J Prev Alz Dis 2024;
Published online April 2, 2024, http://dx.doi.org/10.14283/jpad.2024.71

Original Research

Evaluating the Performance of Different Criteria in Diagnosing AD and Preclinical AD with the Bayesian Latent Class Model

X. Wang[1,2,], G. Niu[1,2,*], J. Zhao[3], H. Zhu[4,5], F. Li[1,2], J. Tian[1,2], Z. Zhang[1], G. Chen[6], Y. He[1,2], Q. Gao[1,2]*

1. Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing, P. R. China; 2. Beijing Municipal Key Laboratory of Clinical Epidemiology, Capital Medical University, Beijing, P. R. China; 3. School of Applied Science, Beijing Information Science and Technology University, Beijing, 100192, China; 4. Department of Maternal and Child Health, School of Public Health, Capital Medical University, Beijing, P. R. China; 5. Laboratory for Gene-Environment and Reproductive Health, Laboratory for Clinical Medicine, Capital Medical University, Beijing, P. R. China; 6. Department of Neurology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China; * Xiaonan Wang and Guanyu Niu contributed equally to this work.

Corresponding Author: Qi Gao, Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, No.10 Xitoutiao, You'anmen Wai, Fengtai District, Beijing, 100069, China. Email: gaoqi@ccmu.edu.cn, Telephone: 086-010-83911497

Abstract

BACKGROUND: The diagnostic criteria for Alzheimer's disease (AD) should be highly sensitive and specific. Clinicians have varying opinions on the different criteria, including the International Working Group-1 (IWG-1), International Working Group-2 (IWG-2), and AT(N) criteria. Few studies had evaluated the performance of these criteria in diagnosing AD and preclinical AD when the gold standard was absent.

METHODS: We estimated and compared the performance of these criteria in diagnosing AD using data from 908 subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI). Additionally, 622 subjects were selected to evaluate and compare the performance of IWG-2 and AT(N) criteria in diagnosing preclinical AD. A novel approach, Bayesian latent class models with fixed effect dependent, was utilized to estimate the diagnostic accuracy of these criteria in detecting different AD statuses simultaneously.

RESULTS: The sensitivity of the IWG-1, IWG-2, and AT(N) criteria in diagnosing AD was 0.850, 0.836, and 0.665. The specificity of these criteria was 0.788, 0.746, and 0.747. The IWG-1 criteria had the highest Youden Index in detecting AD. When diagnosing preclinical AD, the sensitivity of the IWG-2 and AT(N) criteria was 0.797 and 0.955. The specificity of these criteria was 0.922 and 0.720. The IWG-2 criteria had the highest Youden Index.

CONCLUSION: IWG-1 was more suitable than the IWG-2 and AT(N) criteria in detecting AD. IWG-2 criteria was more suitable than AT(N) criteria in detecting preclinical AD.

Key words: Alzheimer's disease, diagnostic accuracy, IWG-1, IWG-2, AT(N).

Abbreviations: AD: Alzheimer's disease; IWG-1: International Working Group-1; IWG-2: International Working Group-2; ADNI: Alzheimer's Disease Neuroimaging Initiative; CSF: cerebrospinal fluid; ADI: Alzheimer's Disease International; NINCDS-ADRDA: Neurological Disorders and Speech Disorders and Stroke - Alzheimer's Disease and Related Disorders Association; IWG: International Working Group; NIA-AA: National Institute on Aging and the Alzheimer's Association; Aβ: beta-amyloid; PET: positron emission tomography; MRI: Magnetic Resonance Imaging; FDG: fluorodeoxyglucose; RAVLT: Rey Auditory Verbal Learning Test; P-tau: phosphorylated tau protein; T-tau: total tau protein; SUVR: standardized uptake ratio; MCMC: Markov chain Monte Carlo; CI: credible intervals; PPV: positive predictive value; NPV: negative predictive value; MCI: mild cognitive impairment.

Introduction

More than 55 million people worldwide are suffering from dementia, and this number is expected to reach 78 million by 2030. Alzheimer's disease (AD) is an irreversible neurodegenerative disease characterized by progressive memory loss and cognitive impairment (1, 2). It is the leading cause of dementia, accounting for 60%-70% of all cases (3). However, the exact causes of AD are not still known. There is no curative treatment for the disease and prevention remains a priority for AD. Preclinical AD is the silent stage of AD, in which abnormal biomarkers (cerebrospinal fluid (CSF) amyloid-$\beta_{1\text{-}42}$, CSF tau, etc.) are present but symptoms are not yet clinically evident in cognitively normal individuals (4). Appropriate intervention at this stage could delay or even prevent the onset of cognitive impairment and dementia. Thus, accurate identification of patients with different stages of AD is of great significance in clinical practice.

Alzheimer's Disease International (ADI) estimates that globally 75% of people with dementia are not diagnosed; this may be as high as 90% in some low and middle-income countries (1). The gold standard for diagnosing AD is still autopsy (5), which is not available for the living. Several different criteria have been proposed for the diagnosis of AD and preclinical AD. The earliest diagnostic criteria for AD was published by the National Institute of Neurological Disorders and Speech Disorders and Stroke - Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) in 1984 (6). It was widely considered as the reference standard to assess the performance of newly developed criteria for diagnosing AD in previous studies. Given the absence of

