



Published in final edited form as:

J Prev Alzheimers Dis. 2015 June ; 2(2): 91–102. doi:10.14283/jpad.2015.67.

Methodological Aspects of the Phase II Study AFF006 Evaluating Amyloid-beta -Targeting Vaccine AFFITOPE® AD02 in Early Alzheimer's Disease – Prospective Use of Novel Composite Scales

S. Hendrix¹, N. Ellison¹, S. Stanworth¹, L. Tierney², F. Mattner², W. Schmidt², B. Dubois³, and A. Schneeberger²

¹Pentara Corporation, Salt Lake City, UT, USA

²AFFiRiS AG, Vienna, Austria

³Cognitive and behavioral disease center and Alzheimer's institute, Research INSERM Unit "Cognition, Neuroimaging, and brain diseases, Salpêtrière University Hospital, Paris, France

Abstract

BACKGROUND—Optimized scales and composite outcomes have been proposed as a way to more accurately measure Alzheimer's disease related decline. AFFITOPE® AD02, is an amyloid-beta (A β)-targeting vaccine to elicit anti-A β antibodies. IMM-AD04, commonly known as Alum, originally designated as a control agent, appeared to have disease-modifying activity in a multicenter, parallel group phase II study in early AD patients.

OBJECTIVES—To develop adapted outcomes for cognition, function and a composite scale with improved sensitivity to decline and treatment effects in early AD (mild plus prodromal AD) based on historical data and to assess these adapted outcomes in this phase II study.

DESIGN—Data from public datasets was analyzed using a partial least squares model in order to identify an optimally weighted cognitive outcome, Adapted ADAS-cog, and an optimally weighted ADL outcome, Adapted ADCS-ADL which were prospectively defined as co-primary endpoints for the study and were also combined into a composite scale. Data from 162 patients in the placebo groups of ADCS studies and 156 mild patients in the ADNI I study were pooled for this analysis. The Adapted ADAS-cog scale considered 13 ADAS-cog items as well as several Neuropsychological test items and CogState items, the Adapted ADCS-ADL considered all ADCS-ADL items. After the pre-specified analyses were complete, additional adapted and composite scales were investigated in a post-hoc manner. Evaluation of the adapted and composite

Corresponding Author: Suzanne Hendrix, Pentara Corporation, Salt Lake City, UT, USA, Telephone: ++1-801-898-724, Fax: +1-801-486-7467, hendrix@pentaracorp.com.

Competing interests: AS, FM, WS, and LT are employees of AFFiRiS, the company that commercializes the AFFITOPE® technology described in the manuscript. SH, NE, and SS are consultants for AFFiRiS through Pentara Corporation. BD declares no conflict of interest.

Data and materials availability: The Phase II clinical trial described in this study is registered at www.clinicaltrials.gov. Identifier: NCT01117818.

Ethical standards: The clinical data described in this study was performed in compliance with Good Clinical Practice (GCP), the Declaration of Helsinki (2013), and local legal and regulatory requirements and applicable international regulations.

scales was performed on Phase II trial data for AFFITOPE® AD02 (AFF006, Clinical Trial Identifier: NCT01117818) and historic data in early AD. Least square means, standard deviations, and least squares mean to standard deviation ratios were compared among adapted and composite scales and traditional scales for the 5 treatment groups in the phase II study and overall for the historic data. Treatment effect sizes and p-values were also compared for the phase II study.

RESULTS—Cognitive items that were selected for the adapted cognitive scale (aADAS-cog) and had the highest weights were Word Recall, Word Recognition, and Orientation. Delayed Word Recall and Digit Cancellation were among the items excluded due to lack of improved sensitivity to decline. Highly weighted ADL items included in the adapted functional scale (aADCS-ADL) were using the telephone, traveling, preparing a meal/snack, selecting clothing, shopping and using appliances. Excluded items were primarily basic ADLs such as eating, walking, toileting and bathing. Comparisons between traditional scales and primary outcome adapted scales show improved sensitivity to group differences with the adapted scales in the phase II trial. Most of the improvement in the sensitivity of the aADAS-cog and the aADCS-ADL is due to a larger treatment difference observed rather than the improved sensitivity to decline in the comparison groups.

CONCLUSION—To our knowledge, this is the first study to prospectively use optimized scales as primary endpoints and to demonstrate the superior power of optimized scales and composites in early disease. Although it is possible that the treatment difference between randomized groups is due to a factor other than the treatment itself, for instance baseline imbalance, the improved power to detect these differences still argues in favor of the adapted scales. The issue of oversensitivity to detect treatment effects is controlled by selection of the alpha level for significance, and in our case will happen less than 5% of the time. Clinical relevance of the treatment difference should be assessed separately from statistical significance, and in this phase II study, is supported by significant or similar sizes of effect on function, behaviour and quality of life outcomes, which are important to patients and caregivers.

Keywords

Aluminum; statistical methods; composite; prodromal; Alzheimer's disease

Introduction

Recent years have witnessed numerous failures in the development of Alzheimer's disease (AD) therapeutics (1, 2). Reasons include the inability to intervene early enough in the disease process as a result of the low specificity of the 1984 diagnostic criteria (3), the complex and thus far not entirely understood pathophysiology of the disease (i.e. A β , the focus of many failed trials (4, 5), might be the wrong target) and a lack of validated biomarkers, among others. Today, progress in the diagnosis of AD along with a better understanding of its natural course now allows for earlier interventions. Therapeutic interventions are now shifting to subjects in earlier disease stages, defining a need for scales that are more sensitive to change in early disease stages such as prodromal AD.

Currently, there is no single scale that can measure AD related decline at early stages. There are over 60 scales relevant for AD covering cognitive impairment, activities of daily living,

behavior, and quality of life (2). Common tests, such as the ADAS-cog, were designed for patients with moderate AD. By definition, patients with pre-dementia AD have only subtle cognitive and functional defects. Consequently, patients in the early stage of the disease perform near the ceiling of traditional scales (6, 7). Multiple studies have found that ADAS-cog lacks an adequate response to MCI (8–11), highlighting the gap in our ability to discern disease progression and potential treatment effects at the earliest stages of the disease. At the same time, there is a growing level of understanding of early AD changes. For example, episodic memory and timed executive functioning are two of the most responsive, early cognitive domains that are changed in the healthy elderly to the pre-dementia AD and mild cognitive impairment (MCI) stage (12, 13) as well language, word finding and orientation difficulties (14–16).

