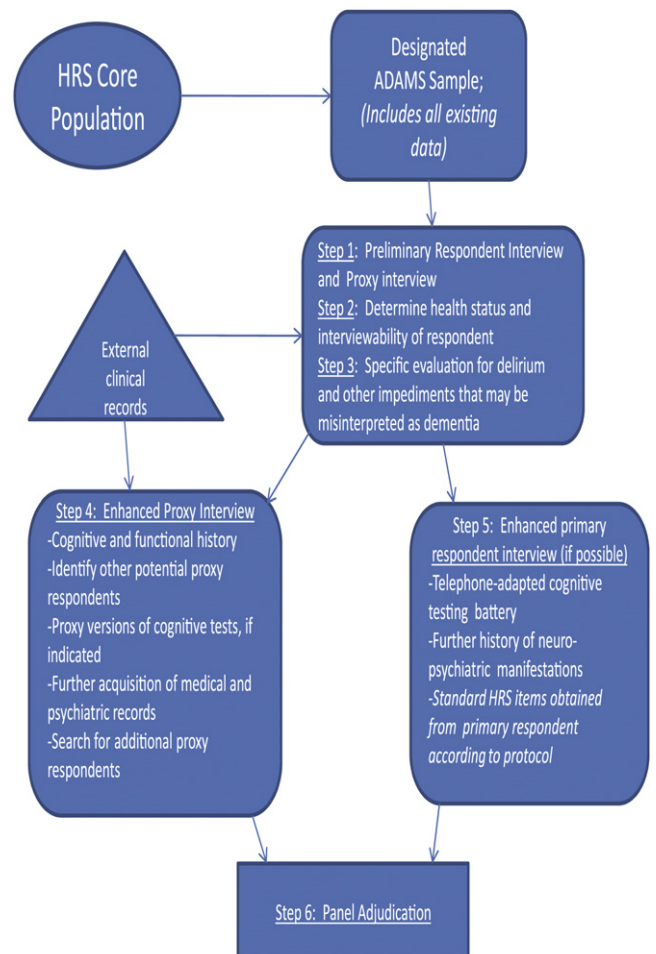


Fig. 2. Subject progress through the alternative protocol for the Aging, Demographics, and Memory Study. For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.



In this section, one such possibility is described—a plan to do more assessment by telephone. The plan, which involves six steps, would require future validation of the proposed methods. The steps are given later in the text and are also depicted in Fig. 2. They attempt to address special issues that begin with the HRS source population. Essentially, designated ADAMS participants would be treated as others in the general HRS protocol. That is, except for possible validation visits, no additional home visits would be conducted.

### 3.1. Step 1

Sample participants would be reached first by telephone. Consents would be obtained from the primary participant and one proxy respondent. A preliminary interview would then be conducted with both the primary and proxy respondents so as to assist in gathering ancillary information and available community medical records. These records might include primary and referral physician records, and insurance claims data, including from Medicare if available.

### 3.2. Step 2

The current physical and cognitive health of the primary respondent would be ascertained using screening items that cover such items as being in bed more than half of each day, use of analgesics and other psychotropic medications, and general adequacy of food intake.

### 3.3. Step 3

Further evaluation of the primary respondent for delirium, mental illness, and other impediments to interviewability would be performed using standardized instruments. Two of the most important conditions that may mimic dementia are depression and delirium. Both conditions can be reliably screened using proxy respondents and established instruments [17,18], particularly emphasizing levels of alertness and consciousness. Determination of, or at least screening for, major concurrent psychological and psychiatric conditions is paramount, because they may be the causes of “pseudo-dementia” as well as a result of the dementing processes [19,20].

If for any reason the primary respondent is deemed unsuitable for cognitive testing or more standardized data collection, the remainder of the information would be collected from the proxy respondent(s), as is done in the full ADAMS protocol. Assembled preliminary information would then be evaluated to decide the remaining steps.

### 3.4. Step 4

If needed, the proxy respondent would be fully interviewed to obtain a clinical and cognitive history of the primary participant, and all additional relevant clinical records would be sought. A proxy cognition evaluation would be conducted if indicated, and possible additional

proxy respondents would be identified for interview if further observations would be informative. Data from previous HRS waves could be particularly important here to provide context to the new data collection. For example, the trajectory of decline in instrumental activities of daily living (IADLs) could help validate new information collected on the emergence of dementia [21].

### 3.5. Step 5

If feasible, the primary respondent would be interviewed, maximizing the content of the full ADAMS protocol that was adapted by telephone. Additionally, if the respondent has a computer available, cognitive testing via the Internet could be explored and could facilitate testing to those that require visualization. Even without visual testing, many cognitive tests have been adapted for telephone use, including assessment of executive function. It may be possible to hone the testing protocol by evaluating various versions or possibly seeking local professional psychological testing services. In addition, a personal history of neurological as well as cognitive symptoms and impairment would be sought, including conditions and exposures relevant to the onset or progression of dementia, such as hypertension, stroke, alcoholism, head trauma, Parkinson's disease, cognitive impairment acquired early in life, and certain hazardous occupational or other environmental exposures.

### 3.6. Step 6

In the final phase of the evaluation, an adjudication panel would review collected information, at first blind to medical records and reports of community diagnoses, and then using all available information, as in the parent ADAMS protocol, yielding the research cognitive diagnosis for analytical purposes.

It is likely that this six-step alternative approach to data collection in ADAMS will yield substantial data collection cost savings for each target respondent studied. Although a formal cost evaluation was not conducted, the savings in travel-related staff and logistical expenses would likely decrease the overall cost by about 20%. It is possible that staff time related to conducting telephone interviews would increase, so the actual cost savings would have to be evaluated through a pilot study.

Despite the efficiencies that are likely to be realized by the aforementioned proposal, there are potential important limitations that should be considered which are as follows: (1) In most instances there will be no direct professional observation of the participant's level of illness or responsiveness to test items or clinical condition. (e.g., there would be no direct observation of participant use of external cues during testing—such as writing down items.); (2) no new biomarker specimen could be directly obtained, although clinical specimens of participants could be obtained by other means; (3) testing may be more difficult than in person if the

respondent has auditory disability or has other difficulty using the telephone; (4) there may be a few desirable cognitive tests that are impractical to administer, such as those requiring direct and visual participation (writing or drawing), and for which computer-aided or proxy-administered testing is not feasible. (How much diagnostic information would be lost must be evaluated in validation studies.); (5) no neurological observation or examination could be performed. This will possibly make diagnosis of some dementia subtypes more difficult.

Clearly, there are some potentially important limitations associated with these proposed survey methods, but how different population prevalence estimates would be vis-à-vis the full ADAMS methodology is unclear, and a well-designed validation and/or pilot study is indicated. A hybrid protocol is also possible, that is when dementia status is uncertain, a home visit with an elaborated protocol could be conducted. Alternatively, in this circumstance, participants could be referred to the nearest appropriate medical center for a fuller evaluation, given that early diagnosis of previously undiagnosed dementia may be of value to the participant and to the family. In the future, new biomarkers related to dementing illnesses may aid in the remote diagnosis of this important, disabling, and fatal disorder. In that regard, the availability in the HRS cohort of genome-wide single nucleotide polymorphism determinations may allow additional diagnostic power.

#### **4. Cost-efficient approach to dementia diagnosis in epidemiological cohort studies**

Longitudinal studies of AD are expensive partly because of the costs associated with clinical classification of AD, dementia, and mild cognitive impairment (MCI). The standard model adopted in most studies is comparable with evaluations performed in tertiary care clinical settings. The core elements of the clinical evaluation—history, neurological examination, and cognitive testing—are supplemented with other procedures including blood work, brain scan, informant interview, and diagnostic case conference. However, the needs of longitudinal research differ from those of clinical practice. In particular, it is imperative to maintain uniformity in procedures across clinicians and time to minimize random variability in clinical classification and to allow for examination of possible change in disease occurrence over time [22]. This is essential to ensure that the relation between an exposure and outcomes, in either observational studies or intervention trials, is not the result of variability in the diagnostic process. Further, procedures should be transparent to facilitate investigation of the effect of specific criteria on diagnostic classification and comparison of findings across studies performed in different cohorts. Clinical decision making systems with these properties have been developed to diagnose psychiatric conditions [23,24]. In the early 1990s at the Rush Alzheimer's Disease Center (Chicago, IL), a clinical classification system was

developed to enhance uniformity, transparency, and efficiency in the diagnosis of AD, dementia, and MCI. Over more than 15 years, this clinical classification system has been implemented in several ongoing longitudinal cohort studies involving thousands of participants and tens of thousands of clinical evaluations.

