



Will biomarker-based diagnosis of Alzheimer's disease maximize scientific progress? Evaluating proposed diagnostic criteria

Medellena Maria Glymour¹ · Adam Mark Brickman² · Mika Kivimaki³ · Elizabeth Rose Mayeda^{1,4} · Geneviève Chêne^{5,6} · Carole Dufouil^{5,6} · Jennifer Jaie Manly²

Received: 13 March 2018 / Accepted: 30 May 2018 / Published online: 9 June 2018
© Springer Nature B.V. 2018

Abstract

A recently published framework for the diagnosis of Alzheimer's disease (AD) in research studies would allow diagnosis on the sole basis of two biomarkers (β -amyloid and pathologic tau), even in people with no objective or subjective memory or cognitive changes. This revision will have substantial implications for future Alzheimer's research, and the changes should be rigorously evaluated before widespread adoption. We propose three principles for evaluating any revision to diagnostic frameworks for AD: (1) does the revision improve the validity of the diagnosis; (2) does the revision improve the reliability or reduce the expense of the diagnosis; and (3) will the revision foster innovative and rigorous research across populations. The new diagnostic framework is unlikely to achieve any of these goals. Instead, it has the potential to handicap future researchers, and slow progress towards identifying effective strategies to prevent or treat AD.

Keywords Alzheimer's disease · Diagnostic criteria · Dementia · Validity · Reliability

A recently published research framework for Alzheimer's disease (AD) proposes that AD can be diagnosed on the sole basis of two biomarkers: β -amyloid and pathologic tau [1]. Such guidance incorporates an important departure from previous criteria for AD in that living individuals with

high amyloid and tau burden would be diagnosed as having AD even if they had no objective or subjective memory or cognitive difficulties. Clinical AD is a multi-factorial syndrome, and the new proposal aspires to untangle the Gordian knot by focusing on a specific neuropathologic process putatively defined as AD.

We argue here that diagnosing AD based on two biomarkers alone, ignoring subjective and objective cognitive assessments, is a mistake until we are certain that these biomarkers are the central causal mechanisms for symptomatic AD. Diagnostic criteria are rules for measuring the presence or absence of a disease, and the standards we typically apply for evaluating measurements are relevant to diagnostic criteria. Measurement innovations can powerfully accelerate research but must be weighed against potential disadvantages. We argue that any changes in a research framework for diagnosing AD should be evaluated on three criteria (Box 1) and adopted only if they: (1) improve the validity of the diagnosis; (2) improve reliability or reduce cost, thereby increasing statistical power achievable in new studies; and (3) foster innovative, rigorous research, by reducing the potential for bias and promoting scientific discoveries. The new research framework as currently proposed is unlikely to fulfill any of these

✉ Medellena Maria Glymour
mglymour@epi.ucsf.edu

¹ Department of Epidemiology and Biostatistics, University of California, San Francisco, 550 16th Street, San Francisco, CA 94158, USA

² Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY, USA

³ Department of Epidemiology and Public Health, University College London, London, UK

⁴ Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, USA

⁵ INSERM U1219-Bordeaux Population Health and CIC1401-EC, ISPED-Bordeaux School of Public Health, INSERM, University of Bordeaux, Bordeaux, France

⁶ CHU de Bordeaux, Pôle de Santé Publique, Bordeaux, France

Box 1 Criteria for evaluating the new research framework for Alzheimer's disease

1. Improve validity of measurement.
 - a. Reduce correlation with disease processes known to be independent of AD.
 - b. Improve prediction of the current gold standard for AD, which includes cognitive decline.
2. Improve achievable statistical power in future studies, either by improving reliability or reducing cost.
 - a. Any increases in cost of new measures should be offset by improved reliability of measures.
 - b. Variance in the new measures should be primarily driven by changes in AD, not by other sources of variation, such as other diseases, lab differences in test implementation, or random fluctuations.
3. Facilitate more rigorous and innovative scientific studies.
 - a. Reduce selection bias due to recruitment or attrition differences.
 - b. Reduce confounding by common causes of AD risk factors and clinical AD.
 - c. Clarify scientific concepts to promote communication between researchers.
 - d. Make research feasible for diverse populations.
 - e. Make research feasible in novel settings.

criteria. On the contrary, adopting this framework will be a setback for the field, muddling current AD research and chilling future scientific discovery.