Several studies have proposed alternative composite scores, which combine multiple cognitive and functional items into a total score to improve scale sensitivity and to potentially allow for shorter or smaller studies (15, 17–19). Both the FDA and the EMA have additionally suggested that they would not exclude the use of a validated composite scale for the MCI disease stage as sufficient proof for market approval (20, 21).

Here, we develop adapted cognitive and functional scores as well as a combined composite score in order to measure AD related decline in a way that improves sensitivity to decline for patients with mild and prodromal AD. We analyzed pooled Alzheimer's Disease Neuroimaging Initiative (ADNI) mild and Alzheimer's Disease Cooperative Study (ADCS) study placebo data to identify optimal items for measuring disease progression in this early AD population resulting in the creation of an adapted cognitive scale and an adapted functional scale that were then combined into a composite scale. We used a phase II clinical study in mild and prodromal AD patients to test the adapted and composite scales. The clinical study used investigated AFFITOPE® AD02, an A β targeting vaccine eliciting anti-A β antibodies. IMM-AD04, commonly known as Alum, which was originally designated as the control agent, appeared to have disease-modifying activity. Additionally, we used the phase II study data to investigate other questions about composites in a post-hoc manner.

Methods

An optimized scale for measuring cognitive decline, adapted ADAS-cog, and an optimized scale for measuring functional decline, adapted ADCS-ADL, were both developed using a partial least squares (PLS) regression model applied to historical data from the ADCS and ADNI I. A composite outcome was also created by summing the two adapted scales to measure overall decline. These adapted and composite scales were designated as primary outcomes in the AFF006 phase II study, providing prospective validation data.

Adapted ADAS-cog (aADAS-cog)

The Alzheimer Disease Assessment Scale-cognitive subscale (ADAS-cog) (22, 23) is a test battery that assesses performance on eleven cognitive tasks or items: orientation, three trials of a 10-word list learning task, three trials of a 12 word recognition task, recall of instructions, comprehension of commands, object and finger naming, word finding difficulty, expressive language, language comprehension, ideational praxis, and

constructional praxis. The ADAS-cog scale extends from 0 to 70, with higher scores indicating greater cognitive impairment. Two additional items are included in the 13-item version of the ADAS-cog: Delayed Word Recall and Digit Cancellation.

The ADAS-cog was used as the basis for the adapted ADAS-cog (aADAS-cog) scale. The following items were considered for inclusion in the aADAS-cog scale: 1) All ADAS-cog13 items; 2) CogState (New Haven, CT, USA): Continuous paired and associate learning (CPAL), Identification Task (IDN), One Back Memory Task (ONB), Detection Task (DET), Go/No Go Task (GONG); 3) Verbal PAL Immediate/Delayed (Neuropsychological Test Battery, NTB); 4) NTB Category Fluency; and 5) NTB Digit Span forward and backward. All 3 historical studies measured the ADAS-cog13. The items from the ADAS-cog13 were included in the PLS model described below, and the additional items were then considered for inclusion based on individual sensitivity to decline from literature references as described below.

ADCS data are publicly available under (<http://www.adcs.org/>). The ADCS cohort included the placebo group from the 3-arm NSAID study (24) and the placebo group from the Homocysteine study (25). The analysis included data from 162 patients in the pooled placebo group who had cognitive data at 18 months.

Additional data used in this analysis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://www.loni.ucla.edu/ADNI/>). The ADNI cohort included multiple diagnostic groups: patients with AD, subjects with MCI, and healthy elderly (cognitively normal) participants. We used the June 4, 2013 sample including data from 156 mild patients who had cognitive data at 18 months.

The following two methods were used to create the aADAS-cog adapted scale: 1) PLS model using the Alzheimer's disease cooperative study (ADCS) nonsteroidal anti-inflammatory drug (NSAID) trial, the ADCS Homocysteine trial and the Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets and 2) individual item sensitivity to decline as reported in the literature.

Adapted ADCS-ADL (aADCS-ADL)

The Alzheimer's Disease Cooperative Study – Activities of Daily Living (ADCS-ADL) scale (26) is an inventory of informant-based items to assess activities of daily living and instrumental activities of daily living. Informants were asked whether patients attempt each of 23 items in the inventory and, if so, to comment on their levels of performance. Each criterion is graded on the level of dependence: patient performs independently (3 points), patient performs with assistance (1–2 points), or patient is unable to perform (0 points). The level of ADL-dependence is graded via the sum of these item scores with a total score of 0 indicating complete dependence and a total score of 78 indicating total independence.

The adapted ADCS-ADL (aADCS-ADL) included the complete set of items from ADCS-ADL in the PLS model, where items were removed or reweighted to improve sensitivity to decline; items from other ADL scales, such as the DAD which was measured only in the ADNI study, were not considered in this adapted scale, due to minimal availability of data.

The PLS model was applied to the pooled data from the ADCS NSAID study and ADCS Homocysteine study; however, the ADNI data was not used since the ADCS-ADL was not collected in that study.

Composite

The aADAS-cog and aADCS-ADL were combined to create one prospectively defined global composite referred to as “Composite.” A simple sum of the two adapted scores, using the weights obtained prior to scaling them to 100 points, was calculated for the Composite. This score was then scaled to a 100 point range. This approach equates the point values of the cognitive and functional components of the scale rather than weighting them equally.

Partial Least Squares Regression Method

The PLS model referred to above was fit using Proc PLS in SAS® v9.4, in order to identify the combination of either cognitive or functional items that correlated best with decline over time. The initial PLS models included change scores for all ADAS-cog13 or ADCS-ADL items as predictors and time since baseline as the response variable. The Variable Importance for Projection (VIP) statistic was calculated for each predictor in the model (27, 28). The VIP summarizes the contribution a variable makes to the model; therefore, if a variable has a small value of VIP, then it is a prime candidate for deletion. The variable with the lowest VIP is dropped and the PLS model is run again (backward selection) continuing until all variables have a VIP of 0.5 or greater (27, 28). The optimized combination of items was designated as this final weighted item combination.

Test clinical data

Adapted and composite scales were developed for use in a randomized, placebo-controlled, parallel group, double-blind, multicenter phase II trial to assess the clinical and immunological activity as well as the safety and tolerability of repeated s.c. administrations of AFFITOPE® AD02 (AFF006, Clinical Trial Identifier: NCT01117818). Complete details on the study design, patient population, and vaccination application can be found in Schneeberger et al (29).