The standard clinical evaluation used in most longitudinal studies at Rush consists of a structured medical history, neurological examination, and cognitive performance testing. Data are collected on laptop computers with forms programmed in Blaise, a Pascal-based data entry system. The medical history focuses on conditions with the potential to impair cognitive function such as depression, cardiovascular disease, and head injury. A complete neurological examination is administered by specially trained nurses [25]. A battery of cognitive tests is administered in an approximately 45-minute session by a research assistant. The evaluation does not include an informant interview, blood work, or brain scan, procedures often included in dementia evaluations conducted at tertiary care centers, nor are routine case conferences used.

A key issue in clinical classification of dementia by the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association criteria [26] is determining whether functioning in different cognitive domains is impaired. To enhance uniformity in these determinations across clinical decision makers and time, an algorithm was developed for rating impairment in the following five cognitive domains: orientation, attention, memory, language, and visuospatial ability [27,28]. The algorithm was designed to mimic the expert clinical judgments of an experienced neurologist (D.A.B.) and neuropsychologist (R.S.W.). Cutoff scores were selected to identify impairment on 11 widely used cognitive measures at four educational levels (0–7 years, 8–11 years, 12–15 years, 16+ years), and rules were developed for converting the test impairment data to ratings of impairment in the five cognitive domains. The algorithm was then pilot tested against expert judgment. Because the agreement was far from perfect, the cutoff scores and algorithm rules were adjusted, its agreement with clinical judgment was retested, and the process was repeated until there was adequate agreement between the algorithm and clinicians.

After a clinical evaluation is completed, the neuropsychologist reviews the cognitive test data plus information on years of education and sensory or motor impairment and then either agrees or disagrees with the impairment ratings of algorithm for each of the five cognitive domains. In the event of disagreement, the neuropsychologist provides a new rating and the reasons for disagreement (e.g., visually impaired). On completion of this process, the cognitive test results are considered consistent with dementia if a minimum of three of the five domains were impaired (and with AD if one of the domains was memory) and inconsistent if zero or one domain was impaired. If two domains were impaired,

the neuropsychologist rates dementia and AD as present (probable or highly probable) or absent (possible or not present). Subsequently, an experienced physician or nurse clinician reviews all clinical data and briefly examines the participant before determining whether the person has experienced a meaningful decline from a previously higher level of cognitive functioning. This clinical determination is based on several factors, particularly self report of memory impairment, previously shown to be related to AD pathology in the ROS [29], and performance on cognitive testing that is lower than expected for a given educational level, as summarized by the educationally adjusted impairment ratings for cognitive tests and domains. If necessary, the system can be implemented in the absence of the brief clinical evaluation.

A second algorithm classified subjects with respect to dementia (history of cognitive decline, at least two impaired cognitive domains on testing), AD (dementia includes memory impairment), and MCI (no dementia, at least one impaired cognitive domain). The clinician agrees or disagrees and in the latter case provides a new rating and reasons for disagreement.

Although diagnoses in this system are guided by algorithms, clinicians make the final diagnostic decisions. Because the neuropsychologist and physician and/or nurse practitioner make their decisions sequentially, case conferences are not routinely required. Further, because the algorithms are based on expert judgment, disagreement is relatively uncommon and clinical effort focuses more on cases with missing data or other specific problems. As a result, substantially less clinician time is expended per case.

In the remainder of the section, data are presented from the ROS [30] and MAP [31] on the agreement of clinical diagnosis of AD from this system with pathological diagnosis [28]. Additional data are then presented on the relation of genetic and experiential risk factors to AD to illustrate the similarity of findings using this system with studies using other more labor-intensive diagnostic systems.

As reported in more detail elsewhere [28], of the 452 completed autopsies in ROS and MAP, 141 were clinically diagnosed with probable AD proximate to death and 128 (91%) met the National Institute on Aging-Reagan pathological criteria for a high or intermediate likelihood of AD. By way of comparison, data were examined from the Rush Alzheimer's Disease Center's memory clinic, where the standard dementia evaluation includes an informant interview, blood work, brain scan, and case conference. Of 428 completed autopsies from the memory clinic, 306 met clinical criteria for probable AD and 286 (93%) had a high or intermediate likelihood of AD on autopsy.

There were 37 completed autopsies from the two epidemiological studies of individuals with a clinical diagnosis of possible AD (i.e., meet criteria for AD and had another condition contributing to cognitive impairment). On autopsy, 23 (62%) had a high or intermediate likelihood of

AD. Of 54 completed autopsies from the memory clinic of persons diagnosed with possible AD, 48 (89%) meet AD pathological criteria.

Overall, the algorithmic clinical diagnosis of AD showed good agreement with the pathological diagnosis. The level of clinical-pathological agreement in the epidemiological cases was roughly comparable with the level observed in persons from a memory clinic whose AD was diagnosed after a more extensive clinical evaluation.

Inheritance of at least one copy of the apolipoprotein E  $\epsilon$ 4 (ApoE- $\epsilon$ 4) allele has been associated with incidence of AD in studies using a comprehensive dementia evaluation [32,33]. In ROS, the association of  $\epsilon$ 4 with incident MCI [34] and AD [35] was consistent with previous AD research and with its association with accelerated cognitive decline in the same cohort.

Research based on comprehensive dementia evaluations has found that persons with higher levels of depressive symptoms are more likely to develop AD [36,37]. The same relationship has been observed in ROS using the algorithmically guided diagnostic system, further supported by the association of depressive symptoms with cognitive decline [38].

Type 2 diabetes mellitus has been linked to an increased risk of dementia in studies using intensive diagnostic systems [39,40]. In ROS, persons with diabetes were more likely to develop AD than persons without diabetes and they experienced more rapid cognitive decline [41].

The relation of engagement in cognitively stimulating activities has been linked to risk of AD in several studies [42,43]. In ROS [44] and MAP [45], persons with higher levels of engagement in cognitively stimulating activities were less likely to develop incident AD or MCI, and they experienced a slower rate of cognitive decline.

These examples support the validity of the algorithmically guided diagnoses of AD and dementia. That is, the association of risk factors with these diagnoses is consistent with external data using more labor-intensive diagnostic systems and with internal data using cognitive decline as a complementary outcome.

Diagnostic classification is a complex decision-making process. The use of algorithms to guide diagnostic classification has several advantages. The most important one is that structuring the decision-making process reduces random variability, drift, and bias across time and clinicians. Further, an algorithm takes advantage of more information. Without such structure, individual clinicians often use heuristics, thereby making consensus diagnostic conferences essential. Thus, a practical advantage is that the system requires far less clinician time, markedly reducing the direct and opportunity costs of the most expensive data collectors in longitudinal cohort studies of aging and AD. The algorithmically guided classification system presented here has been used in several epidemiological studies of aging and AD. Thus far, research suggests that it has adequate performance properties.

## 5. Assessing the reliability and validity of an algorithm for the diagnosis of dementia and MCI

AD has a prodromal phase that involves progressive impairments in cognitive and functional abilities, from a cognitively normal stage to MCI, and eventually, dementia [46–50]. This has led to proposals for making an accurate diagnosis of AD well before criteria for dementia are fulfilled [47–49]. The diagnosis of AD during the predementia stage, including the amnesic MCI (aMCI) stage [51], has considerable potential for enabling pharmacological and nonpharmacological interventions to be introduced when they are likely to be more effective. However, because of the introduction of new criteria [52] for diagnosing these predementia states, it is becoming evident that there is considerable variability among clinicians and research teams for determining thresholds for making these diagnoses [53–58]. These differences in thresholds are likely to translate into considerable variability in predictive potential for a given diagnosis, such as aMCI.

