EDITORIAL

Response to peer commentaries: Composite cognitive and functional measures for early stage Alzheimer's disease trials

We thank all of the authors who contributed the five commentaries in response to our target article, published together in this same volume of the journal as a debate. With respect to Dr. Harrison's commentary, we appreciate that he directly addressed the points raised in our article and generally agreed with our criticisms of composites as they are currently constructed. He points out that global cognitive measures are preferred by drug developers in pharma, perhaps driven by U.S. Food and Drug Administration (FDA) draft guidance. He suggests an empirical approach to composites as, first, to test a range of cognitive domains to identify those impacted by a given treatment, and, second, to comprise a composite of those tests for a pivotal efficacy trial. His key points are that most composites have not presented important psychometric data; and that it is important to establish the psychometric characteristics of a composite; to check that the individual test characteristics are preserved; that after individual scales are identified that their combined use is validated in an appropriate study; and that the measures are appropriate for longitudinal assessment in clinical trials.

Dr. Randolph underscores the important distinction between composites that include assessment of daily function such as the Clinical Dementia Rating Scale Sum of Boxes (CDR-SB), that is endorsed by FDA for early stage Alzheimer's disease (AD) trials,[1] and neuropsychological composites, such as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) and Preclinical Alzheimer Cognitive Composite (PACC), as used in secondary prevention or preclinical AD trials as we discussed.[2] He raises the issue of interpreting clinical meaning for neuropsychological composites by referencing normative data. Indeed, this interpretative issue is why the FDA has accepted the potential for a neuropsychological test or battery to be used as a primary outcome or basis for allowing expedited (provisional) approval for prevention trials and has not accepted a neuropsychological test alone for more advanced mild cognitive impairment (MCI), prodromal, or mild AD trials.

We share Dr. Randolph's concern about the over-reliance and pitfalls in the use of limited neurocognitive data from previous at-risk AD cohorts to derive composites. For example, Dr. Randolph offers that orientation measures that are over-represented in these composites and that appear to drive change in early stage AD are useful as measures of anterograde memory. He extends our comments by highlighting the limited measurement of anterograde memory in clinical trials, and that more direct measures could be used such as list learning and memory for stories.

Drs. Randolph and Duff raise the issue that a composite measure may capture atypical or broader presentations of cognitive impairment associated with AD (eg, posterior cortical atrophy, primary progressive aphasia). However, in clinical trials, the dominant presentation of early stage AD is with memory impairments and one would expect improvement in memory with any effective drug. Composites that are too broad or that assess neuropsychological domains that are unlikely to change due to a drug may dilute an efficacy signal.

We also strongly agree with Dr. Duff that the behavior of tests within a composite may differ compared to their use on a standalone basis because of interference effects. In fact, we are conducting an National Institutes of Health (NIH)–funded instrument development project with a design that attempts to obviate this concern (Novel Measures of Cognition and Function for Preclinical and Prodromal Alzheimer's Disease Trials, R01 AG051346).

Dr. Duff makes two suggestions for next steps. The first is to take advantage of practice effects to predict patient outcomes in trials; and the second is to develop performance-based assessments that require patients to carry out steps in daily activities. He mentions the Naturalistic Action Test.[3] This is also reminiscent of the Direct Assessment of Functional Status (DAFS), which was used in clinical trials in the 1990s.[4] We have long advocated for the use of ecologically relevant performance-based measures, including the University of California San Diego Performance-Based Skills Assessment (UPSA) and presented data about its use in a series of articles.[5,6] The UPSA is now being used in clinical trials of tau antibodies and antivirals, as well as in a Luminosity cognitive training framework. We look forward to the wider use of performance-based instruments in trials.

We agree with Dr. Duff that the presence or absence of practice effects may be a powerful and useful method for assessing AD cognitive stages, selection of patients for clinical trials, and predicting outcomes. Our point, however, is that within a trial that requires serial cognitive assessments, practice effects can result in confounds that produce type 1 and 2 errors, misalign cognitive and biomarker measures, add noise (ie, variability), and lead to interpretative difficulties with regard to the magnitude of any drug versus control effect.