Comparison of adapted to traditional rating scales

Comparisons of adapted and composite scales to traditional scales for the same domains were used to assess the performance of the novel scales in the phase II study. The aADAS-cog was compared to ADAS-cog11 for assessment of cognition, aADCS-ADL was compared to ADCS-ADL for assessment of function, and the Composite was compared to CDR-sb for a global measure of disease progression.

Any comparison of absolute means and standard deviations or means and standard deviations of change scores can be misleading, even if they are standardized to the total range of the scale since the range is somewhat arbitrary and only part of the range is actually relevant to this stage of disease. For these reasons, we compare only the MSDRs within each of the control groups to evaluate the ability of the scales to measure decline sensitively. The standardized treatment effects are also compared to test the assumption of proportionality of

treatment effects and also to see whether scales that are more sensitive to decline are also more sensitive to treatment effects, resulting in non-proportionality of treatment effects.

LSMean to Standard Deviation Ratios (LSMSDRs) for the 5 treatment groups from this phase II study were used to compare the sensitivity to decline of the scales. The expectation was that the MSDRs for the control group(s) or ineffective groups would be larger for the adapted scales than for the traditional scales, resulting in more power to see treatment effects based on the assumption of a proportional treatment effect. Additionally, the standardized treatment difference was compared between novel and traditional scales, using the decline in the comparison group(s) as the reference standard. A proportional treatment effect would be indicated with a similar percent slowing of progression in the active group versus the comparison group(s).

Post-Hoc analysis and resulting scales—After analyzing the phase II study, there was a concern that the Composite was potentially over weighting cognition at the expense of function. In addition, the CogState and NTB items had not been assessed in a historical dataset, leading to the following questions which were addressed post-hoc:

1. What effect did the CogState and NTB items have on the Composite and aADAS-cog;
2. What was the impact of functional (aADCS-ADL) versus cognitive (aADAS-cog) weighting on the Composite?

Additional composite and adapted scales were derived to answer these questions as described in the following sections.

Adapted ADAS-cog 2: The Adapted ADAS-cog 2 (aADAS-cog 2) is a variation of the pre-specified aADAS-cog that excludes CogState and NTB items and rescales the remaining items so that the range of the score is 0–100. This scale was calculated to investigate the effect that CogState and NTB items had on the pre-specified aADAS-cog.

Composite 2: Composite 2 is a variation of Composite that excludes CogState and NTB items while keeping the same individual item weights. This composite was created to investigate the effect of CogState and NTB items on the Composite.

Balanced Composite: The Balanced Composite is the sum of the aADAS-cog and aADCS-ADL after each is scaled to 50 points resulting in equal weight for cognition and function.

Balanced Composite 2: Balanced Composite 2 is the sum of aADAS-cog 2 and aADCS-ADL after each is scaled to 50 points resulting in equal weight for cognition and function.

Empirical Composite: A PLS model was fit to determine the optimal weighting of cognition (aADAS-cog 2) and function (aADCS-ADL). The aADAS-cog 2 and aADCS-ADL were included in the PLS model and the derived weights were used to create the Empirical Composite. The CogState and NTB items were excluded since these items were not available in the historical data used to derive the weighting. This analysis resulted in a 69% weighting of cognition and a 31% weighting of function.

Comparisons of Scales: The additional scales of aADAS-cog 2 and Composite 2 were calculated to determine what effect the CogState and NTB items had on the pre-specified aADAS-cog and Composite. All other post-hoc composites were derived to investigate how different weights on cognition and function affect the performance of a composite score. These post-hoc scales were compared to the pre-specified aADAS-cog, aADCS-ADL, Composite and traditional scales using the same methods and statistics as above applied to the pooled historical datasets as well as the 5 groups from the phase II study.

Results

Weighted item combination for adapted ADAS-cog

The final PLS model identified six weighted items that efficiently measure decline based on the VIP criterion out of the 13 possible cognitive items after seven iterations, each of which removed one item from the adapted scale and optimally weighting the remaining items.

Based on this final model, it was determined that the best combination included the following ADAS-cog items: Word Recall, Orientation, Word Recognition, Recall Instructions, Spoken Language and Word Finding with weights shown in Table 1.

CogState items and Verbal Paired Associates Learning Immediate and Delayed Recall (VPAL) from NTB were not available in historic datasets, although they were available in the phase II study. Based on literature references, some of the items were determined to be sensitive to change and therefore likely to improve the sensitivity of the ADAS-cog combination (30, 31). However, since there was no way to determine the weights from historic datasets, the weight for these items was an average of the weights selected for other items, scaled based on the range of the new item. Weights of each item were then scaled, such that the range of the new adapted scale would be 0 to 100.

The aADAS-cog score is calculated by summing each item in the composite after the item has been multiplied by its associated weight. The weights for all ADAS-cog items (Table 1) were results from the PLS iterations and the other item weights were derived so as to give average weight to these items:

$$\text{AdaptedADAS-cog (aADAScog)} = 2.02 * \text{Word Recall} + 1.65 * \text{Orientation} + 1.74 * \text{Word Recognition} + 0.68 * \text{Recall Instructions} + 0.99 * \text{Spoken Language} + 1.24 * \text{Word Finding} + 6.62 * \text{ONB} + 0.19 * \text{VPAL} + 0.24 * \text{Category Fluency}.$$

Weighted Item results from PLS for Adapted ADCS-ADL scale

The final model selected and assigned weights to 15 items from the 23 possible item options: 1) find his/ her personal belongings; 2) go shopping; 3) performs hobbies/pastimes; 4) obtain a hot/cold beverage for him/herself; 5) make him/herself a meal or snack; 6) talk about current events; 7) watch television; 8) keep appointments; 9) get around (or travel) outside of his/ her home; 10) he/she left alone; 11) use a household appliance; 12) select his/her clothes for the day/ dressing; 13) read a magazine, newspaper or book; 14) use a telephone; and 15) write things down. The following items were excluded from the

composite: 1) eating, 2) walking, 3) toileting, 4) cleaning / clearing dishes 5) garbage/litter 6) bathing 7) grooming and 8) conversation.