Traditionally, the diagnosis of normal cognition, MCI, and dementia is based on a combination of a physician's diagnosis, which itself is based on informant's reports of cognitive and functional impairment and a clinical evaluation, reconciled with a NPDx rendered by a psychologist. The typical ConsDx is labor intensive and influenced by the philosophy, personality, discipline, and inherent biases of the individual clinicians involved [59,60]. Methods that may ensure high inter-individual and inter-site reliability in diagnosing predementia states and mild dementia should require fewer subjects with greater power to obtain reliable results for clinical, epidemiological, and especially longitudinal studies. Algorithmic approaches to making consensus diagnoses have been used successfully, showing high concordance rates to physician reviews, even in population-based studies [61].

Duara et al developed a diagnostic algorithm to identify individuals as cognitively normal, with MCI, or with dementia [62]. This algorithmic approach may be expanded to provide reliable diagnoses of cognitive states that precede MCI. The remainder of this section is devoted to a description of the testing to validate the algorithm.

A total of 532 English- and Spanish-speaking, elderly, community-dwelling subjects, aged 52 to 92 years, were recruited by advertisement for a memory-screening study and from a memory disorders clinic. All subjects were assigned a Physician's Cognitive Diagnosis (PhysDx)—NCI, MCI, or dementia—by the examining physician (who was skilled in diagnosing dementia and MCI) on the basis of the entire clinical history of the subject (obtained in English or Spanish), including his/her functional status, Mini-Mental State Examination (MMSE) score [63], and sub-scores. Factors that could influence the physician's impression about the cognitive and functional abilities of the subject, such as educational and cultural background, visual and hearing deficits, language and speech disorders,

general medical, neurological and psychiatric conditions, and the perceived reliability of the informant were taken into consideration.

Each subject in the study was administered a neuropsychological test battery in his/her native language (English or Spanish). To assess memory, the three-trial Fuld Object Memory Evaluation [64] and Delay Recall of the Wechsler Memory Scale-R [65] were used. Tests of nonmemory function included category fluency (language function) [66], letter fluency (executive and language function) [67], Block Design-WAIS-III (visuospatial skills) [68], Trails B (executive function) [69], and Similarities-WAIS-R (executive function) [68]. Neuropsychological classification (NPDx) was achieved using methods developed by Loewenstein et al [70]. The nomenclature used for NPDx was as follows: NCI, nonamnesic MCI (naMCI, single or multi-domain), aMCI (single or multi-domain), and dementia.

The threshold for MCI was a test score of  $\geq 1.5$  standard deviation (SD) below expected normative values, accounting for age, educational level, and language of administration, and was based on a large co-normed normative database used in previous studies [70–72]. The 1.5 SD cut-score is typically the cutoff used for MCI [51]. The threshold for dementia was a score that was  $\geq 2.0$  SD below expected normative values in at least one memory and at least one nonmemory test. This corresponds to a confirmation of a dementia syndrome at or below the fifth percentile specified by National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association criteria [26,51,73].

The National Alzheimer's Coordinating Center/Uniform Data Set (NACC/UDS) D1 diagnosis nomenclature [52] was used for making the final cognitive diagnosis, using, (a) the traditional ConsDx, and (b), the algorithmic diagnosis (AlgDx).

- (a) The ConsDx was derived by discussions between the physician and neuropsychologist in a consensus meeting. These two individuals reviewed the clinical history of the subject, Clinical Dementia Rating (CDR) scores [73,74], and results of neuropsychological evaluation, taking into account any factors that may have influenced the testing.
- (b) The AlgDx was derived by combining the PhyDx with the NPDx, using a computational algorithm that provided the final cognitive diagnosis, as defined by NACC/UDS nomenclature (Table 1). The validity of the AlgDx was assessed by its concordance with the ConsDx, and by its correspondence to two biomarkers closely associated with the presence of AD, namely, medial temporal atrophy (MTA) [75–78] scores from brain magnetic resonance imaging scans, and *APOE*  $\epsilon 4$  genotype [79] (biomarkers of AD were used because AD, alone or in combination with other causes, is by far the most common cause of progressive aMCI or dementia [80]).



Table 1  
Algorithm for combining the physician's diagnosis with the neuropsychological diagnosis to derive the algorithmic diagnosis

Physician's diagnosis	Neuropsychological diagnosis			
	Normal	Nonamnestic MCI	Amnestic MCI	Dementia
Normal	Normal	Normal	Normal	MCI
MCI	Normal	MCI	MCI	MCI
Dementia	MCI	MCI	Dementia	Dementia

MCI, mild cognitive impairment.

The inter-rater reliability of PhysDx and NPDx was assessed for two separate physician/neuropsychologist teams, who independently assessed the same 30 subjects (10 with NCI, 10 with MCI, and 10 with dementia). The inter-rater reliability for PhysDx, measured by Cohen's weighted kappa [81], was 0.69 (standard error [SE] = 0.11) (agreement was 70% for NCI, 70% for MCI, and 80% for dementia) and for the NPDx was 0.88 (SE = 0.07) (agreement was 89% for NCI, 91% for MCI, and 88% for dementia). Finally, the inter-rater reliability for ConsDx was 0.78 (SE = 0.07) (agreement was 90% for NCI, 70% for MCI, and 80% for dementia).

The concordance of the AlgDx to the same ConsDx categories ranged from 85% to 92%. AlgDx and ConsDx were the same for 88.2% of NCI, 85.1% of aMCI, and 90.9% of dementia cases. Cohen's weighted kappa for agreement was 0.84 (SE = 0.02), which was a high concordance between the two approaches.

Most of this subject sample had MTA data (427 cases) and ApoE-ε4 genotyping (314 subjects). Using the AlgDx classification, post hoc Sidak tests of means indicated that subjects diagnosed with aMCI had higher mean MTA scores (0.944; SD, 0.76) in comparison with NCI subjects (0.632; SD, 0.69), and dementia subjects had higher MTA scores (1.78; SD, 1.10) in comparison with subjects in the other two diagnostic groups ( $F(3,423) = 41.02$ ;  $P < .001$ ). With regard to ε4 frequencies, there were statistically significant differences between groups ( $\chi^2 [df = 3] = 31.33$ ;  $P < .001$ ). Subjects diagnosed with dementia had the highest ε4 frequency (34.9%), followed by aMCI (28.3%) and NCI (12.9%) subjects.

The AlgDX was developed with the goal of having a unified method of incorporating elements from the medical and neuropsychological examinations, with clear decision-making rules used across varying diagnostic teams. The data suggest that the AlgDx provides a simple, reliable, and valid alternative to the classical ConsDx of cognitive impairment. As such, AlgDx may have particular utility for longitudinal, multi-site clinical trials, and population-based studies of MCI and dementia. It is apparent that increased reliability, brevity, convenience, and equivalent validity of the AlgDx makes it an appropriate and a potentially cost-effective approach for diagnosing MCI and other predementia states, as well as early dementia.

The AlgDx of normal cognition, MCI, and dementia is a valid alternative that reduces time, effort, and biases asso-

ciated with the ConsDx. Given the inherent reliability of a fixed algorithm, its user friendly nature and coupled with demonstrated efficiency and avoidance of individual bias, the application of the AlgDx in clinical and epidemiological research is worthy of further study.

## 6. A Web-based approach to diagnostic consensus: Experience of the Monongahela-Youghiogheny Healthy Aging Team Project

For research diagnosis in conditions such as AD, in which there is no single definitive diagnostic test, many clinical research centers rely on a process of data review, adjudication, and consensus by a multi-disciplinary panel of expert clinicians [82–84]. The panel meets in real time to review detailed information on various aspects of the clinical and laboratory assessment of a given patient, discuss the findings, and, render a consensus diagnosis using standardized criteria. This process allows data of each study participant to be individually considered in detail, bringing to bear a wealth of collective clinical expertise and judgment. However, it involves the cost of the time spent by experts in the meeting, the inefficiency of scheduling meetings at a time and location that all experts can attend, and the near impossibility of including experts at different sites.