Received January 11, 2023
Accepted for publication March 3, 2024

viable biological biomarkers, NINCDS-ADRDA heavily relied on patient medical history, clinical experience, and neuropsychological assessments for accurate diagnosis. As understanding of biology of AD and its progression has improved, the International Working Group (IWG) refined the NINCDS-ADRDA criteria and published the IWG-1 criteria in 2007 (7). This criteria represented two significant advancements: firstly, the conceptualization of AD as a dynamic process characterized by developmental changes; and secondly, the incorporation of biological markers in the diagnostic criteria for AD for the first time. The diagnosis of subjects using IWG-1 criteria required the presence of clinical symptoms in conjunction with biomarkers. However, the IWG-1 criteria did not explore the reliability of biological markers further. In 2010, IWG further proposed several emerging concepts regarding AD based on the IWG-1 criteria, such as preclinical AD and prodromal AD (8). In 2014, the IWG categorized biomarkers based on the original IWG-1 criteria and developed the IWG-2 criteria (9). With the growing range of biomarkers associated with AD, the National Institute on Aging and the Alzheimer's Association (NIA-AA) have proposed AT(N) criteria, which emphasize the use of biomarkers for diagnostic individuals in 2018 (10). The AT(N) criteria classify biomarkers into three categories: A, T, and N. The "A" category represents beta-amyloid (Aβ) deposition, which includes amyloid positron emission tomography (PET) and measurements of Aβ levels in the CSF; The "T" category stands for tau pathology, which involves the measurement of tau protein levels in the CSF and the use of tau PET imaging; The "N" category encompasses neurodegeneration and neuronal injury, which include structural Magnetic Resonance Imaging (MRI), fluorodeoxyglucose (FDG) PET scans and CSF. It is important to note that while the AT(N) criteria offered valuable insights into AD diagnosis, they were not intended to replace clinical evaluation.

Due to the absence of a gold standard, few studies have discussed and compared the performance of IWG-1, IWG-2, and AT(N) in diagnosing AD and preclinical AD simultaneously. Usually, the imperfect standard was used to evaluate the accuracy of these criteria in detecting AD or preclinical AD. Bouwman et al. conducted the study on 452 patients recruited from the Alzheimer Center of the VU Medical Center and revealed that the IWG-1 criteria achieved a sensitivity of 95% and a specificity of 86% in detecting AD with the NINCDS-ADRDA criteria serving as the gold standard (11). Wang et al. conducted the study on the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort and showed that the sensitivity of the IWG-2 criteria for diagnosing AD was 84%, with a specificity of 76% using the NINCDS-ADRDA criteria serving as the gold standard (12). Kern et al. compared the AT(N) criteria with the criteria proposed by Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (AIBL) and found that its sensitivity and specificity for diagnosing AD were 67% and 75% (13). To our knowledge, few studies had directly compared the ability of these diagnostic criteria to diagnose preclinical AD. Most studies of preclinical AD have focused on the accuracy of different types of biomarkers in the diagnosis of preclinical AD (14, 15). A direct comparison between the criteria under the same standard is lacking, and it remains unclear which criteria are best for diagnosing AD and preclinical AD. In addition, there is evidence that ignoring imperfect reference standards can lead to substantial bias, which is known in the field of statistics as imperfect gold standard bias (16).

Therefore, we aimed to correct the imperfect gold standard bias, and to evaluate the performance of three diagnostic criteria (IWG-1, IWG-2, and AT(N)) for diagnosing AD, and assess the accuracy of two diagnostic criteria (IWG-2 and AT(N)) in detecting preclinical AD simultaneously without a gold standard.

Materials and Methods

Subjects

Data used in our study were obtained from the ADNI (https://adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. It is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and follow-up of AD. To date, there were 4 cohorts: ADNI-1, ADNI-GO, ADNI-2, and ADNI-3 in this database. Each cohort recruited new participants in North America and completed clinical assessments, lumbar puncture, imaging examinations, and so on. The ADNI was conducted with the approval of the institutional review board at each site, and all subjects had signed an informed consent form (ClinicalTrials.gov registry numbers: ADNI GO: NCT01078636; ADNI 2: NCT0123197; ADNI 3: NCT02854033).

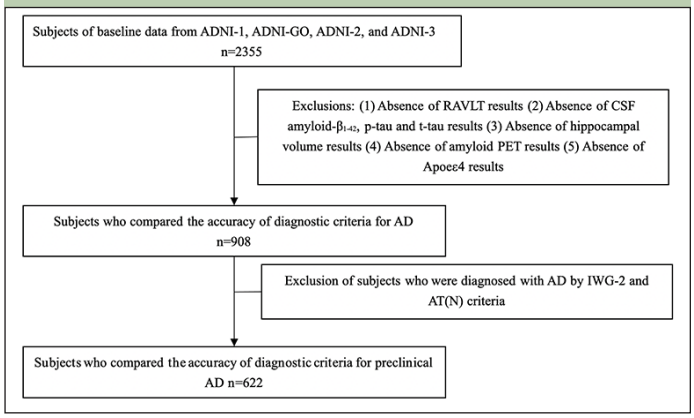
Our study included a total of 2355 individuals from the ADNI-1, ADNI-GO, ADNI-2 and ADNI-3 cohorts. Subjects' baseline data were used in this study. Subjects were excluded if they had never completed cognitive assessment or lumbar puncture or imaging examinations, or were missing gene data. The final sample included 908 individuals to compare the diagnostic accuracy of IWG-1, IWG-2, and AT(N) in detecting AD without a gold standard. Subsequently, we excluded subjects diagnosed with AD by the IWG-2 and AT(N) criteria, leaving 622 subjects, to assess the accuracy of the IWG-2 and AT(N) criteria in diagnosing preclinical AD. The specific process was shown in Figure 1.

Clinical assessment

Clinical assessment for the status of cognitive function was performed according to the Rey Auditory Verbal Learning Test (RAVLT), which could respond well to

the changes in the status of subject's cognitive function (17). The memory tests and corresponding scores can be found on the website (http://adni.loni.usc.edu/uploaddata/). The RAVLT results were converted to Z-scores with results below -1.25 SD as an indication of cognitive impairment, which was consistent with some previous studies (12, 18, 19).

Figure 1. Flow chart of subjects' inclusion and exclusion criteria



Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; RAVLT, Rey Auditory Verbal Learning Test; CSF, cerebrospinal fluid; p-tau, phosphorylated tau at threonine 181; t-tau, total tau; PET, positron emission tomography.

Biomarker assessment

CSF results included amyloid ($A\beta_{1-42}$), phosphorylated tau protein (p-tau), and total tau protein (t-tau). Juan Fortea showed the processing details for CSF (18). We used cut-off values to define abnormal biomarkers, and $A\beta_{1-42}$ <192pg/ml or t-tau > 93pg/ml or p-tau > 23pg/ml were defined as abnormal (20, 21).