Weights of each item were scaled, such that the range of the composite is 0 to 100 (Table 1). The items that were included in the composite were used to calculate the adapted ADCS-ADL by summing the items after individual weights have been applied:

$$\begin{aligned} \text{aADCS-ADL} = & 1.54 * \text{Belongings} + 1.95 * \text{Shopping} + 1.24 * \text{Hobbies} + 2.10 * \\ & \text{Beverage} + 2.02 * \text{Meal} + 1.27 * \text{Current events} + 1.44 * \text{TV} + 1.83 * \text{Keeping} \\ & \text{Appointments} + 2.05 * \text{Travel} + 1.82 * \text{Alone} + 1.91 * \text{Appliance} + 2.72 * \text{Clothes} + \\ & 1.97 * \text{Read} + 3.39 * \text{Telephone} + 1.83 * \text{Writing}. \end{aligned}$$

Composite scale as a global primary study outcome

The composite primary outcome combines both the aADAS-cog and aADCS-ADL to create an outcome that is sensitive to decline in cognition and function. The weights of the items were rescaled so that Composite ranged from 0 to 100 (Table 1). The calculation for the Composite is as follows:

$$\begin{aligned} \text{Composite} = & 1.66 * \text{Word Recall} + 1.35 * \text{Orientation} + 1.42 * \text{Word Recognition} + \\ & 0.55 * \text{Recall Instructions} + 0.81 * \text{Spoken Language} + 1.01 * \text{Word Finding} + 5.42 * \\ & \text{ONB} + 0.15 * \text{VPAL} + 0.19 * \text{Category Fluency} + 0.28 * \text{Belongings} + 0.35 \text{Shopping} \\ & + 0.23 * \text{Hobbies} + 0.38 * \text{Beverage} + 0.37 * \text{Meal} + 0.23 * \text{Current Events} + 1.26 * \text{TV} \\ & + 0.33 * \text{Keeping Appointments} + 0.37 * \text{Travel} + 0.33 * \text{Alone} + 0.35 * \text{Appliance} + \\ & 0.49 * \text{Clothes} + 0.36 * \text{Read} + 0.62 * \text{Telephone} + 0.33 * \text{Writing}. \end{aligned}$$

The percent contribution for each item as well as for cognitive and functional items combined is shown in Table 1. The composite score was weighted higher for cognition than for function, due to the points on the cognitive scale reflecting smaller changes in the course of the disease than the points on the functional scale. This was also based on the stage of disease which was expected to have a larger decline in cognition than in function.

Adapted scales showed minimal or no improvement in the control group MSDRs

Since 4 of the treatment groups in the phase II test study performed similarly and one (2mg IMM-AD04) showed a decrease in the decline rate compared to the other 4, the 4 groups were then treated as “control” groups. This was supported by comparing the decline rates for the traditional scales in the historical pooled placebo mild data (ADCS placebo data from 2 studies pooled with ADNI mild data) to the decline rates in the 4 ineffective treatment groups (Figure 1) and noting that the historical groups declined faster than these 4 groups. The IMM-AD04 2mg group was assumed to have a positive treatment effect, and the effect sizes in the IMM-AD04 2mg group were calculated relative to the 4 “control” groups. The expectation is that a treatment difference would be more easily detected with the adapted scales relative to the traditional scales, primarily due to a larger MSDR in the “control” groups.

MSDRs were similar between aADAS-cog and ADAS-cog11, and also between aADCS-ADL and ADCS-ADL, indicating minimal, if any, improvement in precision of measurement of decline, or possibly an improvement in precision of measurement that was

counteracted by a milder patient population in the AFF006 study compared to the historical pooled mild patient population. This is supported by the observation that the MSDRs for the traditional scales were larger in the historical group compared to the AFF006 study. Alternatively, the reduced decline rate could be due to a small treatment effect in the “control” groups. The MSDR for the CDR-sb is consistently larger than for the composite score, indicating good precision in measurement of decline over time for the CDR-sb. This is consistent with historical studies that have shown that CDR-sb measures decline consistently and often more sensitively than other scales, even in a pre-dementia stage of disease.

Observed treatment differences were not proportional for adapted scales compared to traditional scales

Treatment effect sizes in the 2mg IMM-AD04 group compared to the “control” groups, as measured by the percentage slowing of decline, were larger for the adapted scales with effects of 36% to 53% for aADAS-cog compared to 32% to 51% for ADAS-cog11; 36% to 44% for aADCS-ADL compared to 12% to 37% for ADCS-ADL; and 43% to 56% for the Composite compared to 19% to 37% for the CDR-sb (Table 2). The difference was especially large comparing the Composite to the CDR-sb, indicating that it has minimal sensitivity to group differences, some of which may be due to treatment effects.

The 2mg IMM-AD04 group demonstrated substantially smaller MSDRs (0.2263) than the “control” groups for the aADAS-cog (1 mg IMM-AD04: 0.6843, 25µg 1mg: 0.936, 25µg 2mg: 0.523, and 75µg 2mg: 0.5595), the aADCS-ADL, Composite, the ADAS-cog11 and the ADCS-ADL (Table 2), consistent with a treatment effect or an unusually slowly declining group. The CDR-sb was the only scale that showed similar MSDR in the 2mg IMM-AD04 group (0.6452) and in the 4 comparison groups (1 mg IMM-AD04: 0.8051, 25µg 1mg: 0.6984, 25µg 2mg: 0.7364, and 75µg 2mg: 0.5283).

Adapted scales sensitivity in mildest AD versus worse mild AD

Adapted scales were tested for their sensitivity within different disease stages by assessing patients at “less mild” (MMSE<23) and “mildest” (MMSE 23+) stages from the phase II clinical study (Table 3).

The adapted cognitive scale showed similar sensitivity within the “control” groups as measured by MSDR to the traditional scales within the mildest AD group and also within the less mild group, with the exception that the CDR-sb showed more sensitivity to decline within the less mild group for all 4 “control” groups. Cognitive scales, both adapted and traditional, showed similar decline rates in the mildest and less mild patient populations for all 4 “control” groups, but ADL scales were generally more sensitive to decline within the less mild group compared to the mildest group. The composite scale performed similarly in the mildest and less mild populations, due to its cognitive emphasis, but the CDR-sb declined more in the less mild group, similar to the ADL scales.

Treatment effects for the IMM-AD04 2mg group in the mildest subjects were strong for both adapted and traditional cognitive and functional scales, but the Composite had a much larger treatment effect than the CDR-sb. In the less mild subjects, no cognitive effects were

seen with either the aADAS-cog or ADAS-cog11, but the aADCS-ADL and the Composite had much larger treatment effects than the ADCS-ADL and the CDR-sb.

Post-Hoc Results—Additional scales of aADAS-cog 2 and Composite 2 were calculated to determine what effect the CogState and NTB items had on the pre-specified aADAS-cog and Composite. All other post-hoc composites, Balanced Composite, Balanced Composite 2 and Empirical Composite, were derived to investigate how different weights on cognition and function affect the performance of a composite score.