Research diagnosis poses different challenges in population studies. Participants are often interviewed and examined in their homes or other locations remote from the academic center; assessments are often conducted by raters who are well-trained research personnel but not expert clinicians. When these assessments follow highly standardized protocols, the resulting data can be reviewed by experts who have not personally examined the participants [85,86]. Nevertheless, for multiple experts to review the same data and come to consensus, it is still the norm for a live consensus conference or a teleconference to be implemented, with the same constraints of cost and scheduling.

In this section, an initiative to establish a process of Web-based diagnostic consensus within a population study is reported. The objective was to involve clinical experts in reviewing and rating standardized assessment data in a manner that would eliminate scheduling constraints, minimize cost once the infrastructure was established, and also yield data for analysis beyond the diagnosis itself.

### 6.1. Study site, sample, assessment

The Monongahela-Youghiogheny Healthy Aging Team (MYHAT) Project is a population-based study of the epidemiology of MCI, conducted in a group of small-town communities near Pittsburgh, in southwestern Pennsylvania. All study procedures including the Web-based process are approved annually by the University of Pittsburgh Institutional Review Board. Sampling, recruitment, and assessment of the study cohort have been described previously [87,88].

6.1.1. Selection of variables for online consensus Web site

The assessment protocol was reviewed to select variables that were judged to contribute to the desired ratings and diagnostic impressions. These data points were categorized in the following four groups: (1) demographic and background characteristics: age, gender, education, primary occupation, reading level and estimated intelligence quotient, hearing, and vision; (2) variables relevant to the CDR [73]: subjective complaints, ADLs, instrumental activities of daily living, depressive symptoms, medication management, social engagement, and judgment; (3) variables relevant to cognitive classification: participant's scores on all tests in the neuropsychological battery; (4) variables relevant to etiological diagnosis: health history, medications, physical and neurological examination, and neuroimaging reports if any.

6.1.2. Structure and sequence of Web pages

Designated raters log into a secure Web site and select, from a drop-down menu, one subject at a time for rating. The subject is identified by an identification (ID) number (Study ID) and the annual cycle during which the posted data were collected. Clicking on the Study ID leads the rater

to a series of pages containing data from that participant. A schematic diagram of the sequence is shown in Fig. 3.

The first page viewed displays variables from the first two categories mentioned earlier. The MMSE [63] total score is visible on this page; clicking on the legend "MMSE" opens a table where item-by-item MMSE scores are provided. The same option is available for ADL and other impairment scales. At the bottom of this page, a link is provided to "make your Clinical Dementia Rating." Clicking on this link opens a page where a link is provided to the Washington University CDR scoring algorithm [89] to calculate the CDR summary score on the basis of individual "box scores." Below this is a menu to choose a selection for a required CDR rating of 0, 0.5, 1, 2, or 3; an optional free text field for comments, if any; and a menu to choose a required certainty rating ranging from 1 (not at all certain) to 5 (absolutely certain). Clicking on an icon to save these ratings leads the rater to the next page.

The following page is devoted to cognitive classification, containing the neuropsychological information. Note that the rater has already completed the CDR, based on everyday functioning, before viewing the neuropsychological data. On

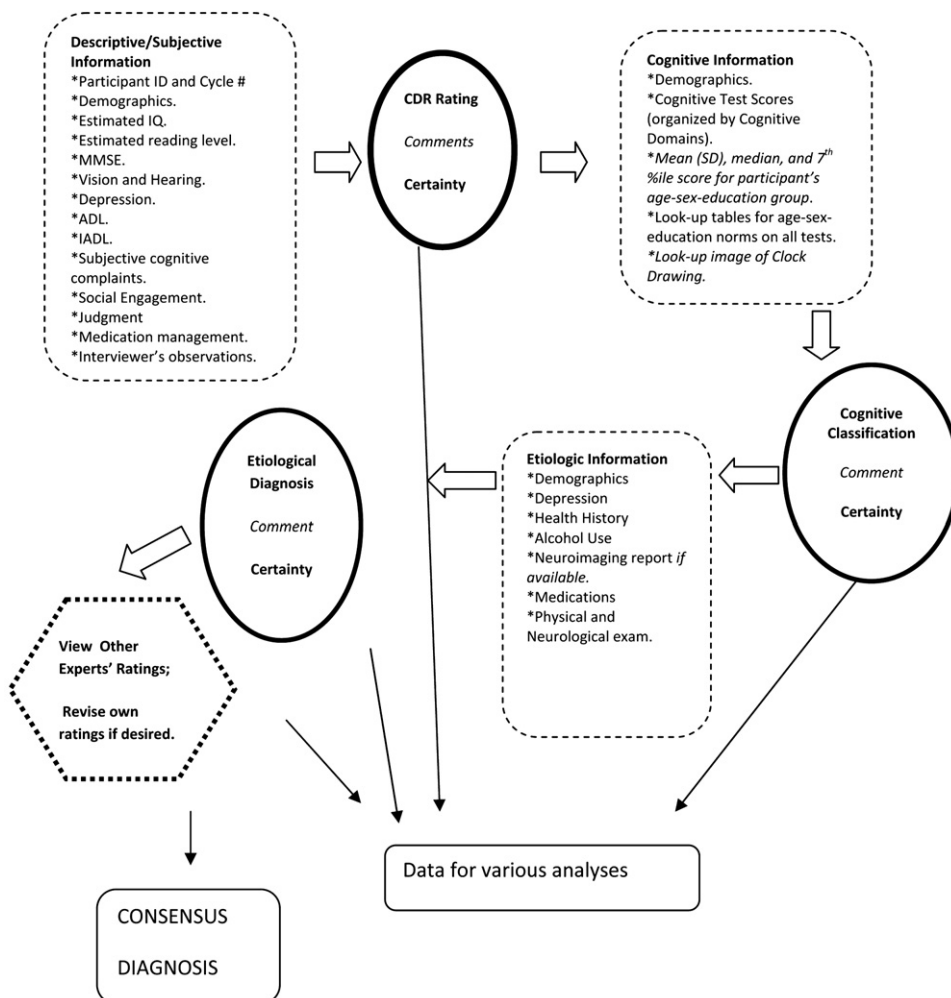


Fig. 3. Schematic diagram of online diagnostic consensus process used in the Monongahela-Youghiogheny Healthy Aging Team Project.

the cognitive page, tests are categorized by the corresponding principal cognitive domain, i.e. attention/processing speed, executive function, memory, language, and visuospatial function. For each test, the table provides the participant's score alongside the mean, standard deviation, and seventh percentile score (equivalent to 1.5 SD below the mean) for the participant's age-gender-education group. If desired, the rater can click on the name of the test to view a table showing MYHAT cohort norms on that test. Clicking on "Clock Drawing" displays a scanned image of that participant's clock drawing. At the bottom of this page is a link to click to "Make your Cognitive Classification." The choices provided include normal, focal aMCI, multi-domain aMCI, focal naMCI, multi-domain naMCI, and moderate to severe cognitive impairment (aka dementia). Again, the rater selects the cognitive classification and certainty level; comments are optional.

The final page is devoted to etiologically relevant information and includes the fourth category of variables listed earlier; by this time the rater has already viewed the CDR and cognitive data and is able to take all information into account when rendering the etiological diagnosis. The rater clicks on "Make your Etiological Diagnosis." Again, the etiological choice and the certainty ratings are required; additional comments, including etiological options not listed in the menu, are optional (Fig. 4).

When the rater has completed all three ratings, clicking on a link to return to the home page saves all these ratings to the database. An asterisk now appears next to the Study ID, indicating that this rater has already dated that case. Now, clicking on the ID allows the rater to view ratings that other raters

have completed on the same case, and to revise his ratings if so desired. The original ratings are also saved.

### 6.1.3. Web development details

The Web site was developed using Microsoft SharePoint with Web parts developed in Visual Studio.Net. (Microsoft, Inc., Seattle, WA, 2008). It is estimated that building the Web site required 320 hours (8 weeks) of programmer time, and, subsequently, 1 hour per week of systems administrator time for server upkeep, security, and Web site modifications. The Web site requires 128-bit secure socket layer encryption as well as full authentication using domain accounts and passwords. Raters are identified automatically within the database and all ratings, including any changes in ratings, confidence or comments are logged. Research subjects are only identified by Study ID. No caching or cookies are stored on the rater's workstation.