For MRI, hippocampal atrophy was an MRI-related marker in our study. We specifically defined hippocampal atrophy as a condition where the volume of the hippocampus measured less than 5.33 cm^3 (22).

The PET data for the amyloid tracer florbetapir (AV-45) were obtained from the AV45 analysis dataset on the website, from the University of California, Berkeley. We used standardized uptake ratio (SUVR) based on the entire cerebellar reference region with a threshold of 1.11, and we defined SUVR< 1.11 as abnormal (http://adni.loni. usc.edu/upload-data/).

Subjects classification

According to the IWG-1 criteria, subjects were classified as 'AD group' if they had cognitive impairment and at least one abnormal biomarker. Conversely, the remaining subjects were assigned to the non-AD group (8) (Table 1).

According to the IWG-2 criteria, subjects were classified as AD group if they had cognitive impairment along with abnormal CSF $A\beta_{1-42}$ and either abnormal p-tau or t-tau. The remaining subjects were classified into the non-AD group. In the diagnosis of preclinical AD, individuals with normal cognitive function but abnormal CSF $A\beta_{1-42}$ and abnormal p-tau or t-tau were classified into the preclinical AD group. The remaining subjects were reclassified into non-preclinical AD group (9) (Table 1).

For the AT(N) criteria in our study, the biomarkers for Aβ plaques (labeled "A") were CSF $A\beta_{1-42}$ and SUVR measured by amyloid PET , and the markers for abnormality were reduced levels of CSF $A\beta_{1-42}$ or reduced SUVR (labeled "A+"); biomarkers of fibrillar tau (labeled "T") is CSF p-tau, with elevated levels of CSF p-tau denoting abnormal values (labeled "T+"); biomarkers of neurodegeneration and neuronal injury (labeled "N") were CSF t-tau and hippocampal volume, with increased levels of CSF p-tau or decreased hippocampal volume indicating outliers (labeled "N+"). According to AT(N) criteria, we classified subjects who were in the A+T+N+, A+T+N-, and A+T-N+ groups as AD group, the remaining subjects were categorized as non-AD group. When diagnosing preclinical AD, we defined A+T-N- as the preclinical AD group, and classified A-T-N-, A-T+N+, A-T+N-, and A-T-N+ as the non-preclinical AD group (10) (Table 1).

Statistical analysis

Quantitative variables were presented as means and SDs, and qualitative variables were shown as numbers (proportions, %). T-test was used to compare quantitative variables following normal distribution. The Wilcoxon rank sum test was used to compare quantitative variables not following normal distribution. The Chi-square test was used to compare the categorical variables.

IWG-2 was revised on the base of IWG-1 criteria, the diagnostic results of these two criteria might be correlated in detecting AD status. Therefore, the latent class model with fixed effect dependence was established to evaluate the accuracy of IWG-1, IWG-2, and AT(N) criteria for diagnosing AD in the absence of a gold standard, details of model can be found in the supplementary materials (Latent class model with fixed effect dependent and Table S3). IWG-2 and AT(N) criteria were independent in detecting preclinical AD. Latent class model with independence assumption was used to assess the performance of IWG-2 and AT(N) in diagnosing preclinical AD. Bayesian method was used to estimate the parameters of the established models. The Bayesian approach combined the likelihood function given the data and prior distributions for the model parameters via Bayes' theorem to get the posterior distribution of the parameters. We followed this approach because prior scientific information about the diagnosis accuracy of three criteria can be incorporated in our estimation. The prior information of the model parameters was obtained using data from previous studies (11-15, 23-25). Since there were no previous studies directly comparing the

Table 1. Classification according to the IWG-1, IWG-2, and AT(N) criteria

Criteria	Definition
IWG-1	
Alzheimer's disease	Cognitive impairment, at least one abnormal Alzheimer's disease biomarker
Normal	No cognitive impairment, normal Alzheimer's disease biomarker
IWG-2	
Alzheimer's disease	Cognitive impairment, abnormal CSF amyloid-β_{1-42} and p-tau or t-tau
Preclinical Alzheimer's disease	Cognitive normal, abnormal CSF amyloid-β_{1-42} and p-tau or t-tau
Normal	Cognitive normal, normal CSF amyloid-β_{1-42}, t-tau and p-tau
AT(N)	
Alzheimer's disease	
(A+T+N+)	Abnormal CSF amyloid-β_{1-42} or amyloid PET, abnormal CSF p-tau, abnormal CSF t-tau or anatomic MRI
(A+T+N−)	Abnormal CSF amyloid-β_{1-42} or amyloid PET, abnormal CSF p-tau, normal CSF t-tau or anatomic MRI
(A+T−N+)	Abnormal CSF amyloid-β_{1-42} or amyloid PET, normal CSF p-tau, abnormal CSF t-tau or anatomic MRI
Preclinical Alzheimer's disease	
(A+T−N−)	Abnormal CSF amyloid-β_{1-42} or amyloid PET, normal CSF p-tau, normal CSF t-tau or anatomic MRI
Normal	
(A−T−N−)	Normal CSF amyloid-β_{1-42} or amyloid PET, normal CSF p-tau, normal CSF t-tau or anatomic MRI

Cognitive impairment is defined as Z-score< -1.25SD. "+" represents abnormal biomarkers and "-" represents normal biomarkers. Abnormal CSF amyloid-β_{1-42} is defined as CSF amyloid-β_{1-42} < 192pg/ml; abnormal CSF t-tau is defined as CSF t-tau > 93pg/ml; abnormal CSF p-tau is defined as CSF p-tau > 23pg/ml; abnormal amyloid PET is defined as Summary SUVR <1.1; abnormal anatomic MRI is defined as hippocampus volume < 5.33cm^3. Abbreviations: IWG, International Working Group; AD, Alzheimer's disease; CSF, cerebrospinal fluid; p-tau, phosphorylated tau at threonine 181; t-tau, total tau; PET, positron emission tomography.

sensitivity and specificity of the IWG-2 and AT(N) criteria for the diagnosis of preclinical AD. However, these two diagnostic criteria mainly used biomarkers to diagnosis of preclinical AD, therefore, we took the sensitivity and specificity of diagnostic models with similar biomarkers to these two criteria as a priori information for diagnosing preclinical AD. The prior densities for sensitivity, specificity, and prevalence were assumed to be independent Beta distributions. The prior densities for the correlations between diagnostic results of IWG-1 and IWG-2 were assumed to be uniform distribution. The detailed hyper-parameters for prior information were shown in Table S3. Gibbs sampling method in Markov chain Monte Carlo (MCMC) was used to estimate the parameters from the approximate posterior distribution. WinBUGS code implemented the inference of parameters in Bayesian latent class model with fixed effect dependence. The posterior distributions were computed based on 10,000 iterations after discarding the initial 5,000 iterations as burn-in and 95% Bayesian credible intervals (95% CI) were evaluated.