Effect of CogState and NTB Items: Results were compared for the aADAS-cog versus the aADAS-cog 2 (which excluded CogState and NTB items). Effect sizes from the aADAS-cog ranged from 36 to 53%, with p-values between 0.022 and 0.254. Effect sizes from the aADAS-Cog 2 ranged from 101 to 102%, with p-values between 0.013 and 0.134.

Similar results are seen when comparing the results from the Composite and Composite 2 (which excluded CogState and NTB items). Effect sizes from the Composite ranged from 43 to 56%, with p-values between 0.032 and 0.227. Effect sizes from Composite 2 ranged from 50 to 63%, with p-values between 0.005 and 0.108.

It appears that the presence of CogState and NTB items hurts the sensitivity to differences of the adapted and composite scales, although it improves (increases) the MSDR within the “control” groups.

Impact of functional versus cognitive weighting on the composite outcome: All post-hoc adapted and composite scales were assessed in the phase II data set and in the historic data (Table 4). Sensitivity to decline and to treatment effects was better for the Optimized Composite compared to Composite and Composite 2. The Balanced Composite and Balanced Composite 2 were not as sensitive to group differences as Composite and Composite 2, even though the MSDRs were higher for the balanced composites, suggesting better sensitivity to decline.

Discussion

We developed adapted cognitive, aADAS-cog, and functional, aADCS-ADL, scales as well as a composite score, Composite, combining both cognition and function, with the goal of establishing scales that are superior to existing ones in measurement of potential decline and treatment effects of patients with early AD. Using this approach, we found that optimizing scale assessment outcomes improved their performance over traditional scales for each domain by demonstrating minimal improvement in the MSDR and an increased signal that the active treatment group had over the “control” groups. The improvement in the MSDR was not as large as anticipated, partly due to bias issues, since the anticipated improvement in MSDR was based on obtaining and testing scales in the same data set. It may also be due to selection of a milder patient population for this study than the pooled mild population that was used for development of the adapted scales. Another possibility is that some of the treatment groups in this study that were assumed to have no effect may actually be demonstrating a slowing of clinical decline.

The increase in treatment effect that was seen was contrary to the usual assumption of proportionality in the treatment effect that is the basis of most sample size calculations. One possible explanation for this increased effect is the possibility that a treatment that only affects AD related decline would be better able to demonstrate a treatment effect on an outcome that is targeted to AD specific decline. The smaller effect seen on the CDR-sb could be due to the CDR-sb measuring non-disease related decline such as normal aging complaints. An AD specific treatment effect would not be expected to slow these types of decline.

Both function and cognition were seen to change, with more cognitive change in the mildest patients, and almost no cognitive change in the less mild patients. More functional change was seen in the less mild patients but was also evident in the mildest half of the patients. The Composite combined cognition and function, but weighted the two unequally, based on assigning cognitive and functional points the same weight. Additional weighting was performed based on the empirically estimated weights of cognition (69%) and function (31%) as well as “balanced” weighting with 50% weight on cognition and 50% weight on function. The disease progression as measured with the Composite in the mild population investigated in this phase II clinical study may be influenced by an overweighting on cognition, however, this did not result in more sensitivity to detect treatment group differences. Changes in function were seen and weighting function equally with cognition resulted in similar, and somewhat stronger, detection of treatment effects, consistent with an AD specific effect rather than a cognition specific effect. The Empirical Composite detected treatment differences with the most sensitivity of any of the composites, presumably due to its weighting cognition and function based on natural weightings of these domains in this stage of disease.

To our knowledge, this is the first study to prospectively use optimized composites as primary endpoints and to demonstrate the superior power of optimized composites in early disease. It was interesting to note that inclusion of the CogState and NTB items in the aADAS-cog scale actually substantially decreased its power to detect group differences, supporting a strictly empirical approach over one based on combining empirical results with literature or expert opinion.

The PLS method employed in this work improves power of the outcome by eliminating items that don't decline over time and optimally weighting declining items. It also incorporates principle component methodology to account for item correlation. Related methodologies to produce composite scores have been successfully applied including the ADAS Tree (32) and ADAS-cog revisited (33).

Historically, many developers of composite scales have assumed proportionality in the treatment effect with the use of more sensitive scales to measure AD, primarily to support statements about increases in power or decreases in sample sizes that could be expected with a more sensitive endpoint. This assumption would imply that the effect sizes would be similar across traditional and adapted scales, but that the p-values for treatment differences would be more significant for the adapted scales due to the increased sensitivity (increased

MSDR of the control group) of the adapted scales relative to the traditional scales. But this is not what is seen in this phase II study.

Based on the MSDRs of the “control” groups in this study, the adapted scores did not perform substantially better than the traditional scales, and in some cases performed worse, which was not too surprising due to the lack of correction for bias in the results of the historical data analysis. However, treatment effects measured were larger for adapted scales vs. traditional scales. This indicates that the impact of an optimized composite on the power of a study may depend more on the ability of the optimized scale to detect treatment effects than on the ability of the scale to measure disease related decline in the control or comparison group. The much larger group differences for the adapted scales and the composite compared to the traditional scales may be due to measuring more disease specific decline, calling into question the common assumption of proportional treatment effects.

These findings are consistent with the theories that led to the development of these adapted scales. For instance, if a treatment shows 20% slowing on a traditional scale that is comprised of 50% relevant item points and 50% irrelevant item points in the particular disease stage, it would be expected to increase to 40% slowing on an adapted scale that only includes the relevant points. In other words, a treatment effect wouldn't be expected to impact the points on the scale that represent noise.

Carefully designed outcome measures for AD can make a big difference in the ability of a clinical study to detect true treatment effects. Current scales leave room for improvement even in a mild AD population, and would be even less effective in earlier stages of disease. Cognition and function are both changing at this stage of the disease, but appear to change at differing rates, bringing into question the idea that they should be similarly sensitive to change. Careful attention to measurement issues in clinical trials will result in improved power for detecting true treatment effects and, at the same time, more confidence in negative results.

Acknowledgments

We would like to thank all the investigators of AFF006 and the DSMB board for their significant contribution to the study. We would also like to thank the Alzheimer's Disease Cooperative Study (ADCS) for data collection and sharing. Additional 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.