## 6.2. Conclusions

The familiar, widely established, live expert consensus process is not usually well-described or highly standardized, with rare published exceptions [85,86]. Previous studies have reported inter-rater agreement on diagnoses by individual clinicians [84] and stability or validity of diagnosis after examination of neuroimaging and neuropathological data [85,86]. Few, if any, reports describe the consensus diagnosis process itself or the relative contribution of its various components. The authors are aware of two previous efforts at Web-based diagnosis. One was designed to examine inter-rater agreement in pathology diagnosis, and was accomplished by adding annotations to an existing

Home Up to MYHAT

Etiological Diagnosis

**First Time Rating Instructions:** Please select an Etiological Diagnosis, enter any comments that you may have in the text area provided, and select an Etiological Confidence Rating. Click on the designated link at the bottom of this page to save your rating.

**Instructions for Cases Already Scored:** If the case you are viewing has already been scored, you may update or correct your rating at this time. If you choose not to update or correct your rating, you must re-enter your last rating shown in the blue bar to proceed to the next step. Also, please note that this consensus system only updates every 3 minutes. So, if you want to make a change to your rating, please be sure to allow for this delay before attempting to change your rating. At this point, you will be returned to the home page of the web site.

**Your saved Etiological rating is Mixed Degenerative + Vascular with confidence level of 3: Reasonably certain**

Reviewer \* gangulm  
 Research ID \* 100001  
 Cycle \* 1  
 Etiological Diagnosis \* Mixed Degenerative + Vascular  
 Etiological Comments Gradual progressive decline, multi-domain amnesic impairment, history of stroke  
 Etiological Confidence \* 3: Reasonably certain

\* indicates a required field

\*Please note that ratings will only be saved if there is BOTH a score and a confidence rating. Comments are optional, but appreciated.

[Click Here to Save Your Rating](#)

Internet 100%

Fig. 4. Screenshot of Web page for etiological diagnosis. The Monongahela-Youghio gheny Healthy Aging Team Project. For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.

Web site where the pathology images were stored [90]. Another described a clinical Web environment for the diagnosis of Alzheimer's and other dementias [91].

As this section is a description of a Web-based process, there are no empirical results to report. Raters include neurologists, psychiatrists, neuropsychologists, and geriatricians. They have uniformly described the process as interesting and user-friendly, and, after an initial learning curve, reported that cases take an average of 10 minutes to complete. Because raters can log into the secure Web site from any location at their convenience, scheduling constraints are eliminated and time is efficiently used. Apart from identifying the diagnoses on which all raters agree, and to select the modal rating or diagnosis where there is less than perfect agreement, data can be used in multiple ways. For example, it is possible to calculate inter-rater agreement across raters and within/across specialties, and to identify rating/diagnostic categories where there is greater and lesser agreement. Researchers can also examine change in etiological diagnosis after additional (e.g., brain magnetic resonance imaging) data are presented. They can attempt to develop diagnostic algorithms on the basis of ratings provided by experts. By empirically identifying the components that predict these ratings, they can even attempt to "deconstruct" to some extent the process of expert clinical judgment.

The Web-based process described here combines the advantages of assessment by nonexperts, judgment and diagnosis by experts, and the convenience of online rating. It offers some alternatives which have potential advantages of standardization, empirical analysis, and efficiency in the use of expert clinician's time, which likely enhances cost-effectiveness.

## 7. Discussion

The public health burden of AD is well documented for the most advanced stages of disease and the importance of identifying cases is critical to establish both prevalence and incidence of disease and to one day conducting clinical trials in disease prevention. In the various sections of this article, different methods of case ascertainment are highlighted through examples of several studies. In each study, the focus is on aspects of the diagnostic process in which cost savings might be realized by means of some methodological modification. However, before savings can be realized there must be a transparent model of cost which acknowledges the resources required for each element of ascertainment. These elements include among others, sample identification and retention, proxy input, clinical, demographic and social data collection, cognitive assessment, functional assessment, and the cost of diagnostic expertise. As we move to earlier stages of detection there may even be an expectation of specific laboratory data for detection which will also incur an as yet unknown cost.

One of the challenges of ascertainment of dementia, acknowledged by many, is that the disease of interest impedes the ability to participate in study activities. The ability to complete assessments and to assess one's own performance is impaired. This leads to the need to use proxies to remove the "self-as-observer" bias for measuring current level of function and to provide comparison to historical performance. Proxies, often family or friends, are a major focus for assessment in surveys such as HRS, but they can have a varying ability to make historical comparisons. This ability to describe a change from previous function can mitigate the lack of previous assessments, which reflects cost saving, but the frequency of contact and the intimacy of knowledge will affect the ability to contribute information of value. The identification of a second individual as well as an evaluation of the quality of information he or she can provide must be factored into the cost of ascertainment. The ADAMS cohort, a resource-intensive sub-study of HRS, demonstrates the under-detection of survey methods, but provides an opportunity to select and validate survey items that might provide greater specificity. A proposal to reduce the cost of the ADAMS evaluation focuses on the use of telephone assessment rather than in-person evaluations. This multi-step process, proposed by Wallace, Langa, and Plassman (section 3), is as yet unvalidated; it begins with intensive clinical data collection and includes screening for depression and delirium. The efficiency of this approach includes selective use of informants in the early steps, telephone-based assessment of cognition, and intensive medical record review. These distance-based assessments limit the ability to collect biological samples but cost reduction is realized through minimizing staff travel, though the assessment time and effort for record review must be considered. Of note, Wilson and Bennett (section 4) describe longitudinal assessment, with algorithmic diagnosis, that does not depend on an informant. Such an approach may be most efficient when the population has high retention and follow-up, as in ROS or MAP.

Several studies focus on detailed data collected by non-professionals, but reviewed by professionals to achieve the diagnosis (sections 4–6). This requires one or more professionals to give input to some or all aspects of the clinical picture but saves the cost of professional assessment of individual cases. The consensus diagnosis typically describes an interdisciplinary team of experts conducting ascertainment by agreement. A feature of the consensus diagnosis is that judgment is used to weigh variables that may have a nonspecific contribution or be part of a complex interaction. For example, a consensus diagnosis may weigh poor cognitive performance in the presence of other medical, physical, or social factors. Technologies are available to permit this activity to be performed without the cost or burden of joint meetings, but through review of iterative opinions as described in the MYHAT Project (section 6). Web-based approaches to data review and decision collection provides a practical approach to examining each step of the diagnostic process. Further cost savings may be achieved by the use of



an algorithm that weighs the input from different assessment domains. Weightings would ideally be built on the actual experience of clinical expertise. For example, examination of how specific medical, physical, or social factors were used to arrive at a consensus diagnosis can inform a diagnostic algorithm. The inclusion of demographic and clinical data can provide additional refinements, and laboratory results may suggest etiology and levels of certainty. ROS and MAP include longitudinal follow-up with autopsy confirmed diagnosis which provides unique validation of the diagnostic algorithm (section 4). Transparency requires acknowledging the initial cost of clinical expertise and neuropathological validation in formulating the algorithm with savings achieved by eliminating clinical expertise for case review.

The work of Duara and Loewenstein (section 5) and Ganguli (section 6) describe categorization within mild forms of impairment, describing amnesic and nonamnesic forms of MCI, conditions defined primarily by normative data. These categories illustrate the ability to detect subtle deficits but will remain in diagnostic infancy until sufficient longitudinal data to understand the predictive value and the full effect of such entities within the healthcare system are available. Nevertheless, the systematic collection of extensive data can provide a diagnostic algorithm for the most important conditions, and may identify the minimum dataset required to make the diagnosis. The simultaneous collection of simple survey tools within these elaborate assessments may provide validation for methods to be used in larger populations.

