Published in final edited form as:

J Alzheimers Dis. 2016 ; 51(2): 463–470. doi:10.3233/JAD-151045.

Evaluation of Cerebrospinal Fluid Assay Variability in Alzheimer's Disease

Matthew T. White¹, Leslie M. Shaw², and Sharon X. Xie^{3,*} for the Alzheimer's Disease Neuroimaging Initiative^{†,†} the National Alzheimer's Coordinating Center[†]

¹Clinical Research Center, Boston Children's Hospital, Boston, Massachusetts , U.S.A.

Department of Psychiatry, Harvard Medical School, Boston, MA, U.S.A

²Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, U.S.A

³Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, U.S.A

SUMMARY

Studies of cerebrospinal fluid (CSF) biomarkers in Alzheimer's disease (AD) have indicated that much of the variability observed in the biomarkers may be due to measurement error. Biomarkers are often obtained with measurement error, which may make the diagnostic biomarker appear less effective than it truly is. In the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, technical replicates of CSF biomarkers are available; the National Alzheimer's Coordinating Center database contains longitudinal replicates of CSF biomarkers. We focus on the area under the receiver operating characteristic curve (AUC) as the measure of diagnostic effectiveness for differentiating AD from normal cognition using CSF biomarkers and compare AUC estimates obtained by a more standard, naïve method (which uses a single observation per subject and ignores measurement error) to a maximum likelihood (ML) based method (which uses all replicates per subject and adjusts for measurement error). The choice of analysis method depends upon the noise to signal ratio (i.e., the magnitude of the measurement error variability relative to the true biomarker variability); moderate to high ratios may significantly bias the naïve AUC estimate, and the ML-based method would be preferred. The noise to signal ratios were low for the ADNI biomarkers but high for the tTau and pTau biomarkers in NACC. Correspondingly, the naïve and ML-based AUC estimates were nearly identical in the ADNI data but dissimilar for the tTau and pTau biomarkers in the NACC data. Therefore, using the naïve method is adequate for analysis of CSF biomarkers in the ADNI study, but the ML method is recommended for the NACC data.

*sxie@mail.med.upenn.edu

[†]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) and the National Alzheimer's Coordinating Center (NACC) database (www.alz.washington.edu). As such, the investigators within the ADNI and the NACC contributed to the design and implementation of ADNI/NACC 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/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

6. Conflict of Interest

The authors declare that they have no conflict of interest.

Keywords

Alzheimer's disease; Biomarkers; Diagnostic testing; Maximum likelihood; Measurement error; Replicate data

1. Introduction

Biomarker research in Alzheimer's disease (AD) has become increasingly important in recent years as biomarkers have been recognized in signaling the onset of the disease before the emergence of measurable clinical symptoms. The Alzheimer's Disease Neuroimaging Initiative (ADNI) and National Alzheimer's Coordinating Center (NACC) studies are designed to track the longitudinal changes in neuroimaging, biological markers, clinical, and neuropsychological measures in order to accelerate the development of new treatments. It has been known that measurements of the cerebrospinal fluid (CSF) concentrations of amyloid beta ($A\beta_{1-42}$), total tau (tTau), and phosphorylated tau (pTau) have high variability across different assays, laboratories, and time [1,2,3,4]. Shaw et al. [5] reported a seven-center inter-laboratory standardization study using ADNI participants for CSF $A\beta_{1-42}$, tTau, and pTau measures with a within-center percent coefficient of variation (%CV) ranging from 5.3% to 10.8% and inter-center %CV ranging from 13.1% to 17.9%.

In the ADNI study, each CSF sample was separated into several aliquots; at each study visit, a new CSF sample from each subject was obtained and analyzed, and one aliquot from each of the subject's previous visits was re-analyzed in the lab. Due to reagent changes in 2008, the raw CSF measures obtained after 2007 were rescaled to the levels of the 2007 values using regression analyses (see details on the ADNI website: <http://www.adni-info.org/>). Each subject therefore has replicate measures of $A\beta_{1-42}$, tTau, and pTau at each time point. Currently, researchers typically use one replicate of each biomarker (e.g., the initial replicate or a randomly selected replicate) in data analyses. In the NACC study, CSF samples are obtained from each subject at approximately yearly intervals. Researchers typically use one replicate of each biomarker (e.g., the initial replicate or most recent replicate) in data analyses.