Validity of current and newly proposed diagnostic guidelines

Validity is defined as the extent to which a measurement assesses the construct of interest, as opposed to other, potentially correlated constructs. What is the target of interest in AD research? Conceptually, we can separate the biological changes in the brain from the clinical consequences of those changes such as cognitive and functional deterioration. It is the cognitive and functional outcomes that distress patients and families and pose major health care challenges. From a public health and clinical decision-making perspective, the clinical syndrome of AD is the relevant outcome.

One argument for adopting a biomarker-based AD research framework is that the clinical syndrome of AD is influenced by multiple pathologies and using a biomarker-based criterion may help us focus on a limited set of physiologic processes. A narrower set of physiologic processes indicated by the selected biomarkers might be easier to understand, interrupt, or reverse compared to the complex, intersecting processes contributing to clinical AD. This reasoning only advances science if we have comprehensive biomarkers for the relevant neuropathological process and we have demonstrated that process is necessary and sufficient for the eventual development of clinical AD. Unfortunately, we cannot say these conditions are met with respect to AD. The amyloid cascade hypothesis remains unverified as illustrated by several pharmacological treatments that show reductions in amyloid burden but no

improvements on clinical manifestations [2]. Amyloid is characteristic of patients with clinical AD and correlates with future cognitive deterioration, but many people (about 30%) have substantial amyloid burden and no detectable cognitive consequences. Most people who are biomarker positive will never develop the clinical disease [3]. Conversely, many people (about 25%) who meet the clinical diagnostic criteria for AD have no or limited amyloid burden [4]. The limited available evidence suggests the correspondence of these biomarkers with clinical and cognitive outcomes may differ by age and race, perhaps due to the differential importance of vascular disease [5–8].

Additionally, we do not know if amyloid is the initial etiologic insult that leads to AD, or merely a biomarker of another pathologic process. It is critical to understand this before designating these biomarkers as sufficient to define the disease. For many diseases, there are strongly predictive biomarkers that are not biological mechanisms. For example, elevated C-reactive protein (CRP) is a strong marker of coronary heart disease risk, but it is not a biological mechanism [9]. Until we understand the essential biological mechanisms of AD, we cannot be sure that amyloid and tau are valid measures of the most relevant biological processes leading to clinical AD. The important new approaches developed to measure amyloid appear to be valid measures of the presence of amyloid in the brain, but they are *not* yet proven to be valid measures of the clinical syndrome of AD. The best evidence to date suggests that clinical AD culminates when multiple pathways converge [10–12].

The importance of amyloid in the original case description of Alois Alzheimer is sometimes invoked to justify using amyloid imaging as a gold standard in contemporary diagnoses. This argument is specious because

Table 1 Ratio of power in hypothetical studies using amyloid imaging versus cognitive assessments, considering tradeoffs in affordable sample size versus reliability under alternative scenarios

Ratio of cost of cognitive assessment to cost of amyloid imaging	Percent of variance in cognitive assessments due to AD					
	30%			50%		
	Percent of variance in amyloid burden due to AD					
	50%	70%	90%	50%	70%	90%
1	1.46	1.74	1.79	1.00	1.18	1.27
5	0.45	0.62	0.74	0.30	0.41	0.51
10	0.26	0.35	0.45	0.18	0.24	0.32
20	0.18	0.24	0.24	0.13	0.17	0.18
40	0.16	0.20	0.20	0.10	0.13	0.14