Although diagnosis by consensus or algorithm may save the cost of the professional assessment of each individual, the cost of collecting assessments remains high. Cognitive testing requires training and takes time and resources. Functional assessment and other clinical and demographic data are rarely available in a systematic way. Of note, neither administrative datasets nor general medical records are likely to contain information about subtle cognitive deficits or mild functional impairment (unless cognitive impairment is a presenting complaint). Thus, specific assessment above and beyond medical record is required for ascertainment of cognitive loss or dementia. The work of Wilson and Bennett (section 4), Duara and Loewenstein (section 5), and Ganguli (section 6) describe collection of extensive assessments with highly trained staff. Maintaining a staff that can provide reliable standardized assessment is a very large part of the ascertainment expense. Weir (section 2) and Wallace, Langa, and Plassman (section 3) describe ways to reduce this expense, including the use of survey methods and telephone-based assessments. The sensitivity of these approaches for ascertainment of dementia is well established, although not without limitations. However, the effectiveness in detecting MCI is not well established. Among the challenges of these methods is the capture of information from an aging population, often with sensory deficit. About 40% of people aged >65 years suffer hearing loss, impeding telephone interactions; visual impairment can interfere with completion

of written materials; and general caution about interacting with unknown surveyors can reduce the possibility of participation among the elderly. Estimates of participation of targeted populations can be as low as 50% to 65% and although increased sample size can address this for establishing prevalence, true incidence and randomized trials are particularly challenged. Thus, the cost savings of these approaches needs to be weighed against any need to increase sample size.

True incidence requires longitudinal surveillance and re-assessment over time. Even among those with measurable cognitive impairment, incidence is as low 5% to 15% annually. Prevention studies of nonimpaired individuals required a doubling of observation time to achieve the needed conversion rates. The cost of maintaining the cohort can be considerable, particularly when it may be a transient population. Although impairment may reduce mobility, relocation for proximity to family or for higher level care is likely. Of note, little has been described of the cost of recruitment or retention. The resources and cost of obtaining consent, and of re-evaluation to ensure continued capacity to consent, in a population at risk for cognitive change are also unrecognized. Yet, these are necessary costs when methodologies include performance-based evaluation, acquiring medical records, and using a proxy informant. Many cost-effective approaches to ascertainment focus on reducing manpower for data collection. These include removing the expert clinician from the assessment and using technologies to collect data, all of which reduce the human contact between participant and researcher. Although these may be cost saving and even reduce the burden on the participant they may also inadvertently reduce the visibility of the project and undermine the importance of the effort contributed by participants. Of particular importance is building the partnership with participants to create shared commitment for the outcomes of these studies. When the motivation for follow-up is shared by the participant and project staff, the effort to track and follow cases may be reduced and retention may be higher.

Great progress has been made in the willingness and ability to detect cognitive impairment and dementia in aging populations. Factors such as the aging of the baby boomer generation, proposed increases in retirement age, and improved longevity highlight the healthcare imperative to understand and address these diseases. Our growing experience with evaluation of cognition and function in diverse and real world populations is leading to efficient methods for characterizing cognitive impairment and diagnosing dementia. Efficient and effective measures of performance have been developed and are widely used. Moving forward, a transparent and informed approach is needed to evaluate both the savings and potential unintended costs of ascertainment methods. Efforts to maximize these efficiencies can reduce the cost of diagnosis but must be balanced against the cost of underestimates of disease. These methods will be critical to conduct disease prevention trials and in fact have already proven effective in such studies. Although the most novel approaches to prevention postulate

a presymptomatic stage of disease which would be defined by a biomarker it is not clear what economy this will provide. Thus far, no biomarker predicts progression or incident dementia better than memory impairment in the otherwise asymptomatic individual. Although the development of precise diagnostic laboratory tests may reduce uncertainty of diagnosis, it is unclear that it will reduce the cost of ascertainment.

## Acknowledgments

This combined effort was supported by National Institute on Aging grants U01AG009740, P30AG010161, R01AG-015819, R01AG017917, P50AG025711, R01AG030561, R01AG023651, K24AG022035, P50AG005138. This work was also supported by the State of Florida Department of Elder Affairs, Tallahassee, Florida; the Johnnie Byrd, Sr., Alzheimer's Center and Research Institute and the School of Medicine, University of South Florida, Tampa, Florida.

Drs. Wallace, Langa, and Plassman thank Michelle Birt Leeds for her excellent assistance in assembling and editing parts of this manuscript. Dr. Ganguli thanks Lynda Rose, Cathy Kanczes, Jack Doman, Joni Vander Bilt, and Kathryn McMichael at the University of Pittsburgh.

The authors have no conflicts to disclose. The sponsors had neither a role in the analysis or interpretation of these data, nor in the content of the article. Appropriate approval procedures were used concerning human subjects.

## References

- [1] Brookmeyer R, Gray S, Kawas C. Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *Am J Public Health* 1998;88:1337–42.
- [2] Sloane PD, Zimmerman S, Suchindran C, Reed P, Wang L, Boustani M, Sudha S. The public health impact of Alzheimer's disease, 2000–2050: potential implication of treatment advances. *Ann Rev Public Health* 2002;23:213–31.
- [3] Evans DA, Grodstein F, Loewenstein D, Kaye J, Weintraub S. Reducing case ascertainment costs in U.S. population studies of Alzheimer's disease, dementia, and cognitive impairment—Part 2. *Alzheimer Dement* 2011;7:110–23.
- [4] Ofstedal MB, Fisher GG, Herzog AR. Documentation of cognitive functioning measures in the Health and Retirement Study. HRS Documentation Report DR-006 (March 2005). Available at: <http://hrsonline.isr.umich.edu/sitedocs/userg/dr-006.pdf>.
- [5] 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;19:1015–22.
- [6] Langa KM, Chernew ME, Kabeto MU, Herzog AR, Ofstedal MB, Willis RJ, et al. National estimates of the quantity and cost of informal caregiving for the elderly with dementia. *J Gen Intern Med* 2001;16:770–8.
- [7] Langa KM, Larson EB, Wallace RB, Fendrick AM, Foster NL, Kabeto MU, et al. Out-of-pocket health care expenditures among older Americans with dementia. *Alzheimer Dis Assoc Disord* 2004;18:90–8.
- [8] Langa KM, Plassman BL, Wallace RB, Herzog AR, Heeringa SG, Ofstedal MB, et al. The Aging, Demographics, and Memory Study: study design and methods. *Neuroepidemiology* 2005;25:181–91.
- [9] Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, et al. Prevalence of cognitive impairment without dementia in the United States. *Ann Intern Med* 2008;148:427–34.
- [10] Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, et al. Prevalence of dementia in the United States: the Aging, Demographics, and Memory Study. *Neuroepidemiology* 2007;29:125–32.
- [11] Taylor DH, Østbye T, Langa KM, Weir D, Plassman BL. The accuracy of Medicare claims as an epidemiological tool: the case of dementia revisited. *J Alzheimers Dis* 2009;17:807–15.
- [12] Juster F, Suzman R. An overview of the Health and Retirement Study. *J Hum Resour* 1995;30:S7–56.
- [13] Gure T, Kabeto M, Plassman B, Piette J, Langa K. Differences in functional impairment across subtypes of dementia. *J Gerontol A Biol Sci Med Sci* 2010;65:434–41.
- [14] Okura T, Plassman B, Steffens D, Llewellyn D, Potter G, Langa K. Prevalence of neuropsychiatric symptoms and their association with functional limitations in older adults in the United States: the Aging, Demographics, and Memory Study. *J Am Geriatr Soc* 2010;58:330–7.
- [15] Rogers MA, Plassman BL, Kabeto M, Fisher GG, McArdle JJ, Llewellyn DJ, Potter GG, Langa KM. Parental education and late-life dementia in the United States. *J Geriatr Psychiatry Neurol* 2009;22:71–80.
- [16] Mehta K, Stewart A, Langa K, Yaffe K, Moody-Ayers S, Williams B, et al. “Below average” self-assessed school performance and Alzheimer's disease in the Aging, Demographics, and Memory Study. *Alzheimers Dement* 2009;5:380–7.
- [17] Schuurmans M, Deschamps P, Markham S, Shortridge-Baggett L, Duursma S. The measurement of delirium: review of scales. *Res Theory Nurs Pract* 2003;17:207–24.
- [18] Chen C, Liu C, Liang H. Comparison of patient and caregiver assessments of depressive symptoms in elderly patients with depression. *Psychiatry Res* 2009;166:69–75.
- [19] Steffens DC, Fisher GG, Langa KM, Potter GG, Plassman BL. Prevalence of depression among older Americans: the Aging, Demographics and Memory Study. *Int Psychogeriatr* 2009;21:879–88.
- [20] Panza F, Frisardi V, Capurso C, D'Introno A, Colacicco A, Imbimbo B, et al. Late-life depression, mild cognitive impairment, and dementia: possible continuum? *Am J Geriatr Psychiatry* 2010;18:98–116.
- [21] Barberger-Gateau P, Fabrigoule C, Helmer C, Rouch I, Dartigues J. Functional impairment in instrumental activities of daily living: an early clinical sign of dementia? *J Am Geriatr Soc* 1999;47:456–62.
- [22] Hebert LE, Bienias JL, Aggarwal NT, Wilson RS, Bennett DA, Shah RC, et al. Change in risk of Alzheimer's disease over time. *Neurology* 2010;75:786–91.
- [23] Wing JK. The use of the Present State Examination in general population surveys. *Acta Psychiatr Scand* 1980;62:230–40.
- [24] Copeland JR, Dewey ME, Griffiths-Jones HM. A computerized psychiatric diagnostic system and case nomenclature for elderly subjects: GMS and AGE-CAT. *Psychol Med* 1986;16:89–99.
- [25] Bennett DA, Shannon KM, Beckett LA, Goetz CG, Wilson RS. Metric properties of nurses' ratings of Parkinsonian signs with a modified Unified Parkinson's Disease Rating Scale. *Neurology* 1997;49:1580–7.
- [26] McKhann G, Drachman D, Folstein M, Katzman R, Price D, 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;34:939–44.
- [27] Bennett DA, Wilson RS, Schneider JA, Evans DA, Beckett LA, Aggarwal NT, et al. Natural history of mild cognitive impairment in older persons. *Neurology* 2002;59:198–205.
- [28] Bennett DA, Schneider JA, Aggarwal NT, Arvanitakis Z, Shah RC, Kelly JF, et al. Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology* 2006;27:169–76.