When using highly variable covariates to predict cognitive rate of decline, dementia conversion, and measures of diagnostic accuracy such as the area under the receiver operating characteristic curve (AUC), statistical estimates can be biased if the high variability of the covariate (e.g., CSF biomarkers) is ignored [6,7,9,10,11,12]. In this situation, we need more sophisticated statistical approaches that use all replicates of a covariate rather than just one replicate in order to correctly estimate the covariate effect [6,7,9,10,11,13,14,15]. If the variability of the covariate is low, the statistical estimates have little bias, which can effectively be ignored in practice; in this case, using a single replicate in the analysis is allowable.

In this paper, we focus specifically on the impact of variability (i.e., variability among technical replicates due to imperfect lab conditions in the ADNI study or temporal/laboratory variability among longitudinal replicates in the NACC study) on estimates of diagnostic performance and aim to evaluate whether using one replicate of the CSF value

instead of using all replicates is sufficient to perform unbiased statistical analysis. We examine the ability of baseline CSF A β ₁₋₄₂, tTau, and pTau to differentiate AD from cognitively normal subjects and compare the AUC estimates derived using a single replicate to those derived using all replicates of the CSF measures.

2. Methods

2.1 ADNI study

Data used in the preparation of this article were obtained from the ADNI database (<http://adni.loni.usc.edu>). The ADNI study was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as to lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date, these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see <http://www.adni-info.org/>.

CSF samples were collected from all consenting subjects. All biomarker samples were collected in the morning before breakfast and after an overnight fast. Only water was permitted until blood draws and the lumbar puncture were completed. The lumbar puncture was performed with a 20- or 24-gauge spinal needle as described in the ADNI procedures manual (<http://www.adni-info.org/>). In brief, CSF was collected into tubes provided to each site, transferred into polypropylene tubes and frozen on dry ice within 1 hour after collection, and shipped overnight to the ADNI Biomarker Core laboratory at the University of Pennsylvania Medical Center on dry ice. Aliquots (0.5ml) were prepared from these samples after thawing at room temperature for 1 hour and gentle mixing. The aliquots were then stored in barcode labeled polypropylene vials at -80°C. Written informed consent was obtained for participation in these studies, and institutional review board approval was obtained at each participating center. Further details are described by Shaw et al. [5,16].

2.2 NACC study

Additional data used in the preparation of this article were obtained from the NACC database (<http://www.alz.washington.edu>). The NACC was established in 1999 and maintains a cumulative database including clinical evaluations, neuropathology data (when available), MRI imaging, and CSF biomarkers. Clinical evaluation and CSF biomarker data used in this manuscript were obtained from a single Alzheimer's Disease Center (ADC) as there is currently one ADC which has contributed CSF biomarker results to the NACC database. Data were obtained from visits between September 2005 and May 2015.

For the longitudinal NACC data, it is possible for participants' AD severity to worsen over time. To help meet the assumption of temporal stability within subject (needed for the statistical model in Section 2.3), only observations within 15 months of the participant's initial study visit (up to 2 months prior and up to 13 months after the initial clinic visit) were included. If subjects had only one visit or transitioned in diagnosis between the initial visit and their second (1 year) visit, then only CSF observations within 2 months before and 2 months after the initial clinic visit were included in the analyses. These ranges were chosen for clinical and practical reasons, namely that the disease severity should remain relatively stable in the given time frame and that the yearly study visit and CSF sample collection were not performed at the same visit.

2.3 Statistical analysis

The ADNI and NACC samples consist of 377 and 142 subjects, respectively, who were diagnosed as having AD or normal cognition at baseline and who had at least one tTau, pTau, or $A\beta_{1-42}$ biomarker measurement. All analyses were conducted separately for the ADNI and NACC samples. Baseline demographic and clinical characteristics of the samples were summarized overall and by disease status (AD versus cognitively normal). Comparisons between disease groups were made using Fisher's exact test (for categorical variables) and the Wilcoxon rank-sum test (for continuous variables).