References

1. Brooks LG, Loewenstein DA. Assessing the progression of mild cognitive impairment to Alzheimer's disease: current trends and future directions. *Alzheimers Res Ther.* 2010; 2:28. [PubMed: 20920147]
2. Robert P, et al. Review of Alzheimer's disease scales: is there a need for a new multi-domain scale for therapy evaluation in medical practice? *Alzheimers Res Ther.* 2010; 2:24. [PubMed: 20796301]
3. Lopez OL, McDade E, Riverol M, Becker JT. Evolution of the diagnostic criteria for degenerative and cognitive disorders. *Curr Opin Neurol.* 2011; 24:532–541. [PubMed: 22071334]
4. Doody RS, et al. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med.* 2014; 370:311–321. [PubMed: 24450890]
5. Salloway S, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med.* 2014; 370:322–333. [PubMed: 24450891]
6. Cano SJ, et al. The ADAS-cog in Alzheimer's disease clinical trials: psychometric evaluation of the sum and its parts. *J Neurol Neurosurg Psychiatry.* 2010; 81:1363–1368. [PubMed: 20881017]
7. Karin A, et al. Psychometric evaluation of ADAS-Cog and NTB for measuring drug response. *Acta Neurol Scand.* 2014; 129:114–122. [PubMed: 23763450]
8. Doraiswamy PM, Kaiser L, Bieber F, Garman RL. The Alzheimer's Disease Assessment Scale: evaluation of psychometric properties and patterns of cognitive decline in multicenter clinical trials of mild to moderate Alzheimer's disease. *Alzheimer Dis Assoc Disord.* 2001; 15:174–183. [PubMed: 11723368]
9. Hobart J, et al. Putting the Alzheimer's cognitive test to the test II: Rasch Measurement Theory. *Alzheimers Dement.* 2013; 9:S10–S20. [PubMed: 23253779]
10. Samtani MN, et al. An improved model for disease progression in patients from the Alzheimer's disease neuroimaging initiative. *J Clin Pharmacol.* 2012; 52:629–644. [PubMed: 21659625]
11. Samtani MN, et al. Disease progression model in subjects with mild cognitive impairment from the Alzheimer's disease neuroimaging initiative: CSF biomarkers predict population subtypes. *Br J Clin Pharmacol.* 2013; 75:146–161. [PubMed: 22534009]
12. Caselli RJ, et al. The neuropsychology of normal aging and preclinical Alzheimer's disease. *Alzheimers Dement.* 2014; 10:84–92. [PubMed: 23541188]
13. Howieson DB, et al. Trajectory of mild cognitive impairment onset. *J Int Neuropsychol Soc.* 2008; 14:192–198. [PubMed: 18282317]
14. Albert MS, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011; 7:270–279. [PubMed: 21514249]
15. Hort J, et al. Spatial navigation deficit in amnesic mild cognitive impairment. *Proc Natl Acad Sci U S A.* 2007; 104:4042–4047. [PubMed: 17360474]
16. Ostberg P, Fernaeus SE, Hellstrom K, Bogdanovic N, Wahlund LO. Impaired verb fluency: a sign of mild cognitive impairment. *Brain Lang.* 2005; 95:273–279. [PubMed: 16246735]
17. Hendrix S, Wells B. Time Course of Cognitive Decline in Subjects With Mild Alzheimer's Disease Based on ADAS-cog Subscales and Neuropsychological Tests Measured in ADNI. Abstract P4-096. *Alzheimer Dement.* 2010; 6:e50.
18. Hendrix S, et al. A new tool for optimizing responsiveness to decline in early AD - abstract OC12. *J of Nutr Health Aging.* 2012; 16:805.
19. Monteiro IM, et al. Addition of a frequency-weighted score to the Behavioral Pathology in Alzheimer's Disease Rating Scale: the BEHAVE-AD-FW: methodology and reliability. *Eur Psychiatry.* 2001; 16(Suppl 1):5s–24s. [PubMed: 11520474]
20. European Medicines Agency. [Accessed on 10 March 2015] Concept paper on need for revision of the guideline on medicinal products for the treatment of Alzheimer's disease and other dementias. 2013 Oct 24. EMA/CHMP/617734; 2013. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/10/WC500153464.pdf.

21. U.S. Department of Health and Human Services, Food and Drug Administration. Draft Guidance for Industry. [Accessed on 10 March 2015] Alzheimer's disease: Developing drugs for the treatment of early stage disease. 2013 Feb. <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm338287.pdf>.
22. Mohs RC, et al. Development of cognitive instruments for use in clinical trials of antimentia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord.* 1997; 11(Suppl 2):S13–S21. [PubMed: 9236948]
23. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry.* 1984; 141:1356–1364. [PubMed: 6496779]
24. Aisen PS, Schafer KA, Grundman M, Pfeiffer E, Sano M, et al. Effects of rofecoxib or naproxen vs placebo on Alzheimer disease progression: a randomized controlled trial. *JAMA.* 2003; 289:2819–2826. [PubMed: 12783912]
25. Aisen PS, Schneider LS, Sano M, Diaz-Arrastia R, van Dyck CH, et al. High-dose B vitamin supplementation and cognitive decline in Alzheimer disease: a randomized controlled trial. *JAMA.* 2008; 300:1774–1783. [PubMed: 18854539]
26. Galasko D, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord.* 1997; 11(Suppl 2):S33–S39. [PubMed: 9236950]
27. Wold, H. Estimation of Principal Components and Related Models by Iterative Least Squares,”. In: Krishnaiah, PR., editor. *Multivariate Analysis.* New York: Academic Press; 1966.
28. Wold S. *PLS for Multivariate Linear Modeling, QSAR: Chemometric Methods in Molecular Design. Methods and Principles in Medicinal Chemistry.* 1994
29. Schneeberger A, et al. Results from a phase II study to assess the clinical and immunological activity of AFFITOPE® AD02 in patients with early Alzheimer's disease. *J Prev Alz Dis.* 2015; 2(2):103–114.
30. Lowndes GJ, et al. Recall and recognition of verbal paired associates in early Alzheimer's disease. *J Int Neuropsychol Soc.* 2008; 14:591–600. [PubMed: 18577288]
31. Maruff P, et al. Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC Psychol.* 2013; 1:30. [PubMed: 25566378]
32. Llano DA, Laforet G, Devanarayan V. Derivation of a new ADAS-cog composite using tree-based multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. *Alzheimer Dis Assoc Disord.* 2011; 25:73–84. [PubMed: 20847637]
33. Raghavan N, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimers Dement.* 2013; 9:S21–S31. [PubMed: 23127469]