- [29] Barnes LL, Schneider JA, Boyle PA, Bienias JL, Bennett DA. Memory complaints are related to Alzheimer disease pathology in older persons. *Neurology* 2006;67:1581–5.
- [30] Wilson RS, Bienias JL, Evans DA, Bennett DA. Religious Orders Study: overview and change in cognitive and motor speed. *Aging Neuropsychol Cogn* 2004;11:280–303.
- [31] Bennett DA, Schneider JA, Buchman AS, Mendes de Leon CF, Bienias JL, Wilson RS. The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. *Neuroepidemiology* 2005;25:163–75.
- [32] Tang MX, Stern Y, Marder J, Bell K, Gurland B, Lantigua R, et al. The APOE  $\epsilon 4$  allele and the risk of Alzheimer disease among African Americans, Whites, and Hispanics. *JAMA* 1998;279:751–5.
- [33] Khachaturian AS, Corcoran CD, Mayer LS, Zandi PP, Breitner JC, Cache County Investigators. Apolipoprotein E  $\epsilon 4$  count affects age at onset of Alzheimer disease, but not lifetime susceptibility: the Cache County Study. *Arch Gen Psychiatry* 2004;61:518–24.
- [34] Boyle PA, Buchman AS, Wilson RS, Kelly JF, Bennett DA. The ApoE  $\epsilon 4$  allele is associated with incident mild cognitive impairment among community-dwelling older persons. *Neuroepidemiology* 2009;34:43–9.
- [35] Wilson RS, Schneider JA, Barnes LL, Beckett LA, Aggarwal NT, Cochran EJ, et al. The apolipoprotein E  $\epsilon 4$  allele and decline in different cognitive systems during a 6-year period. *Arch Neurol* 2002;59:1154–60.
- [36] Berger AK, Fratiglioni L, Forsell Y, Winblad B, Backman L. The occurrence of depressive symptoms in the preclinical phase of AD: a population-based study. *Neurology* 1999;53:1998–2002.
- [37] Devanand DP, Sano M, Tang MX, Taylor S, Gurland BJ, Wilder D, et al. Depressed mood and the incidence of Alzheimer's disease in the elderly living in the community. *Arch Gen Psychiatry* 1996;53:175–82.
- [38] Wilson RS, Barnes LL, Mendes de Leon CF, Aggarwal NT, Schneider JS, Bach J, et al. Depressive symptoms, cognitive decline, and risk of AD in older persons. *Neurology* 2002;59:364–70.
- [39] Luchsinger JA, Tang MX, Stern Y, Shea S, Mayeux R. Diabetes mellitus and risk of Alzheimer's disease and dementia with stroke in a multiethnic cohort. *Am J Epidemiol* 2001;154:635–41.
- [40] Peila R, Rodriguez BL, Launer LJ. Honolulu-Asia Aging Study. Type 2 diabetes, APOE gene, and the risk for dementia and related pathologies: the Honolulu-Asia Aging Study. *Diabetes* 2002;51:1256–62.
- [41] Arvanitakis Z, Wilson RS, Bienias JL, Evans DA, Bennett DA. Diabetes mellitus and risk of Alzheimer disease and decline in cognitive function. *Arch Neurol* 2004;61:661–6.
- [42] Scarmeas N, Levy G, Tang MX, Manly J, Stern Y. Influence of leisure activity on the incidence of Alzheimer's disease. *Neurology* 2001;57:2236–42.
- [43] Verghese J, Lipton RB, Katz MJ, Hall CB, Derby CA, Kuslansky G, et al. Leisure activities and the risk of dementia in the elderly. *N Engl J Med* 2003;348:2508–16.
- [44] Wilson RS, Mendes de Leon CF, Barnes LL, et al. Participation in cognitively stimulating activities and risk of incident Alzheimer's disease. *JAMA* 2002;287:742–8.
- [45] Wilson RS, Scherr PA, Schneider JA, Tang Y, Bennett DA. Relation of cognitive activity to risk of developing Alzheimer disease. *Neurology* 2007;69:1911–20.
- [46] Bowen J, Teri L, Kukull W, McCormick W, McCurry SM, Larson EB. Progression to dementia in patients with isolated memory loss. *Lancet* 1997;349:763–5.
- [47] Grober E, Hall CB, Lipton RB, Zonderman AB, Resnick SM, Kawas C. Memory impairment on free and cued selective reminding predicts dementia. *Neurology* 2000;54:827–32.
- [48] Howieson DB, Dame A, Camicioli R, Sexton G, Payami H, Kaye JA. Cognitive markers preceding Alzheimer's dementia in the healthy oldest old. *J Am Geriatr Soc* 1997;45:584–9.
- [49] Fabrigoule C, Rouch I, Taberly A, Letenneur L, Commenges D, Mazaux JM, et al. Cognitive process in preclinical phase of dementia. *Brain* 1998;121:135–41.
- [50] Tabert MH, Albert SM, Borukhova-Milov L, Camacho Y, Pelton G, Liu X, et al. Functional deficits in patients with mild cognitive impairment: prediction of Alzheimer's disease. *Neurology* 2002;58:758–64.
- [51] Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med* 2004;256:183–94.
- [52] Beekly DL, Ramos EM, Lee WW, Deitrich WD, Jacka ME, Wu J, et al. The National Alzheimer's Coordinating Center (NACC) database: the Uniform Data Set. *Alzheimer Dis Assoc Disord* 2007;21:249–58.
- [53] Graham JE, Rockwood K, Beattie BL, McDowell I, Eastwood R, Gauthier S. Standardization of the diagnosis of dementia in the Canadian Study of Health and Aging. *Neuroepidemiology* 1996;15:246–56.
- [54] Luis CA, Barker WW, Loewenstein DA, Crum TA, Rogava E, Kawarai T, St George-Hyslop P, Duara R. Conversion to dementia among two groups with cognitive impairment. A preliminary report. *Dement Geriatr Cogn Disord* 2004;18:307–13.
- [55] Bennett DA, Schneider JA, Bienias JL, Evans DA, Wilson RS. Mild cognitive impairment is related to Alzheimer disease pathology and cerebral infarctions. *Neurology* 2005;64:834–41.
- [56] Fillenbaum GG, Peterson B, Morris JC. Estimating the validity of the clinical Dementia Rating Scale: the CERAD experience. Consortium to Establish a Registry for Alzheimer's Disease. *Aging (Milano)* 1996;8:379–85.
- [57] Rockwood K, Strang D, MacKnight C, Downer R, Morris JC. Interrater reliability of the clinical dementia rating in a multicenter trial. *J Am Geriatr Soc* 2000;48:558–9.
- [58] Schafer KA, Tractenberg RE, Sano M, Mackell JA, Thomas RG, Gamst A, et al. Reliability of monitoring the clinical dementia rating in multicenter clinical trials. *Alzheimer Dis Assoc Disord* 2004;18:219–22.
- [59] Hogervorst E, Bandelow S, Combrinck M, Irani SR, Smith AD. The validity and reliability of 6 sets of clinical criteria to classify Alzheimer's disease and vascular dementia in cases confirmed post-mortem: added value of a decision tree approach. *Dement Geriatr Cogn Disord* 2003;16:170–80.
- [60] Jobst KA, Barnetson LP, Shepstone BJ. Accurate prediction of histologically confirmed Alzheimer's disease and the differential diagnosis of dementia: the use of NINCDS-ADRDA and DSM-III-R criteria, SPECT, X-ray CT, and APO E4 medial temporal lobe dementias. The Oxford Project to Investigate Memory and Aging. *Int Psychogeriatr* 1998;10:271–302.
- [61] Liu H, Harker JO, Wong AL, Maclean CH, Bulpitt KJ, Mittman BS, et al. Case finding for population-based studies of rheumatoid arthritis: comparison of patient self-reported ACR criteria-based algorithms to physician-implicit review for diagnosis of rheumatoid arthritis. *Semin Arthritis Rheum* 2004;33:302–10.
- [62] Duara R, Loewenstein DA, Greig M, Acevedo A, Potter E, Appel J, et al. Reliability and validity of an algorithm for the diagnosis of normal cognition, mild cognitive impairment, and dementia: implications for multicenter research studies. *Am J Geriatr Psychiatry* 2010;18:363–70.
- [63] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the physician. *J Psychiatr Res* 1975;12:189–98.
- [64] Fuld PA. Fuld Object-Memory Evaluation. Wood Dale, IL: Stoelting Co; 1981.
- [65] Wechsler. The Wechsler Memory Scale-Revised. San Antonio, TX: The Psychological Corporation; 1987.
- [66] Spreen O, Strauss E. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. 2nd ed. New York, NY: Oxford University Press; 1998.
- [67] Acevedo A, Loewenstein DA, Barker W, Harwood DF, Luis C, Bravo M. Category fluency test: normative data for English- and Spanish-speaking elderly. *J Int Neuropsychol Soc* 2000;6:760–9.