We assume the biomarkers follow a classical measurement error model, in which each observed value of the biomarker is equal to the subject's true biomarker value plus some error (see [18] for full model specifications and assumptions). This model assumes that the errors are additive, independent, and normally distributed. In the ADNI study, biomarkers were found to be positively skewed and were log transformed to meet the normality assumption. In the NACC study, the biomarkers exhibited different degrees of skewness; the log of tTau, inverse of pTau, and square root of $A\beta_{1-42}$ were used for analysis. For each biomarker, the additive error assumption was assessed in each group by plotting the intra-individual standard deviation against the mean [19]. The plots (not shown) did not show any structure, indicating that the transformation was reasonable and that the additive error assumption was appropriate. For each biomarker, analyses of model residuals in each group were used to assess the joint normality and independence assumptions for the error terms. Normal quantile-quantile plots of the residuals (not shown) were approximately linear, and Mardia's tests for multivariate skewness and kurtosis [20] were not significant except for pTau in the cognitively normal group in the ADNI data. Since the model residuals are correlated by definition, Fitzmaurice, Laird, and Ware [21] recommend using a semi-

variogram to assess the adequacy of the selected model for the covariance structure. The semi-variograms for the ADNI data (not shown) were centered at 1 and showed no systematic pattern, indicating that the assumed covariance structure (i.e., uncorrelated errors) was appropriate. This suggests that the normality assumption is appropriate after transformation. It is not possible to create semi-variograms for data with 2 or fewer observations per subject; in the selected NACC data, fewer than 5% of participants had more than 2 longitudinal observations. It was therefore not possible to use semi-variograms to assess the covariance structure for the NACC data.

For normally distributed biomarkers, the AUC A can be written as a function of the population mean and variance from each disease group; see [14] for details on the estimation of A and its confidence interval. In a more standard (naïve) method, we select one replicate per subject and use the sample mean and variance of each group in the AUC estimation, which has been used by previous researchers who analyzed biomarker data from the ADNI study. This method ignores measurement error and could potentially lead to substantial bias in the estimation of A . In a more advanced method that uses all replicates, we use a maximum likelihood (ML) approach to estimate the population mean and variance as well as the measurement error variance in each group; see page 86 of [17] and [18] for additional information. This ML approach allows us to separate the true between-subject variability from the measurement error variability, which leads to unbiased estimation of A .

In our analysis, we utilize the repeated baseline CSF measures from the ADNI study and the longitudinal CSF measures from the NACC study for comparison of the naïve and ML methods; a large discrepancy in the estimated AUCs would indicate that the noise to signal ratio (i.e., the ratio of the measurement error variance to the true biomarker variance) is too high, motivating the need to utilize the more sophisticated ML approach in the analysis. We assess the naïve method under conditions where the first, last, or a randomly-selected observation from each subject is used for analysis, since these are commonly done to reduce the complexity of biomarker data.

3. Results

Demographic and clinical characteristics of the study subjects are presented in Table 1. In the ADNI study, the cognitively normal group had significantly higher proportions of female subjects, subjects with education beyond high school, and non-White subjects than the AD group; no differences were statistically significant in the NACC study. There were no differences in ethnicity between groups in either study. In both studies, the cognitively normal group scored significantly higher on the Mini-mental state examination than the AD group. Among the biomarkers in each study, the cognitively normal group had significantly higher levels of $A\beta_{1-42}$ and significantly lower levels of tTau and pTau than the AD group.

Table 2 presents the distribution of the number of observations per subject for each of the CSF biomarkers. In ADNI, we see that about 40% of subjects have 1 observation for $A\beta_{1-42}$ and tTau while almost 80% of subjects have 1 observation for pTau; in NACC, about 70% of subjects overall have 1 observation for each biomarker. Each biomarker, however, has

adequate replicate data for the maximum likelihood-based estimation of the biomarker mean and variance as well as the measurement error variance for each group.