We agree with Dr. Duff's general point that because most composites use the same or very similar measures they will perform similarly, have the same limitations, and that new "from the ground up" measures would be helpful. We alluded to our approach to new measures in a paper on practice effects,[7] measures that informed the design of our previously mentioned instrument development study.

We do not agree with the assertion of Drs. Rentz and Papp that composites, the PACC in particular, "by definition maximize signal to noise ratio," and capture "more subtle" cognitive change in preclinical AD. Such definitions of subtlety and efficiency imply that it is sufficient to derive composites by simply choosing a combination of tests and weightings to gain the largest signal-to-noise ratio; and that such a composite can be applied to any given drug trial without regard to its construct validity and potential relevance to drug mechanisms. Such a composite and weighting may or may not be consistent with sensitivity to change in the new sample or change due to the drug being tested. In fact, given shrinkage from a discovery sample to a validation sample these composites will likely be less sensitive. It is possible that the signal will be diluted in a composite insofar as the signal is a single cognitive domain (eg, memory). We further suggest that the "phenotypic heterogeneity" that Drs. Rentz and Papp describe for preclinical and prodromal AD is mainly memory, as poor memory performance is definitional for preclinical AD and for both amnestic and multi-domain prodromal AD.

Moreover, the PACC and similar approaches rely on the assumption that trajectories of change decline and are constant over the 1.5- to 5-year duration of a given clinical trial, and that no one can improve on a domain. A majority of patients with prodromal AD, of course, would decline over a long trial. However, a substantial minority would, nonetheless, not decline and may improve somewhat on their scores. Moreover, ceiling effects of the scales may attenuate any improvement, constrain the dynamic range of change, and lead to a misleading increase in signal to noise. Although the PACC contains individual tests that are validated, there is no validation of the composite overall. Using psychometrics of individual tests from available studies and cohorts is not the same as knowing the psychometrics of the composite itself as Dr. Harrison pointed out critically it has not been demonstrated that a single measure might not be more effective. We demonstrated that limited memory measures may significantly predict progression from MCI to AD in an Alzheimer's Disease Neuroimaging Initiative (ADNI) sample.[8]

Drs. Sano and Zhu asserted that we used FDA draft regulatory guidance as a "straw person" in order to create "a false or at least partial narrative around" FDA guidance. We are perplexed by this comment as the FDA was explicit in approving the CDR-SB as the primary outcome in prodromal AD trials and it is in fact used as such in early stage Alzheimer's phase 3 amyloid antibody trials.[1,9] Moreover, the FDA clearly states that neuropsychological tests can be used as primary outcomes for accelerated (provisional) approval in secondary prevention trials as well. In addition, we did not state that existing composites are not useful, but that in some clinical trials some composites might not serve their intended purpose.

Drs. Sano and Zhu seem to criticize us for the paper we did not write, for not discussing composites in general, theory-driven composites, or assessments that might capture "the true characterization of clinical change." They give as an example the neuropsychological tests used in the NIA National Alzheimer Coordinating Center's Uniform Data Set for the Alzheimer's Disease Research Centers and state—without evidence—that baseline performance on these tests are "more potent variables in determining the trajectory" than even clinical disease stage. (The tests included in the Uniform Data Set 2 (UDS 2) version are Logical Memory I & II, Digit Span forward & backward, Category Fluency, Trails A&B, WAIS-R Digit Symbol, and Boston Naming Test). We suggest that Drs. Sano and Zhu are conflating the design of a neuropsychological test battery comprising individual tests to assess the performance of people on individual neurocognitive functions, with a neuropsychological composite of tests intended to be combined in order to provide a single overall score for use as a primary outcome in a clinical trial.