Numbers above 1 indicate more would be learned for the same financial resources using amyloid imaging. Calculations assume enough resources for 500 individuals in the cognition study, an effect size of 0.25 standard deviations, and an alpha criterion of .05. For reference, power to detect an effect with the cognitive outcome under these circumstances would be 54% if 30% of the variance in cognition was due to AD pathologic processes; the power would be 78% if 50% of the variance was due to AD. The variance explained by AD pathology may be less than 100% for either cognitive assessments or amyloid burden due to unmeasured causes (e.g., Parkinson's disease) or measurement error (e.g., due to differences in interviewer, machine, software, quality control rules). As an example of interpretation of the estimates in the table, if 30% of the variance in cognitive assessments were due to the underlying AD pathologic process, while 90% of the variance in amyloid burden were due to this AD pathology, but amyloid imaging costs 20 times as much as cognitive assessments, the power of an amyloid based study would be only 24% as high as the power of a study using only cognitive assessments

the first patient identified with AD—a woman of 51 years—was brought to Alzheimer's attention because she was experiencing severe cognitive impairment and further was found to have substantial co-occurring neuropathological changes [13].

Improving efficiency of research

New diagnostic approaches often provide breakthroughs because they improve the efficiency of research and help us make faster scientific progress. For example, a new diagnostic approach that made identification of AD cases less expensive or less burdensome would allow us to enroll larger sample sizes and achieve more precise effect estimates. Alternatively, increasing measurement accuracy would allow us to learn more from smaller samples, for example allowing for more targeted recruitment in invasive clinical trials. Reliability is the extent to which variation in an instrument reflects variation in the construct of interest. More reliable outcome assessments provide more precise effect estimates with the same sample size. Increasing reliability improves statistical efficiency, but that advantage is eroded if the new measure is more expensive or burdensome.

We can calculate the net impact of proposed new measures with a few assumptions about the relative cost versus relative reliability (i.e., percent of variance in measurement that is due to a hypothetical pathologic process defining AD) of alternative measures. Consider a

study of whether an exposure, for example, physical activity, reduces risk of AD. Given fixed resources to conduct the study, are we better off using neuropsychological assessments or using amyloid imaging to assess whether the exposure influences the pathologic process of clinical AD? Which approach will maximize power? In Table 1, we show the net impact on power under alternative assumptions. In most plausible scenarios with current technology, power is much worse with imaging measures, and adopting these measures may increase risk of missing important causes of clinical AD. The same calculations can be applied to evaluate alternative biomarkers, and the ratios may improve with technological innovations to reduce cost.

Fostering rigorous and innovative scientific research

A serious potential negative aspect of the proposed research framework is that it may narrow our scientific vision, instead of expanding it. In part due to numerous failed trials, interest in investing research on β -amyloid and pathologic tau as therapeutic targets are diminishing in the pharmaceutical industry. Measuring proposed AD biomarkers is burdensome and expensive, cannot be done in the home of research participants or in most clinical settings, and may be perceived skeptically by many potential research participants, particularly those within populations where the benefit of early diagnosis would be

of the greatest advantage. The result of adopting this research framework will be even fewer study participants from communities already underrepresented in research, including racial/ethnic minorities, low socioeconomic status individuals, and people in rural or medically underserved communities. These categories include the majority of US residents and the vast majority of all living humans. In other words, the proposed framework will work best for a small slice of white, highly educated, people in middle- and high-income countries who live within close proximity to a major research university. By restricting the diversity of research participants, we also restrict the types of risk factors we can evaluate, and scientifically, this limits the generalizability and relevance of research results. Given this, combined with the high unit cost for biomarker diagnosis, search for upstream risk factors in the general population will become increasingly difficult. We will not be able to assess, for example, whether AD is influenced by many geographic or environmental exposures or risk factors that vary primarily across demographic groups.