Table 3 presents the AUC estimate and 95% confidence interval for differentiating AD from cognitively normal subjects, calculated under the naïve and ML approaches for each of the CSF biomarkers. We evaluated the naïve approach under scenarios where the first, last, or a randomly selected observation was used for analysis, since these scenarios are commonly seen in practice. In ADNI, the average noise to signal ratios across the AD and cognitively normal subjects were low: 0.10, 0.05, and 0.09 for the $A\beta_{1-42}$, tTau, and pTau biomarkers, respectively. We see that there is little difference among the analytical approaches, indicating that using one replicate per subject is adequate to assess the diagnostic performance of CSF biomarkers. In the NACC data, the average noise to signal ratio across the AD and cognitively normal subjects was low for the $A\beta_{1-42}$ biomarker but high for the tTau and pTau biomarkers: 0.36, 2.61, and 2.36, respectively. As expected, the naïve and ML approaches yielded similar results for the $A\beta_{1-42}$ biomarker but different results for tTau and pTau. Measurement error has been shown to negatively bias the naïve AUC estimate [9]; the large noise to signal ratios for tTau and pTau may explain the large differences between the naïve and ML AUC estimates. The confidence intervals for the ML AUC estimates are large, however, since there are relatively few subjects with replicate data in the NACC study.

As a sensitivity analysis within the ADNI study, we repeated our analysis for the $A\beta_{1-42}$ and tTau biomarkers on all subjects with 2 observations ($N = 160$ and 142 , respectively; see Table 2). The results (data not shown) were consistent with those presented above.

4. Discussion

In this paper, we aimed to determine whether a standard (naïve) approach using one replicate per subject in the analysis of diagnostic biomarkers from the ADNI and NACC studies were adequate or if a more sophisticated ML approach incorporating all replicates for each subject was needed. When measurement error variability is low, the two methods yield very similar results, so the standard approach is acceptable; however, when measurement error variability is high, the two methods will not agree and the ML approach is preferred. We found that the two methods produced nearly identical results in ADNI, supporting the use of a single replicate in the ADNI data analysis. The results are similar due to the small amount of noise (measurement error variability) observed in the biomarkers, controlled in part by the rescaling procedure described in Section 1. In NACC, however, there was a high degree of noise relative to the estimated “true” variability in the tTau and pTau biomarkers, which lead to differing estimates of the naïve- and ML-based AUC.

There are two potential limitations in using the rescaling procedure used in ADNI, however. The first issue is that the rescaling procedure induces correlation between the 2007 CSF value and subsequent CSF values obtained after 2008. Extensions to the current maximum likelihood approach to account for correlated measurement errors are needed. The second issue with the rescaling procedure is that the accuracy of the rescaled values depends upon the quality of the regression model. If the regression model fits poorly, the rescaled values will be incorrect, potentially biasing the estimation of the population parameters and thus the

AUC. The regression models used in the ADNI study all fit the data well, with an R^2 of about 0.9 in each model. For cases where rescaling is needed but the model does not fit well, the investigator may consider using a non-linear regression model to better capture the relationship among the observations. Note, however, that any transformation applied to the data should be a monotone, one-to-one function and that the same transformation should be applied to both the diseased and non-diseased groups.

We recommend that researchers evaluate the noise to signal ratio for each CSF biomarker when choosing an analysis approach. Faraggi [8] evaluated confidence intervals around naïve AUC estimates when the noise to signal ratio ranged between 0.5 and 4.0 and found that even for a ratio of 0.5, the confidence interval did not contain the true AUC value in an adequate proportion of simulated datasets. We therefore do not recommend analyzing a single replicate when the noise to signal ratio is 0.5 or above since the naïve AUC estimate may be negatively biased in this case. Our current study suggests that the rigorous protocol for assay analysis in the ADNI study may help to control the error variability of the biomarkers, allowing the use of a single replicate for analysis. We currently recommend using the maximum likelihood approach for the NACC data. However, additional information is needed to confirm this conclusion. When CSF data from additional ADCs are available, we will repeat this analysis to see if our conclusions hold.