Instead of the two kinds of composites currently used in early stage trials, that is, metrical combinations of neuropsychological tests with or without assessments of daily function, Drs. Sano and Zhu propose a "global impression of disease risk stage or severity" that is derived from a "composite that captures cognition, behavior, and function." Minus the cant, they are simply arguing for a kind of global impression rating that includes disruptive behaviors as part of the overall assessment of illness severity. We have no objection to this in principle, but Drs. Sano and Zhu do not provide details, whether this is indeed a composite in which metrics for the three areas are combined, or whether they are suggesting something more impressionistic. Moreover, they do not provide information on how this would work, whether it would serve its intended purpose in trials, be sensitive to change, or clinically interpretable. Their main point seems to be that psychiatric symptoms and disruptive behaviors are not measured as primary outcomes in studies of prodromal AD, and should be combined with assessments of daily function and cognition to create one score. We would again ask how this would be done, as disruptive behaviors do not progress on a continuum or with any ordinality in people with cognitive impairment, and we would not expect or design a potentially disease modifying drug to treat apathy, depression, anxiety, agitation, delusions, and hallucinations, in addition to preserving cognitive function.

We reiterate the need to distinguish types of composites. Composites of selected neuropsychological tests are one example; composites that combine dimensions such as neuropsychological domains, daily and social function, and clinical assessment is another. Finally, we might offer that regardless of the particular composite and individual cognitive tests used, a new treatment with a clinically important effect will have to show clear and consistent—not marginal—effects on most of the outcomes in an adequate and well-controlled trial.

Lon S. Schneider[1]
Terry E. Goldberg[2]

[1]USC Keck School of Medicine, Los Angeles, California
[2]Columbia University Medical Center, New York, New York

Correspondence

Terry E. Goldberg, Columbia University Medical Center, 1051
Riverside Drive,
New York, NY 10032.
Email: teg2117@cumc.columbia.edu

REFERENCES

1. Kozauer N, Katz R. Regulatory innovation and drug development for early-stage Alzheimer's disease. *N Engl J Med.* 2013;368(13):1169-1171.
2. Schneider LS, Goldberg TE. Composite cognitive and functional measures for early stage Alzheimer's disease trials. *Alzheimers Dementia.* 2020; e12017. https://doi.org/10.1002/dad2.12017
3. Giovannetti T, Bettcher BM, Brennan L, et al. Characterization of everyday functioning in mild cognitive impairment: a direct assessment approach. *Dement Geriatr Cogn Disord.* 2008;25(4):359-365.
4. Loewenstein DA, Amigo E, Duara R, et al. A new scale for the assessment of functional status in Alzheimer's disease and related disorders. *J Gerontol.* 1989;44(4):P114-P121.
5. Gomar JJ, Harvey PD, Bobes-Bascaran MT, Davies P, Goldberg TE. Development and cross-validation of the UPSA short form for the performance-based functional assessment of patients with mild cognitive impairment and Alzheimer disease. *Am J Geriatr Psychiatry.* 2011;19(11):915-922. https://doi.org/10.1097/JGP.0b013e3182011846.PMID:22024615
6. Goldberg TE, Koppel J, Keehlisen L, et al. Performance-based measures of everyday function in mild cognitive impairment. *Am J Psychiatry.* 2010;167(7):845-853. https://doi.org/10.1176/appi.ajp.2010.09050692. Epub 2010 Apr 1.PMID:20360320
7. Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS. Practice effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimers Dementia.* 2015;1(1):103-111.
8. Gomar JJ, Bobes-Bascaran MT, Conejero-Goldberg C, Davies P, Goldberg TE; Alzheimer's Disease Neuroimaging Initiative. Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. *Arch Gen Psychiatry.* 2011;68(9):961-969. https://doi.org/10.1001/archgenpsychiatry.2011.96.PMID:21893661
9. FDA. United States Food and Drug Administration. Early Alzheimer's Disease: Developing Drugs for Treatment Guidance for Industry. February 2018. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM596728.pdf (Accessed February 21, 2018).