While AD research is already highly selective, the 2-biomarker framework would exacerbate this problem. The strong selection may also weaken the rigor of studies and threaten internal validity. In observational studies, when treatment cannot be randomized, selection bias can lead to spurious correlations [14]. To illustrate how this could happen, consider the possibility that people with a family history of AD may be exceptionally motivated to participate in AD research. Such individuals may be willing to drive a long distance to the clinic and undergo uncomfortable or invasive procedures to participate. Subtle memory changes may make the person even more motivated to participate in research. In contrast, people with no family history of AD may ignore early feelings of memory decline because the possibility of developing AD is less salient to them. Such individuals may differentially decline study participation. This phenomenon may bias effect estimates or create entirely spurious associations between familial risk factors (e.g., genetics) and AD incidence [15]. As the barriers to participation grow, the selection bias introduced by such phenomena is also likely to grow.

We acknowledge that the current diagnostic criteria for clinical AD have important challenges, for example arising from the need to disentangle developmental from neurodegenerative processes. Early identification of AD using the current clinical criteria could be influenced by reserve, such that for two individuals with identical neuropathology meeting a specific biomarker diagnosis for AD, only one may meet clinical diagnostic criteria (or they may meet clinical criteria at very different ages) because of different early life experiences and cognitive development. To design targeted and effective interventions, however, it may be important to distinguish determinants of cognitive

development from determinants of neurodegeneration, even if both processes influence clinical cognitive impairment in late life. This is an important motivation for revisiting current criteria and should influence statistical analyses in AD research [16]. Unfortunately, the recently published biomarker-only criterion, based on an incomplete biological understanding of disease, does not solve this challenge.

An additional rationale offered for the new framework is the promise of identifying cases earlier in progression, before cognitive symptoms are manifest [17]. There is hope that such early identification will improve the potential for interventions to prevent disease development and identify a more appropriate population for trial enrollment. This is a specious argument because there is no need to redefine the disease in order to allow trialists to preferentially enroll people with heavy amyloid or tau burden. Further, future innovations in treatment strategies may not need to target amyloid positive individuals. More fundamentally, there is insufficient evidence that amyloid is the earliest detectable physiologic change foreshadowing incident clinical AD. In vivo amyloid assessments have not been available long enough to demonstrate that they show changes earlier than more easily assessed markers, such as cognition or even non-specific physiologic changes such as declines in body mass index (BMI) [18, 19]. In the Dominantly Inherited Alzheimer Network (DIAN), which simulated longitudinal data by using age of anticipated symptom onset, trajectories of logical memory for carriers began to deteriorate relative to trajectories of non-carriers at nearly an identical period as CSF $a\beta42$ began to diverge [20]. The similar timing of early changes in cognition and the biomarkers was especially notable given the known limitations in sensitivity of the logical memory test as a single indicator. It is not yet clear at the population level that we are gaining *any* early notice by using amyloid biomarkers. We might better serve the goal of early detection by improving reliability and range of cognitive assessments.

Redefining AD as equivalent to brain amyloidosis will create a Tower of Babel in current research. Although there are compelling hypotheses about the role of amyloid, as noted above, we still face substantial uncertainty regarding the biological mechanisms linking amyloid and cognitive decline. Is it possible that amyloid is a byproduct of a disease process, indicating a neurodegenerative process but not the underlying cause of the neurodegenerative process? Could amyloid result from cellular efforts to recover from or reduce the impact of a cellular injury? Perhaps amyloid is a factor that increases cellular vulnerability to other biological insults but has little effect in otherwise healthy cells? If any of these might be true, it is not appropriate to use amyloid burden as a primary criterion. Such a definition could result in adoption of expensive interventions—

many of which have significant side effects—that alter amyloid burden but have no benefits for cognitive or functional well-being.

Why not adopt this new framework and change it later as we learn more?