We encourage investigators to incorporate aspects of the ADNI protocol in new biomarker studies, where appropriate, in order to minimize the impact of measurement error on estimates of biomarker performance. If measurement error is still an issue after implementing these procedures, however, there are methods available (see, for example, [14] and [18]) to statistically adjust for measurement error in the analysis. In summary, investigators should strive to achieve better control over measurement error, which will reduce the number of replicates needed per subject in the study design, reducing patient burden and lowering study costs.

Acknowledgments

The authors thank the three anonymous reviewers for their helpful comments which greatly strengthened the manuscript.

This work is supported by NIH grants including AG-10124 (the Alzheimer's Disease Core Center), AG-32953, AG-17586, and NS-053488 (the Morris K. Udall Parkinson's Disease Research Center of Excellence).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Steven Ferris, PhD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG016570 (PI David Teplow, PhD), P50 AG005131 (PI Douglas Galasko, MD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P50 AG005136 (PI Thomas Montine, MD, PhD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), and P50 AG005681 (PI John Morris, MD).

References

1. Dean RA, Shaw LM. Use of cerebrospinal fluid biomarkers for diagnosis of incipient Alzheimer disease in patients with mild cognitive impairment. *Clin Chem*. 2010; 56:7–9. [PubMed: 19926774]
2. Lewczuk P, Beck G, Ganslandt O, Esselmann H, Deisenhamer F, Regeniter A, Petereit HF, Tumani H, Gerritzen A, Oschmann P, Schroder J, Schonknecht P, Zimmermann K, Hampel H, Burger K, Otto M, Hausteiner S, Herzog K, Dannenberg R, Wurster U, Bibl M, Maler JM, Reubach u, Kornhuber J, Wiltfang J. International quality control survey of neurochemical dementia diagnosis. *Neurosci Lett*. 2006; 409:1–4. [PubMed: 17045397]
3. Lewczuk P, Beck G, Esselmann H, Bruckmoser R, Zimmerman R, Fiszer M, Bibl M, Maler JM, Kornhuber J, Wiltfang J. Effect of sample collection tubes on cerebrospinal fluid concentrations of tau proteins and amyloid beta peptides. *Clin Chem*. 2006; 52:332–334. [PubMed: 16449222]
4. Verwey NA, van der Flier WM, Blennow K, Clark C, Sokolow S, De Deyn PP, Galasko D, Hampel H, Hartmann T, Kapki E, Lannfelt L, Mehta PD, Parnetti L, Petzold A, Pirttila T, Saleh L, Skinningsrud A, Swieten JC, Verbeek MM, Wiltfang J, Younkin S, Scheltens P, Blankenstein MA. A worldwide multicentre comparison of assays for cerebrospinal fluid biomarkers in Alzheimer's disease. *Ann clin Biochem*. 2009; 46:235–240. [PubMed: 19342441]
5. Shaw LM, Vanderstichele H, Knapik-Czajka M, Figurski M, Coart E, Blennow K, Soares H, Simon AJ, Lewczuk P, Dean RA, Siemers E, Potter W, Lee VM, Trojanowski JQ. for the Alzheimer's Disease Neuroimaging Initiative. Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. *Acta Neuropathol*. 2011; 121:597–609. [PubMed: 21311900]
6. Coffin M, Sukhatme S. Receiver operating characteristic studies and measurement errors. *Biometrics*. 1997; 53:823–837. [PubMed: 9333348]
7. Coffin, M.; Sukhatme, S. A parametric approach to measurement errors in receiver operating characteristic studies. In: Jewell, NP.; Kimber, AC.; Ting Lee, ML.; Whitmore, GA., editors. *Lifetime Data: Models in Reliability and Survival Analysis*. Kluwer; Boston: 1995.
8. Faraggi D. The effect of random measurement error on receiver operating characteristic (ROC) curves. *Stat Med*. 2000; 19:61–70. [PubMed: 10623913]
9. Schisterman EF, Faraggi D, Reiser B, Trevisan M. Statistical inference for the area under the receiver operating characteristic curve in the presence of random measurement error. *Am J Epidemiol*. 2001; 154:174–179. [PubMed: 11447052]
10. Perkins NJ, Schisterman EF. The Youden index and optimal cut-point corrected for measurement error. *Biom J*. 2005; 47:428–441. [PubMed: 16161802]
11. Perkins NJ, Schisterman EF, Vexler A. Generalized ROC curve inference for a biomarker subject to a limit of detection and measurement error. *Stat Med*. 2009; 28:1841–1860. [PubMed: 19340817]
12. Long Q, Flanders WD, Fedirko V, Bostick RM. Robust statistical methods for analysis of biomarkers measured with batch/experiment-specific errors. *Stat Med*. 2010; 29:361–370. [PubMed: 20020422]
13. Li Y, Koval JJ, Donner A, Zuo GY. Interval estimation for the area under the receiver operating characteristic curve when data are subject to error. *Stat Med*. 2010; 29:2521–2531. [PubMed: 20684006]
14. Reiser B. Measuring the effectiveness of diagnostic markers in the presence of measurement error through the use of ROC curves. *Stat Med*. 2000; 19:2115–2129. [PubMed: 10931515]