Statistical analysis was done via R 4.1.3 with significance set at P<0.05. Bayesian data analysis was performed with WINBUGS 1.4.3.

Results

Sample Demographics and classification results in diagnosing AD

Among the 908 participants in comparing the diagnostic accuracy of IWG-1, IWG-2, and AT(N) for detecting AD, there were 421 females and 487 males with a mean age of 72.65±7.12 years (Table 2). Table 3 showed the classification and characteristics of subjects according to the criteria. 99 subjects with a mean age of 74.23±7.78 years had AD according to the IWG-1 criteria, and 77 subjects with a mean age of 72.41±7.00 years according to the IWG-2 criteria were assigned to the AD group. According to AT(N) criteria, 267 subjects were assigned with a mean age of 73.33±7.06 years to the AD group, of which 124 were female (Table 3, Figure S1). When AD was diagnosed according to IWG-1 and IWG-2 criteria, AD patients had a shorter mean duration of education compared with non-AD patients (15.67 vs 16.31 years, 15.68 vs 16.25 years), and there was a substantial decrease in RAVLT scores (18.08 vs 40.18, 17.87 vs 39.54). When AD was diagnosed according to three criteria, AD patients had higher CSF p-tau, t-tau, and SUVR,lower CSF Aβ and hippocampal volume.

Sample Demographics and classification results in diagnosing preclinical AD

A total of 622 individuals, comprising 292 females and 330 males, were enrolled in this study to evaluate the accuracy of IWG-2 and AT(N) criteria for preclinical AD. The mean age of participants was 72.37±7.10 years (Table 2). Table 4 presented the classification and characteristics of subjects based on diagnostic criteria for preclinical AD. Of these individuals, 480 subjects with a mean age of 71.50±7.20 met IWG-2 criteria and were assigned to a preclinical AD group, while 357 subjects with a mean age of 74.87±7.06 met AT(N) criteria and were assigned to a preclinical AD group (Table 4, Figure S2). When subjects

Table 2. Demographic information of different population groups

		Diagnosis of AD population (n=908)	Diagnosis of preclinical AD population (n=622)
Age (Mean±SD, years)		72.65(7.12)	72.37(7.10)
Female (n, %)		421(46.37)	292(46.95)
Education (Mean±SD,years)		16.24(2.65)	16.30(2.64)
RAVLT(Mean±SD)		37.77(12.55)	41.19(11.68)
Aβ (Mean±SD, pg/ml)		242.94(88.59)	282.63(77.74)
P-tau (Mean±SD, pg/ml)		24.76(13.25)	19.82(9.13)
T-tau (Mean±SD, pg/ml)		78.56(45.33)	61.52(30.27)
Hippocampus(Mean±SD, mm^3)		7496.29(1031.33)	7699.39(997.69)
Apoeε4	0	514	437
	1	315	164
	2	79	21
Summary SUVR (Mean±SD)		1.23(0.25)	1.14(0.20)

Table 3. Demographic and clinical outcomes for the diagnosis of AD according to IWG-1, IWG-2 and AT(N) criteria

		IWG-1			IWG-2			AT(N)		
		AD (n=99)	Non-AD (n=809)	P value	AD (n=77)	Non-AD (n=831)	P value	AD (n=267)	Non-AD (n=641)	P value
Age (Mean±SD, years)		74.73(7.78)	72.41(7.00)	<0.05	73.97(7.95)	72.53(7.03)	0.06	73.33(7.06)	72.39(7.13)	<0.05
Female (n, %)		27(27.27)	394(48.70)	<0.05	23(29.87)	398(47.89)	<0.05	124(46.44)	297(46.33)	0.97
Education (Mean±SD, years)		15.67(2.75)	16.31(2.63)	<0.05	15.68(2.70)	16.25(2.64)	<0.05	16.16(2.61)	16.27(2.66)	0.48
RAVLT(Mean±SD)		18.08(3.34)	40.18(11.05)	<0.05	17.87(3.35)	39.54(11.44)	<0.05	31.13(11.00)	40.53(12.12)	<0.05
Aβ (Mean±SD, pg/ml)		182.48(60.28)	250.34(88.69)	<0.05	162.97(36.27)	250.45(88.36)	<0.05	154.18(26.86)	279.91(78.53)	<0.05
P-tau (Mean±SD, pg/ml)		32.45(14.67)	23.82(12.76)	<0.05	34.78(14.89)	24.00(12.71)	<0.05	36.2(14.3)	19.99(9.31)	<0.05
T-tau (Mean±SD, pg/ml)		114.64(55.54)	74.14(41.87)	<0.05	121.76(54.34)	75.02(42.25)	<0.05	117.88(49.84)	62.18(31.07)	<0.05
Hippocampus(Mean±SD, mm^3)		6911.31(981.30)	7567.87(1014.95)	<0.05	6896.64(939.40)	7564.74(1022.29)	<0.05	7094.91(1009.84)	7663.48(994.15)	<0.05
Apoeε4	0	38	476	-	16	498	-	77	437	-
	1	46	269		46	269		138	177	
	2	15	64		15	64		52	27	
Summary SUVR (Mean±SD)		1.40(0.22)	1.21(0.24)	<0.05	1.45(0.18)	1.21(0.24)	<0.05	1.44(0.21)	1.15(0.21)	<0.05

Abbreviations: AD, Alzheimer's disease; RAVLT, Rey Auditory Verbal Learning Test, Aβ, β-amyloid$_{1-42}$; p-tau, phosphorylated tau at threonine 181; t-tau, total tau; Summary SUVR, based on the whole cerebellum reference region (cortical composite region intensity normalized by the FreeSurfer-defined whole cerebellum); SUVR, standard uptake value ratios; IWG, International Working Group.

were diagnosed according to the IWG-2 and AT(N) criteria, higher Aβ, and lower p-tau, t-tau and SUVR were found compared to non-preclinical AD patients. Patients with preclinical AD were younger than those without preclinical AD according to the IWG-2 criteria.