Scientific hypotheses are constantly tested, revised, or falsified. The biomarker-based diagnostic framework is sometimes framed as a hypothesis, with the expectation that although imperfect, it is a step forward and we can improve it as we go along. Adopting a new criterion for a potentially fatal disease is not analogous to positing a testable scientific hypothesis in other settings, however. Redefining a disease criterion changes the course of research and makes it *more difficult* to evaluate alternative hypotheses. Premature adoption of new criteria, based on incomplete biological understanding and only accessible to a few highly advantaged individuals, could have many adverse consequences.

We should learn here from the numerous public health episodes in which serious mistakes occurred because we misunderstood the biology of biomarkers linked to disease. For example, premature ventricular complexes (PVCs) in post-myocardial infarction (post-MI) patients predict increased risk for death. For years it seemed reasonable to assume drug suppression of PVCs would decrease the risk of post-MI mortality and standard practice was to attempt to suppress such events. The CAST trial showed later that such suppression increased mortality [21]. Prostate specific antigen (PSA) was widely adopted as a standard screen for prostate cancer before we had clear understanding of how to distinguish tumors that were likely to be indolent versus aggressive [22]. Based on premature widespread adoption of PSA criteria, over a million men were needlessly treated for a cancer that would have remained innocuous; prostate cancer treatment often has severe consequences on quality of life, such as impotence and incontinence. PSA is an important example, because once enshrined as a diagnostic instrument, major financial and professional incentives create pressure to retain the standards [23]. A mistake that creates financial incentives for the status quo is very difficult to correct, and the strong financial incentives will render objective scientific discussion and discovery more difficult. Changing the diagnostic criteria for AD will have numerous financial consequences which should be considered when evaluating proposed changes. Such a change will also have immense personal consequences for the individuals diagnosed, most of whom will never manifest symptoms. AD is among the most feared diseases of late life [24]. In 2010, 66% of respondents in the US based Health and Retirement Study believed that individuals with

Alzheimer's disease are not capable of making informed decisions about their own care [25]; the influence of stigma associated with AD on biomarker positive people is unknown.

Conclusions

Biomarkers present an incredible opportunity to evaluate, test, and revise our understanding of the biological mechanisms that may underlie cognitive change and neurodegeneration. If we make a very limited set of biomarkers definitional of disease, ignoring the cognitive syndrome, we throw away that opportunity. Research should be focused on prevention and treatment of the AD clinical syndrome, which is primarily manifested as deteriorations in memory and other cognitive abilities. We should therefore identify and verify a stronger mechanistic link between proposed biomarker and ultimate clinical manifestation of the disease before codifying the biomarkers as diagnostic of the disease itself. Without such a link, a new biomarker-only-based research framework will fail to accelerate scientific progress in preventing and treating AD. We fear this is one step towards a move to define these biomarkers as acceptable surrogate outcomes in clinical trials for prevention or treatment of AD [26]. Such a move would be a financial boon for many, but a tragic betrayal of the interests of millions of people whose lives are affected by clinical AD.

References

1. Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimer's Dement J Alzheimer's Assoc.* 2018;14(4):535–62. <https://doi.org/10.1016/j.jalz.2018.02.018>.
2. Ricciarelli R, Fedele E. The amyloid cascade hypothesis in Alzheimer's disease: it's time to change our mind. *Curr Neuroparmacol.* 2017;15(6):926–35.
3. Brookmeyer R, Abdalla N. Estimation of lifetime risks of Alzheimer's disease dementia using biomarkers for preclinical disease. *Alzheimer's and Dementia.* 2018.
4. Alzheimer's Association. 2017 Alzheimer's disease facts and figures. *Alzheimer's Dementia.* 2017;13(4):325–73.
5. Gu Y, Razlighi QR, Zahodne LB, et al. Brain amyloid deposition and longitudinal cognitive decline in nondemented older subjects: results from a multi-ethnic population. *PLoS ONE.* 2015;10(7):e0123743.
6. Howell JC, Watts KD, Parker MW, et al. Race modifies the relationship between cognition and Alzheimer's disease cerebrospinal fluid biomarkers. *Alzheimers Res Ther.* 2017;9(1):88. <https://doi.org/10.1186/s13195-017-0315-1>.
7. Gottesman RF, Schneider AL, Zhou Y, et al. The ARIC-PET amyloid imaging study Brain amyloid differences by age, race, sex, and APOE. *Neurology.* 2016;87(5):473–80.