15. Tosteson TD, Buonaccorsi JP, Demidenko E, Wells WA. Measurement error and confidence intervals for ROC curves. *Biom J.* 2005; 47:409–416. [PubMed: 16161800]
16. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VM, Trojanowski JQ. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol.* 2009; 65:403–413. [PubMed: 19296504]
17. Searle, SR.; Casella, G.; McCulloch, CE. *Variance Components.* John Wiley & Sons, Inc; New Jersey: 1992.
18. White MT, Xie SX. Adjustment for measurement error in evaluating diagnostic biomarkers by using an internal reliability sample. *Stat Med.* 2013; 32:4709–4725. [PubMed: 23765915]
19. Carroll, RJ.; Ruppert, D.; Stefanski, LA. *Measurement Error in Nonlinear Models.* Chapman & Hall/CRC; Boca Raton: 1995.
20. Mardia KV. Measures of multivariate skewness and kurtosis with applications. *Biometrika.* 1970; 57:519–530.
21. Fitzmaurice, GM.; Laird, NM.; Ware, JH. *Applied Longitudinal Analysis.* John Wiley & Sons, Inc; New Jersey: 2004.

Table 1

Demographic and clinical characteristics of the study subjects. Note: CN = Cognitively normal, AD = Alzheimer's disease, MMSE = Mini-Mental State Examination, IQR = Interquartile range. P-values are calculated separately for each study and are based on Fisher's exact test (for categorical variables) or the Wilcoxon rank-sum test (for continuous variables).

Characteristic	Alzheimer's Disease Neuroimaging Initiative				National Alzheimer's Coordinating Center			
	Total (N = 377)	CN (N = 189)	AD (N = 188)	p-value	Total (N = 142)	CN (N = 95)	AD (N = 47)	p-value
Gender, n (%)				0.0233				0.5912
Male	198 (52.5)	88 (46.6)	110 (58.5)		79 (55.6)	51 (53.7)	28 (59.6)	
Female	179 (47.5)	101 (53.4)	78 (41.5)		63 (44.4)	44 (46.3)	19 (40.4)	
Education, n (%)				< 0.0001				0.1232
High School	57 (15.1)	17 (9.0)	40 (21.3)		13 (9.2)	6 (6.3)	7 (14.9)	
> High School	320 (84.9)	172 (91.0)	148 (78.7)		129 (90.0)	89 (93.7)	40 (85.1)	
Race, n (%)				0.0090				0.8202
White	345 (91.5)	165 (87.3)	180 (95.7)		131 (92.3)	88 (92.6)	43 (91.5)	
Black	23 (6.1)	18 (9.5)	5 (2.7)		1 (0.7)	1 (1.1)	0	
Other	9 (2.4)	6 (3.2)	3 (1.6)		10 (7.0)	6 (6.3)	4 (8.5)	
Ethnicity				0.1714				> 0.9999
Hispanic or Latino	14 (3.7)	10 (5.3)	4 (2.1)		3 (2.1)	2 (2.1)	1 (2.1)	
Not Hispanic or Latino	362 (96.3)	179 (94.7)	183 (97.9)		138 (97.9)	92 (97.9)	46 (97.9)	
Unknown	0	0	0		1	1	0	
MMSE, Median (IQR)	26 (24, 29)	29 (29, 30)	24 (22, 25)	< 0.0001	29 (24, 30)	29 (29, 30)	20 (17, 24)	< 0.0001
Aβ ₁₋₄₂ , Median (IQR)	153.2 (128.9, 223.0)	216.0 (161.3, 247.8)	133.6 (115.9, 151.6)	< 0.0001	266.5 (183.7, 355.9)	313.7 (228.4, 425.5)	185.2 (148.2, 239.6)	< 0.0001
tTau, Median (IQR)	81.5 (54.0, 121.7)	58.0 (45.0, 83.1)	113.8 (82.0, 158.0)	< 0.0001	53.8 (42.6, 68.5)	46.3 (39.4, 56.6)	68.5 (57.9, 87.1)	< 0.0001
pTau, Median (IQR)	33.0 (21.0, 47.9)	22.0 (17.0, 35.0)	41.0 (32.0, 58.0)	< 0.0001	35.1 (27.1, 46.8)	30.5 (25.1, 36.7)	49.5 (40.1, 62.1)	< 0.0001