Cross-classification results of these criteria in diagnosing AD and preclinical AD

Table S1 showed the cross-classification results of the three diagnostic criteria for diagnosing AD among the participants. The value "1" represented a diagnosis of AD and "0" represented a diagnosis of non-AD. A total of 19 individuals were diagnosed with AD by all three diagnostic criteria, while 600 were diagnosed with non-AD by all three diagnostic criteria. 58 individuals were diagnosed with AD by the IWG-1 and IWG-2 criteria but not by the AT(N) criteria, and 209 individuals were diagnosed with AD by the AT(N) criteria but not by the IWG-1 and IWG-2 criteria.

Table S2 presented the cross-classification results of the IWG-2 and AT(N) for diagnosing preclinical AD among the participants. The table showed that 329 subjects were diagnosed with preclinical AD by IWG-2 and AT(N) criteria, while 114 subjects were diagnosed with non-preclinical AD by the two diagnostic criteria.

Accuracy of three criteria in diagnosing AD

The prevalence, sensitivity, and specificity of the three diagnostic criteria for AD were presented in Table 5. The positive predictive value (PPV), negative predictive value (NPV), and Youden index of the three diagnostic criteria for AD were presented in Figure 2. The prevalence of AD was estimated at 1.5% (95%Cl 0.008 to 0.026). Among the three diagnostic criteria for diagnosing AD, the IWG-1 criteria exhibited the highest sensitivity with a value of 0.850 (95% CI: 0.798 to 0.930) and a specificity of 0.788 (95% CI: 0.716 to 0.848). The specificity of the AT(N) criteria was similar to that of the IWG-2 criteria. It is worth noting that all three diagnostic criteria had low

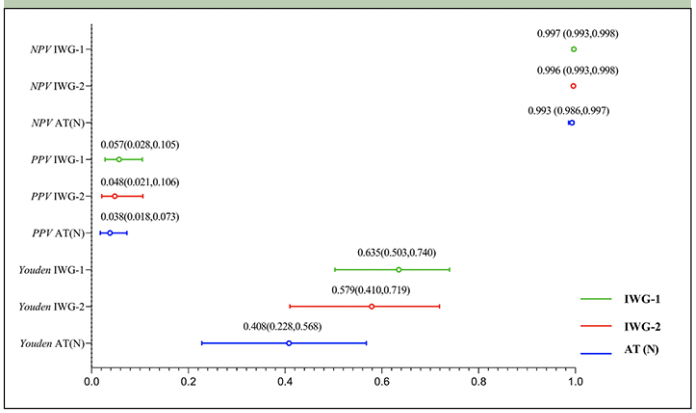
Table 4. Demographic and clinical outcomes for the diagnosis of preclinical AD according to IWG-2 and AT(N) criteria

		IWG-2			AT(N)		
		Preclinical AD (n=480)	Non-preclinical AD (n=142)	P value	Preclinical AD (n=357)	Non-preclinical AD (n=265)	P value
Age (Mean±SD, years)		71.50(7.20)	74.94(5.70)	<0.05	74.87(7.06)	73.04(7.12)	0.05
Female (n, %)		226(47.08)	66(46.48)	0.90	165(46.22)	127(47.92)	0.67
Education (Mean±SD,years)		16.43(2.62)	15.85(2.68)	<0.05	16.11(2.73)	16.35(2.62)	0.46
RAVLT(Mean±SD)		42.30(11.10)	40.87(10.76)	<0.05	36.67(9.12)	40.00(11.63)	0.06
Aβ (Mean±SD, pg/ml)		294.55(77.25)	240.34(62.51)	<0.05	293.01(84.15)	268.65(65.76)	<0.05
P-tau (Mean±SD, pg/ml)		19.04(8.20	22.09(11.58)	<0.05	15.43(3.53)	25.71(10.88)	<0.05
T-tau (Mean±SD, pg/ml)		58.07(25.72)	70.26(36.67)	<0.05	49.87(16.10)	77.22(37.08)	<0.05
Hippocampus(Mean±SD, mm^3)		7719.21(963.66)	7518.49 (1094.48)	<0.05	7703.02(936.80)	7633.48(1047.82)	0.27
Apoeε4	0	295	142	-	268	169	
	1	164	0		76	88	
	2	21	0		13	8	
Summary SUVR (Mean±SD)		1.09(0.17)	1.32(0.19)	<0.05	1.07(0.18)	1.23(0.21)	<0.05

Abbreviations: AD, Alzheimer's disease; RAVLT, Rey Auditory Verbal Learning Test, Aβ, β-amyloid$_{1-42}$; p-tau, phosphorylated tau at threonine 181; t-tau, total tau; Summary SUVR, based on the whole cerebellum reference region (cortical composite region intensity normalized by the FreeSurfer-defined whole cerebellum); SUVR, standard uptake value ratios; IWG, International Working Group.

PPV, suggesting a higher likelihood of false positives. Conversely, the NPV for the IWG-1 and IWG-2 criteria were similar . Additionally, the IWG-1 criteria had the largest Youden index, indicating its superior overall diagnostic performance.