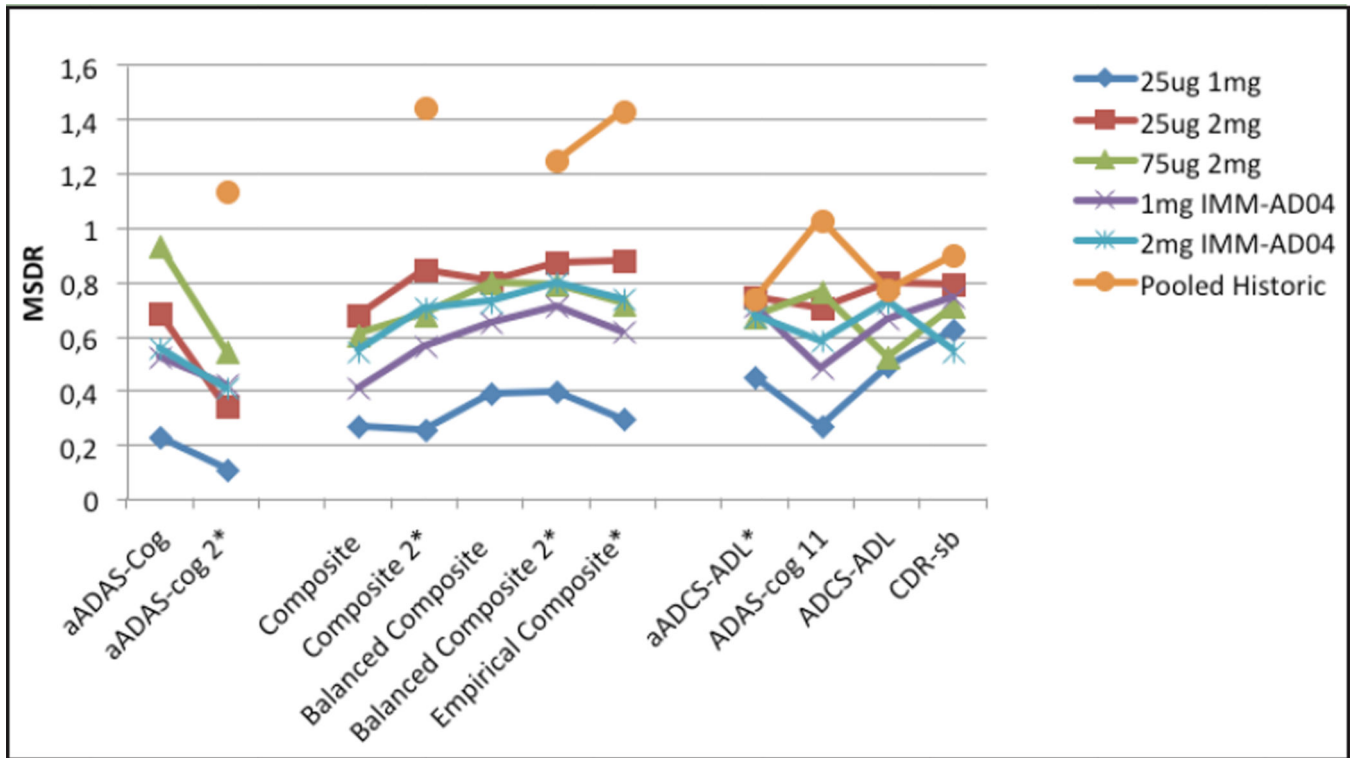


Figure 1. MS DR for Traditional and Composite Scales for AFF006 and Historic Data MS DR change from baseline at 18 months. Composite 2 is Composite without CogState and NTB items. *Estimates for Pooled Historic are biased due to using same data set to derive and assess this scale.

Table 1

Item contribution to the adapted and composite scales. Max: maximum.

Scale	Item	Contribution	Weight	Max	Weighted Max
aADAS-cog	ADAS-cog Word Recall	20.2%	2.02	10	20.25
	ADAS-cog Orientation	13.2%	1.65	8	13.22
	ADAS-cog Word Recognition	20.8%	1.74	12	20.82
	ADAS-cog Recall Instructions	3.4%	0.68	5	3.38
	ADAS-cog Spoken Language	4.9%	0.99	5	4.95
	ADAS-cog Word Finding	6.2%	1.24	5	6.20
	CogState One-Back Memory	10.4%	6.62	1.571	10.40
	NTB VPAL	10.4%	0.19	56	10.40
	NTB Category Fluency	10.4%	0.24	44	10.40
	Total	100%			100
aADCS-ADL	ADCS-ADL Belongings	4.6%	1.54	3	4.62
	ADCS-ADL Shopping	7.8%	1.95	4	7.81
	ADCS-ADL Hobbies	3.7%	1.24	3	3.72
	ADCS-ADL Beverage	6.3%	2.10	3	6.30
	ADCS-ADL Meal	8.1%	2.02	4	8.08
	ADCS-ADL Current Events	3.8%	1.27	3	3.82
	ADCS-ADL TV	4.3%	1.44	3	4.31
	ADCS-ADL Keeping Appointments	5.5%	1.83	3	5.50
	ADCS-ADL Travel	8.2%	2.05	4	8.20
	ADCS-ADL Alone	5.5%	1.82	3	5.45
	ADCS-ADL Appliance	7.7%	1.91	4	7.65
	ADCS-ADL Clothes	8.1%	2.72	3	8.15
	ADCS-ADL Read	3.9%	1.97	2	3.93
	ADCS-ADL Telephone	17.0%	3.39	5	16.96
	ADCS-ADL Writing	5.5%	1.83	3	5.50
	Total	100%			100
Composite	ADAS-cog Word Recall	16.6%	1.66	10	16.57

Scale	Item	Contribution	Weight	Max	Weighted Max
	ADAS-cog Orientation	10.8%	1.35	8	10.82
	ADAS-cog Word Recognition	17.0%	1.42	12	17.04
	ADAS-cog Recall Instructions	2.8%	0.55	5	2.76
	ADAS-cog Spoken Language	4.1%	0.81	5	4.05
	ADAS-cog Word Finding	5.1%	1.01	5	5.07
	CogState One-Back Memory	8.5%	5.42	1.571	8.51
	NTB VPAL	8.5%	0.15	56	8.51
	NTB Category Fluency	8.5%	0.19	44	8.51
	ADCS-ADL Belongings	0.8%	0.28	3	0.84
	ADCS-ADL Shopping	1.4%	0.35	4	1.42
	ADCS-ADL Hobbies	0.7%	0.23	3	0.68
	ADCS-ADL Beverage	1.1%	0.38	3	1.14
	ADCS-ADL Meal	1.5%	0.37	4	1.47
	ADCS-ADL Current Events	0.7%	0.23	3	0.69
	ADCS-ADL TV	0.8%	0.26	3	0.78
	ADCS-ADL Keeping Appointments	1.0%	0.33	3	1.00
	ADCS-ADL Travel	1.5%	0.37	4	1.49
	ADCS-ADL Alone	1.0%	0.33	3	0.99
	ADCS-ADL Appliance	1.4%	0.35	4	1.39
	ADCS-ADL Clothes	1.5%	0.49	3	1.48
	ADCS-ADL Read	0.7%	0.36	2	0.71
	ADCS-ADL Telephone	3.1%	0.62	5	3.08
	ADCS-ADL Writing	1.0%	0.33	3	1.00
	Total	100%			100