Table 2

Distribution of the number of observations per subject for each of the CSF biomarkers within the Alzheimer's Disease Neuroimaging Initiative (ADNI) and National Alzheimer's Coordinating Center (NACC) studies. Note: CN = Cognitively normal, AD = Alzheimer's disease. In the ADNI study, all subjects had at least one observation for Aβ₁₋₄₂; 7 subjects were missing tTau, and 1 subject was missing pTau.

Study	Group	Marker	Observations per Subject				
			1	2	3	4	5
ADNI	CN	Aβ ₁₋₄₂	74	56	19	23	17
		tTau	72	56	20	23	16
		pTau	129	19	23	17	0
	AD	Aβ ₁₋₄₂	86	84	3	15	0
		tTau	81	86	2	14	0
NACC	CN	pTau	170	4	14	0	0
		Aβ ₁₋₄₂	73	22	0	0	0
		tTau	73	22	0	0	0
		pTau	73	22	0	0	0
		Aβ ₁₋₄₂	25	16	6	0	0
	AD	tTau	25	16	6	0	0
		pTau	25	16	6	0	0

Estimate and 95% confidence interval for the AUC of each biomarker for differentiating AD from cognitively normal subjects, calculated under the naïve and maximum likelihood (ML) approaches. Note: Results are presented separately for the Alzheimer's Disease Neuroimaging Initiative (ADNI) and National Alzheimer's Coordinating Center (NACC) studies. The naïve approach is based on a single observation from each subject; the table presents results where the first, last, or a randomly-selected observation is used. The ML approach incorporates all replicate observations in the AUC estimation.

Table 3

Study	Marker	Naïve Approach			ML Approach
		First Obs.	Last Obs.	Random Obs.	
ADNI	Aβ ₁₋₄₂	0.84 (0.80, 0.88)	0.86 (0.82, 0.89)	0.84 (0.80, 0.88)	0.86 (0.82, 0.90)
	τ _{Tau}	0.83 (0.78, 0.87)	0.82 (0.78, 0.86)	0.83 (0.78, 0.87)	0.83 (0.79, 0.87)
	p _{Tau}	0.78 (0.73, 0.82)	0.78 (0.73, 0.82)	0.78 (0.73, 0.82)	0.79 (0.74, 0.84)
NACC	Aβ ₁₋₄₂	0.78 (0.69, 0.85)	0.79 (0.71, 0.85)	0.78 (0.69, 0.85)	0.82 (0.73, 0.89)
	τ _{Tau}	0.80 (0.72, 0.86)	0.82 (0.74, 0.88)	0.80 (0.72, 0.86)	0.94 (0.78, 0.99)
	p _{Tau}	0.85 (0.77, 0.91)	0.85 (0.77, 0.91)	0.86 (0.79, 0.92)	0.97 (0.82, 1.00)