Figure 2. The performance of the different diagnostic criteria for the diagnosis AD

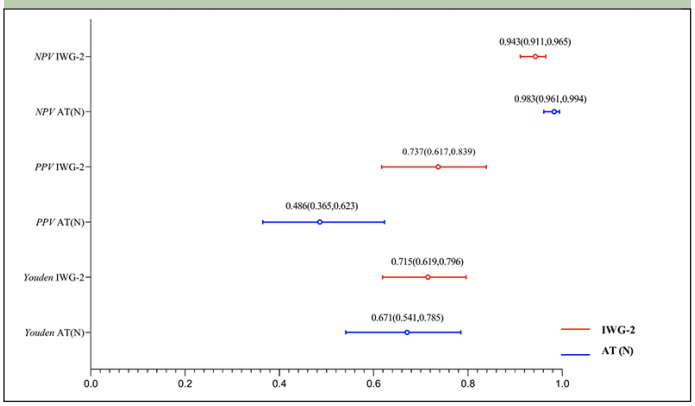


Abbreviations: AD, Alzheimer's disease; NPV, Negative predictive value; PPV, Positive predictive value; Youden, Youden Index; CI, Confidence interval; IWG, International Working Group

Accuracy of two criteria in diagnosing preclinical AD

Table 5 showed the prevalence, sensitivity, and specificity for the IWG-2 and AT(N) criteria in the diagnosis of preclinical AD. The PPV, NPV, and Youden index for the IWG-2 and AT(N) criteria in the diagnosis of preclinical AD were presented in Figure 3. The prevalence of preclinical AD was estimated to be 21.6% (95% CI: 0.168 to 0.271). The AT(N) criteria exhibited the highest sensitivity of 0.955(95% CI: 0.905 to 0.983) for detecting preclinical AD, indicating its ability to identify a significant proportion of individuals with the condition. In comparison, the IWG-2 criteria demonstrated the highest specificity of 0.922 (95% CI: 0.879 to 0.954). The IWG-2 criteria displayed the highest PPV of the three diagnostic criteria, with a value of 0.737 (95% CI: 0.617 to 0.839), indicating a relatively higher likelihood of true positives. The AT(N) criteria exhibited the highest NPV at 0.983 (95% CI: 0.961 to 0.994), suggesting a lower probability of false negatives. The IWG-2 criteria yielded the larger Youden index compared to AT(N) criteria, with a value of 0.715 (95% CI: 0.619 to 0.796), indicating its superior overall diagnostic performance.

Figure 2. The performance of the different diagnostic criteria for the diagnosis A preclinical AD



Abbreviations: AD, Alzheimer's disease; NPV, Negative predictive value; PPV, Positive predictive value; Youden, Youden Index; CI, Confidence interval; IWG, International Working Group

Table 5. Estimation for the performance of these criteria in diagnosing AD and preclinical AD

Parameters	Disease status			
	AD		Preclinical AD	
	Estimated value	95% CI	Estimated value	95% CI
Prevalence	0.015	(0.008,0.026)	0.216	(0.168,0.271)
Se IWG-1	0.850	(0.738,0.930	-	-
Se IWG-2	0.836	(0.756,0.900)	0.797	(0.705,0.869)
Se AT(N)	0.665	(0.505,0.803)	0.955	(0.905.0.983)
Sp IWG-1	0.788	(0.716,0.848)	-	-
Sp IWG-2	0.746	(0.595,0.865)	0.922	(0.879,0.954)
Sp AT(N)	0.747	(0.652,0.825)	0.720	(0.594,0.829)

Discussion

This was the first study to simultaneously compare the performance of IWG-1, IWG-2 and AT(N) criteria in diagnosing AD and to compare the performance of IWG-2 and AT(N) criteria in diagnosing preclinical AD without a gold standard. The findings revealed that IWG-1 criteria exhibited the highest Youden index, suggesting superior overall diagnostic performance in detecting AD. Conversely, IWG-2 criteria demonstrated the largest Youden index in diagnosing preclinical AD, indicating their superior ability to accurately detect individuals in the early stages of the disease. These results provided valuable insights into the comparative strengths of these diagnostic criteria for distinguishing AD vs non-AD and preclinical AD vs non- preclinical AD.

Our study estimated the prevalence of AD and preclinical AD to be 1.5% and 21.6%, respectively, which was the similar as the findings reported in other studies (1, 26). Notably, when comparing our results to the data presented in the 2018 World Alzheimer's Report (27), a significant increase in the number of the AD individuals becomes evident. This indicated a significant upward trend in the prevalence of AD cases over time.

The IWG-1 criteria exhibited the highest sensitivity and specificity among the three diagnostic criteria for detecting AD. The IWG-2 criteria were revised from the IWG-1 criteria, and emphasized the critical role of Aβ protein in the AD disease process. Nevertheless, the sensitivity and specificity of the IWG-2 criteria for the diagnosis of AD were lower than those of the IWG-1 criteria. This discrepancy highlighted the imperfection of relying only on the combination of Aβ protein and tau protein as supplemental information. This phenomenon might be due to incomplete understanding of the underlying pathologic process of AD, so more biomarkers need to be explored to refine criteria. The specificities of the IWG-2 and AT(N) criteria were comparable, yet their sensitivities diverged. While both criteria involved biomarker categorization, their methodologies differed significantly. Specifically, the IWG-2 criteria classified biomarkers into two distinct categories, whereas the AT(N) criteria employed a three-category system. Notably, this divergence become more pronounced when considering their utilization of biomarkers: the IWG-2 criteria treated biomarkers as a supplementary information, while the AT(N) criteria relied exclusively on biomarkers for subject diagnosis. This likely accounted for the heightened sensitivity observed in the IWG-2 criteria compared to the AT(N) criteria.

The IWG-1 criteria also demonstrated greater NPV, PPV, and Youden Index compared to other diagnostic criteria in detecting AD. Although the IWG-2 criteria were adapted from the IWG-1 criteria, the differences between the NPV and PPV of the two criteria were not significant. The IWG-2 criteria were more specific in terms of the types of biomarkers used in the diagnosis of AD. However, there was a decrease in the NPV but no increase in the PPV. The NPV and PPV of the AT(N) criteria were smaller than the other two criteria, suggesting that the diagnosis of AD still need to rely on scales to test the subject's cognition. The larger NPV and smaller PPV observed across all three criteria might be related to the low prevalence of AD. While more individuals were being diagnosed with AD, particularly among those aged 60 years and above, it still represented a relatively small proportion of the overall population. Consequently, this low prevalence contributed to lower PPV for all three criteria (28).