8. Gottesman RF, Schneider AL, Zhou Y, et al. Association between midlife vascular risk factors and estimated brain amyloid deposition. *JAMA*. 2017;317(14):1443–50.
9. Ridker PM. A test in context: high-sensitivity C-reactive protein. *J Am Coll Cardiol*. 2016;67(6):712–23.
10. Lee S, Zimmerman ME, Viqar F, et al. Are white matter hyperintensities a core feature of Alzheimer's disease or just a reflection of amyloid angiopathy? Evidence from the Dominantly Inherited Alzheimer Network (DIAN). *Alzheimer's Dement J Alzheimer's Assoc*. 2016;12(7):P226.
11. Bennett DA. Mixed pathologies and neural reserve: implications of complexity for Alzheimer disease drug discovery. *PLoS Med*. 2017;14(3):e1002256.
12. Schneider JA, Arvanitakis Z, Leurgans SE, Bennett DA. The neuropathology of probable Alzheimer disease and mild cognitive impairment. *Ann Neurol*. 2009;66(2):200–8.
13. Stelzmann RA, Norman Schnitzlein H, Reed Murtagh F. An English translation of Alzheimer's 1907 paper, "Über eine eigenartige Erkrankung der Hirnrinde". *Clin Anat*. 1995;8(6):429–31.
14. Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615–25.
15. Fardo DW, Gibbons LE, Mukherjee S, et al. Impact of home visit capacity on genetic association studies of late-onset Alzheimer's disease. *Alzheimer's and Dementia*. 2017.
16. Weuve J, Proust-Lima C, Power MC, et al. Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. *Alzheimer's Dementia*. 2015;11(9):1098–109.
17. Gordon BA, Blazey TM, Su Y, et al. Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: a longitudinal study. *Lancet Neurol*. 2018;17(3):241–50.
18. Müller S, Preische O, Sohrabi HR, et al. Decreased body mass index in the preclinical stage of autosomal dominant Alzheimer's disease. *Sci Rep*. 2017;7(1):1225.
19. Singh-Manoux A, Dugravot A, Shipley M, et al. Obesity trajectories and risk of dementia: 28 years of follow-up in the Whitehall II Study. *Alzheimer's and Dementia*. 2017.
20. Bateman RJ, Xiong C, Benzinger TL, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med*. 2012;367(9):795–804. <https://doi.org/10.1056/NEJMoa1202753>.
21. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: the Cardiac Arrhythmia Suppression Trial. *N Engl J Med*. 1991;324(12):781–8.
22. Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. London: The Cochrane Library; 2013.
23. Otis Webb Brawley M, Goldberg P. How we do harm: a doctor breaks ranks about being sick in America. New York: St. Martin's Press; 2012.
24. Blendon RJ, Benson JM, Wikler EM, et al. The impact of experience with a family member with Alzheimer's disease on views about the disease across five countries. *Int J Alzheimer's Disease*. 2012. <https://doi.org/10.1155/2012/903645>.
25. Health and Retirement Study. Core 2010 public use dataset, experimental module on Alzheimer's disease attitudes. Ann Arbor, MI: Produced and distributed by the University of Michigan with funding from the National Institute on Aging (Grant Number NIA U01AG009740). 2010.
26. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Early Alzheimer's disease: developing drugs for treatment guidance for industry: draft guidance, revision 1. Rockville Maryland U.S. Department of Health and Human Services. 2018.