- [68] Wechsler D. The Wechsler Adult Intelligence-Revised. New York, NY: Psychological Corporation; 1981.
- [69] Army Individual Test Battery. Manual of Directions and Scoring. Washington, DC: War Department, Adjutant General's Office; 1944.
- [70] Loewenstein DA, Acevedo A, Agron J, Martinez G, Duara R. The use of amnesic and nonamnesic composite measures at different thresholds in the neuropsychological diagnosis of MCI. *J Clin Exp Neuropsychol* 2007;29:300–7.
- [71] Acevedo A, Loewenstein DA, Agrón J, Duara R. Influence of socio-demographic variables on neuropsychological test performance in Spanish-speaking older adults. *J Clin Exp Neuropsychol* 2007; 29:530–44.
- [72] Loewenstein DA, Acevedo A, Ownby R, Agrón J, Barker WW, Issacson RS, et al. Using different memory cut-offs to assess MCI. *Am J Geriatr Psychiatry* 2006;4:911–9.
- [73] Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993;43:2412–4.
- [74] Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *Int Psychogeriatr* 1997;9(Suppl 1):173–6.
- [75] Bobinski M, Wegiel J, Wisniewski HM, Tarnawski M, Bobinski M, Reisberg B, et al. Neurofibrillary pathology—correlation with hippocampal formation atrophy in Alzheimer disease. *Neurobiol Aging* 1996;17:909–19.
- [76] Jack CR Jr, Dickson DW, Parisi JE, Xu YC, Cha RH, O'Brien PC, et al. Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. *Neurology* 2002;58:750–7.
- [77] Järvenpää T, Laakso MP, Rossi R, Koskenvuo M, Kaprio J, Räähä I, et al. Hippocampal MRI volumetry in cognitively discordant monozygotic twin pairs. *J Neurol Neurosurg Psychiatry* 2004;75:116–20.
- [78] Duara R, Loewenstein DA, Potter E, Appel J, Greig MT, Urs R, et al. Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer's disease. *Neurology* 2008;71:1986–92.
- [79] Ye S, Huang Y, Müllendorff K, Dong L, Giedt G, Meng EC, et al. Apolipoprotein (apo) E4 enhances amyloid beta peptide production in cultured neuronal cells: apoE structure as a potential therapeutic target. *Proc Natl Acad Sci USA* 2005;102:18700–5.
- [80] Barker WW, Luis CA, Kashuba A, Luis M, Harwood DG, Loewenstein D, et al. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the state of Florida brain bank. *Alzheimer Dis Assoc Disord* 2002;16:203–12.
- [81] Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psych Bull* 1968;70:213–20.
- [82] Blacker D, Albert MS, Bassett SS, Go RC, Harrell LE, Folstein MF. Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. The National Institute of Mental Health Genetics Initiative. *Arch Neurol* 1994;51:1198–204.
- [83] Lopez OL, Becker JT, Klunk WE, Saxton JA, Hamilton RL, Kaufer DI, et al. Research evaluation and diagnosis of probable Alzheimer's disease over the last two decades. *Neurology* 2000;55:1854–62.
- [84] Lopez OL, Litvan I, Catt KE, Stowe R, Klunk WE, Kaufer DI, et al. Accuracy of four clinical diagnostic criteria for the diagnosis of neurodegenerative dementias. *Neurology* 1999;53:1292–9.
- [85] Lopez OL, Kuller LH, Fitzpatrick A, Ives D, Becker JT, Beauchamp N. Evaluation of dementia in the Cardiovascular Health Cognition Study. *Neuroepidemiology* 2003;22:1–12.
- [86] Shumaker SA, Legault C, Rapp SR, Thal L, Wallace RB, Ockene JK, et al. Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. *JAMA* 2003; 289:2651–62.
- [87] Ganguli M, Snitz B, Vander Bilt J, Chang CC-H. How much do depressive symptoms affect cognition at the population level? The Monongahela-Youghiogheny Healthy Aging Team (MYHAT) study. *Int J Geriatr Psychiatry* 2009;24:1277–84.
- [88] Ganguli M, Snitz BE, Lee CW, Vanderbilt J, Saxton JA, Chang CCH. Age and education effects and norms on a cognitive test battery from a population-based cohort: the Monongahela-Youghiogheny Healthy Aging Team. *Aging Ment Health* 2010;14:100–7.
- [89] Available at: <http://alzheimer.wustl.edu/cdr/default.htm>.
- [90] Zapletal E, Le Bozec C, Degoulet P, Guinebretiere J-M, Jaulent M-C. Specifications and implementation of a new exchange format to support computerized consensus in pathology. In: Fieschi M, ed. *MED-INFO 2004*. Amsterdam, The Netherlands: IOS Press; 2004. p. 693–7.
- [91] Araujo CPS, Del Pino MAP, Baez PG, Lopez PF. Clinical web environment to assist the diagnosis of Alzheimer's disease and other dementias. *Proceedings of the 4th WSEAS Conference on Applied Informatics and Communications*. WSEAS Trans Comput 2004; 3:2083–8.