When diagnosing preclinical AD, on the one hand, the sensitivity and PPV of the IWG-2 were higher than the AT(N) criteria, indicating that IWG-2 was more likely to identify patients with preclinical AD. On the other hand, the specificity and NPV of IWG-2 were lower than the AT(N) criteria, suggesting that the AT(N) criteria was more likely to identify patients without preclinical AD. The difference between the IWG-2 and AT(N) criteria lied in their requirements for diagnosis. The IWG-2 criteria required not only the presence of Aβ protein pathology but also the presence of t-tau or p-tau proteins, while the AT(N) criteria only required the presence of Aβ protein pathology and did not consider other biomarkers or genetic factors. Our results suggested that not only Aβ protein, but also other biomarkers should be considered in the diagnosis of preclinical AD. While the IWG-2 criteria might be less capable of recognizing

non-preclinical AD, its Youden index was higher than that of the AT(N) criteria, indicating its superior overall diagnostic performance. This suggested that the IWG-2 criteria might be more suitable for diagnosing preclinical AD than the AT(N) criteria, but further research is needed to improve both criteria.

In summary, our findings indicated that clinical cognitive assessment results remain the core diagnostic criteria in the diagnosis of AD. It was evident that the identified biomarkers did not encompass the entire pathological process of AD and preclinical AD. Moreover, the specific physiological role of each biomarker in the disease process remained unclear, rendering biomarker models insufficient as standalone diagnostic guidelines for AD diagnosis. Whereas biomarkers played a significant role in the diagnosis of preclinical AD. As the existing biomarker models for diagnosis were imperfect, future research should focus on exploring combinations of different biomarkers to enhance diagnostic accuracy. Overall, while clinical cognitive assessment remained pivotal in AD diagnosis, further investigation into the integration of biomarkers and cognitive assessment is crucial for advancing our understanding and diagnostic capabilities in both AD and preclinical AD.

These three diagnostic criteria served distinct applications in different scenarios. IWG-1 criteria was particularly practical as a definitive clinical criteria, allowing for more accurate identification of patients and minimizing the risk of missed diagnoses in AD diagnosis. In a clinical setting, when determining whether a patient is suffering from preclinical AD, IWG-2 criteria should be employed. In a research setting, the AT(N) criteria was more appropriate, as also concluded by Jack et al (29). These criteria facilitated the exploration of combined biomarker models, expediting the application of biomarkers in clinical diagnosis. This promotes efficient integration of biomarkers into the diagnostic process of AD.

When considering the application of diagnostic criteria, it is essential to take into account not only their accuracy but also their practicality. The collection of biomarkers required for diagnosis was a complex process, with certain tests such as cerebrospinal fluid collection being invasive and not universally acceptable to all patients. Moreover, imaging tests were prohibitively expensive, making them inaccessible to certain families. Therefore, it is imperative to adopt different criteria based on specific situations, ensuring their suitability, and optimizing their utility to better serve patients. By carefully considering both accuracy and practical considerations, diagnostic criteria can be applied in a more effective manner, catering to the diverse needs of patient populations.

Our study also has some limitations. Firstly, we relied on a scale to assess patients' cognitive function, without further refining each item of the scale to determine the cognitive status of the subjects. Future research can strengthen this aspect by conducting more detailed item-level analysis to accurately determine cognitive status. Secondly, it is important to note that our findings were based on data obtained from the ADNI database, and caution should be exercised in generalizing these results to other settings or populations. The specific characteristics and demographics of the ADNI cohort may introduce biases that need to be considered when interpreting the findings. Additionally, our study was cross-sectional, which limited our ability to establish causality or make predictions about disease progression. Future studies utilizing longitudinal designs are needed to assess the comparative diagnostic abilities of different criteria in predicting the conversion of preclinical AD to dementia or the progression from cognitively normal individuals to mild cognitive impairment.

Considering these limitations, further research should aim to refine assessment methodologies, validate findings in diverse settings, and conduct longitudinal studies to explore the predictive capabilities of different diagnostic criteria in tracking disease progression and prognosis.

Conclusion

In conclusion, the IWG-1 criteria was recommended for screening purposes in the diagnosis of AD, while the IWG-2 criteria was more appropriate for precise diagnosis. Specifically, the IWG-2 criteria demonstrated superior applicability compared to AT(N) criteria in the diagnosis of preclinical AD. It is important to note that the AT(N) criteria was primarily applicable within research settings and have not yet been widely implemented in clinical practice.

Acknowledgements: Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wpontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Sources of Funding: This work was supported by National Natural Science Foundation of China (Grant No. 81903408), Beijing Excellent Talents Training Founding Project (Grant No. 2018000020124G136), R&D Program of Beijing Municipal Education Commission (Grant KM202011232018), and Key Research and Cultivation Project of Scientific Research on Campus of Beijing Information Science and Technology University (Grant 2021YJPY236). The funder/sponsor had no role in the design and conduct of the study, collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Statements and Declarations: Ethics approval and consent to participate Subjects for this study were obtained from the ADNI. The ADNI was conducted with the approval of the review boards of the respective research institutions, and all subjects signed an informed consent form.

Availability of data and materials: The datasets generated during the current study are available in the ADNI repository, http://adni.loni. usc.edu/upload-data/.

Conflict of interest: The authors declare no that they have competing interests.

Authors Contributions: X.N.W.: conceptualization, funding acquisition, methodology, data analysis, writing original draft and review & editing. G.Y.N.: methodology, data analysis, writing-review & editing. J.X.Z.: conceptualization, methodology, and writing—review and editing. H.P.Z.: methodology and writing—review and editing. F.J.L.: methodology, data analysis. J. T.: methodology, data analysis. Z.Y.Z: methodology and writing—review and editing. G.Q.C:

writing—review and editing. Y.H.: conceptualization, methodology, project administration, validation, writing—review and editing. Q.G.: conceptualization, methodology, project administration, validation, writing—review and editing. All authors read and approved the final manuscript.

References

1. 2021 Alzheimer's disease facts and figures. Alzheimers Dement 2021;17: 10.1002/alz.12328
2. Lane CA, Hardy J, Schott JM. Alzheimer's disease. Eur J Neurol 2018;25: 10.1111/ene.13439
3. Realdon O, Rossetto F, Nalin M, et al. Technology-enhanced multi-domain at home continuum of care program with respect to usual care for people with cognitive impairment: the Ability-TelerehABILITation study protocol for a randomized controlled trial. BMC Psychiatry 2016;16: 10.1186/s12888-016-1132-y
4. Dubois B, Hampel H, Feldman HH, et al. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. Alzheimers Dement 2016;12: 10.1016/j.jalz.2016.02.002
5. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. J Neuropathol Exp Neurol 2012;71: 10.1097/NEN.0b013e31824b211b
6. McKhann G, Drachman D, Folstein M, et al. 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: 10.1212/wnl.34.7.939
7. Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. The Lancet Neurology 2007;6: 10.1016/s1474-4422(07)70178-3
8. Dubois B, Feldman HH, Jacova C, et al. Revising the definition of Alzheimer's disease: a new lexicon. The Lancet Neurology 2010;9: 10.1016/s1474-4422(10)70223-4
9. Dubois B, Feldman HH, Jacova C, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. The Lancet Neurology 2014;13: 10.1016/s1474-4422(14)70090-0
10. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement 2018;14: 10.1016/j.jalz.2018.02.018
11. Bouwman FH, Verwey NA, Klein M, et al. New research criteria for the diagnosis of Alzheimer's disease applied in a memory clinic population. Dement Geriatr Cogn Disord 2010;30: 10.1159/000315542
12. Wang HF, Tan L, Cao L, et al. Application of the IWG-2 Diagnostic Criteria for Alzheimer's Disease to the ADNI. J Alzheimers Dis 2016;5110.3233/JAD-150824
13. Kern S, Zetterberg H, Kern J, et al. Burnham SC, Coloma PM, Li QX, et al. Application of the NIA-AA Research Framework: Towards a Biological Definition of Alzheimer's Disease Using Cerebrospinal Fluid Biomarkers in the AIBL Study. J Prev Alzheimers Dis 2019;6: doi: 10.14283/jpad.2019.25
14. Adamczuk K, Schaeverbeke J, Vanderstichele HM, et al. Diagnostic value of cerebrospinal fluid Abeta ratios in preclinical Alzheimer's disease. Alzheimers Res Ther 2015;7: doi: 10.1186/s13195-015-0159-5
15. Chatterjee P, Pedrini S, Ashton NJ, et al. Diagnostic and prognostic plasma biomarkers for preclinical Alzheimer's disease. Alzheimers Dement 2022;18: doi: 10.1002/alz.12447
16. Zhou XH, Obuchowski NA, McClish DK (2011) Statistical Methods in Diagnostic Medicine, John Wiley & Sons,Inc.
17. Hale C, Last BS, Meier IB, et al. The ModRey: An Episodic Memory Test for Nonclinical and Preclinical Populations. Assessment 2019;26: 10.1177/1073191117723113
18. Petersen RC, Aisen P, Boeve BF, et al. Mild cognitive impairment due to Alzheimer disease in the community. Annals of neurology 2013;7410.1002/ana.23931
19. Jack CR, Jr., Knopman DS, Weigand SD, et al. An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease. Annals of neurology 2012;71: 10.1002/ana.22628
20. Fortea J, Vilaplana E, Alcolea D, et al. Cerebrospinal fluid β-amyloid and phospho-tau biomarker interactions affecting brain structure in preclinical Alzheimer disease. Annals of neurology 2014;76: 10.1002/ana.24186
21. Trojanowski JQ, Vandeerstichele H, Korecka M, et al. Update on the biomarker core of the Alzheimer's Disease Neuroimaging Initiative subjects. Alzheimers Dement 2010;6: 10.1016/j.jalz.2010.03.008
22. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Annals of neurology 2009;65: 10.1002/ana.21610
23. Prestia A, Caroli A, Wade SK, et al. Prediction of AD dementia by biomarkers following the NIA-AA and IWG diagnostic criteria in MCI patients from three European memory clinics. Alzheimers Dement 2015;11: 10.1016/j.jalz.2014.12.001
24. Lee J, Jang H, Kang SH, et al. Cerebrospinal Fluid Biomarkers for the Diagnosis and Classification of Alzheimer's Disease Spectrum. J Korean Med Sci 2020;35: doi: 10.3346/jkms.2020.35.e361
25. Blennow K, Hampel H, Weiner M, Zetterberg H. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat Rev Neurol 2010;6: doi: 10.1038/nrneurol.2010.4
26. Parnetti L, Chipi E, Salvadori N, D'Andrea K, Eusebi P. Prevalence and risk of progression of preclinical Alzheimer's disease stages: a systematic review and meta-analysis. Alzheimers Res Ther 2019;1110.1186/s13195-018-0459-7
27. World Alzheimer Report 2018 - The state of the art of dementia research: New frontiers[J]. NEW FRONTIERS
28. Priest PC, Jennings LC, Duncan AR, Brunton CR, Baker MG. Effectiveness of border screening for detecting influenza in arriving airline travelers. American journal of public health 2015;105: Suppl 410.2105/AJPH.2012.300761r
29. Jack CR, Jr., Therneau TM, Weigand SD, et al. Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging-Alzheimer's Association Research Framework. JAMA Neurol 201910.1001/jamaneurol.2019.1971

How to cite this article: X. Wang, G. Niu, J. Zhaoi, et al. Evaluating the Performance of Different Criteria in Diagnosing AD and Preclinical AD with the Bayesian Latent Class Model. J Prev Alz Dis 2024; http://dx.doi.org/10.14283/jpad.2024.71