Table 2

MSDRs, Effect sizes and p-values at 18 Months

Statistic	Treatment	aADAS-cog	aADCS-ADL	Composite	ADAS-Cog II	ADCS-ADL	CDR-sb
LSMEAN/SD (L.S MSDR)	2mg IMM-AD04	0.3139	-0.5473	0.3094	0.3091	-0.5915	0.7834
	Pooled Excluding 2mg IMM-AD04	0.7779	-0.8325	0.6869	0.7109	-0.7616	0.8727
	1mg IMM-AD04	0.8168	-0.8355	0.7872	0.8358	-0.8752	0.9322
	25ug 1mg	1.0782	-0.7048	0.7167	0.8051	-0.5585	0.8571
	25ug 2mg	0.594	-0.9816	0.5551	0.5476	-0.7819	0.969
	75ug 2mg	0.7371	-0.8347	0.7336	0.7336	-0.9075	0.7925
MSDR	Historical Pooled MCH	0.3549*	-0.3469	0.3532*	0.3491	-0.5017	0.5363
	1 Historical Pooled Mild	1.1336*	-0.7409	1.4405*	1.0309	-0.7757	0.8973
2mg IMM- AD04 Effect Size vs. each group	Pooled Excluding 2mg IMM-AD04	46%	40%	52%	44%	31%	29%
	1mg IMM-AD04	42%	43%	53%	39%	32%	23%
	25ug 1mg	53%	36%	55%	51%	12%	37%
	25ug 2mg	36%	44%	43%	32%	36%	32%
	75ug 2mg	48%	39%	56%	48%	37%	19%
p-value vs. 2mg IMM-AD04	1mg IMM-AD04	0.1899	0.0850	0.0996	0.2396	0.2320	0.3315
	25ug 1mg	0.0220	0.1512	0.0484	0.0351	0.6863	0.0459
	25ug 2mg	0.2543	0.0414	0.2266	0.3431	0.1108	0.1168
	75ug 2mg	0.0553	0.0885	0.0315	0.0603	0.0803	0.3436

* Uses original weights from PLS model - excluding NTB and CogState items; 1. For functional and global scales ADNI was excluded from the Historical Pooled groups.

Table 3

Statistics for Mildest (MMSE 23+) vs. Less Mild (MMSE <23)

Statistic	MMSE	Efficacy Outcome	2mg IMM-AD04	1mg IMM-AD04	25ug 1mg	25ug 2mg	75ug 2mg
LSMEAN	Mildest (23+)	aADAS-cog	-0.2	7.6	7.3	3.6	8.9
		aADCS-ADL	-6.5	-11.6	-9.0	-11.4	-11.9
		Composite	0.4	6.0	7.3	3.3	8.5
	Less Mild (<23)	ADAS-cog 11	-0.3	4.8	5.0	2.6	6.7
		ADCS-ADL	-3.8	-7.1	-3.9	-7.4	-8.3
		CDR-sb	0.8	1.6	1.8	1.9	1.9
		aADAS-cog	11.9	7.7	11.6	12.6	8.6
		aADCS-ADL	-15.4	-23.7	-22.8	-26.2	-20.9
		Composite	8.7	11.0	9.9	12.9	9.6
		ADAS-cog 11	9.1	7.0	9.8	9.6	7.2
LSMEAN/SD	Mildest (23+)	ADCS-ADL	-11.3	-13.4	-12.4	-14.8	-13.5
		CDR-sb	3.1	3.2	3.8	3.5	2.8
		aADAS-cog	-0.0173	0.9453	0.9001	0.3750	0.7759
	Less Mild (<23)	aADCS-ADL	-0.4574	-0.7891	-0.5190	-0.9179	-0.6471
		Composite	0.0392	0.6417	0.9136	0.2918	0.7914
		ADAS-cog 11	-0.0444	0.9736	0.7576	0.3476	0.8251
		ADCS-ADL	-0.4612	-0.8755	-0.3827	-0.9391	-0.7004
		CDR-sb	0.6061	0.8081	0.7347	0.7661	0.8120
		aADAS-cog	0.8368	0.7257	1.2636	1.0161	0.7741
		aADCS-ADL	-0.7805	-1.0561	-0.9887	-1.1471	-1.1957
2mg IMM-AD04 Effect Size vs. each group	Mildest (23+)	Composite	0.6765	0.9910	0.6302	1.0777	0.7090
		ADAS-cog 11	0.6537	0.8216	0.9133	0.9143	0.7222
	Less Mild (<23)	ADCS-ADL	-0.8947	-1.0128	-0.7898	-0.8168	-1.3107
		CDR-sb	1.2157	1.2030	1.0026	1.3208	0.9211

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Statistic	MMSE	Efficacy Outcome	2mg IMM-AD04	1mg IMM-AD04	25ug 1mg	25ug 2mg	75ug 2mg
		Composite		93%	95%	88%	95%
		ADAS-cog 11		106%	106%	112%	104%
		ADCS-ADL		46%	3%	49%	54%
		CDR-sb		50%	56%	58%	58%
	Less Mild(<23)	aADAS-cog		-55%	-3%	6%	-38%
		aADCS-ADL		35%	32%	41%	26%
		Composite		21%	12%	33%	9%
		ADAS-cog 11		-30%	7%	5%	-26%
		ADCS-ADL		16%	9%	24%	16%
		CDR-sb		3%	18%	11%	-11%

Table 4

Composites with Alternate Weighting of Cognition and Function

Pre-Specified Scales	Statistic	2mg IMM-AD04	1mg IMM-AD04	25ug 1mg	25ug 2mg	75ug 2mg	Pooled Historic ¹
aADAS-cog	LSMEAN	4.3	7.4	9.1	6.7	8.3	
	MSDR (MEAN/SD)	0.2278	0.6852	0.9306	0.5253	0.5575	
	Effect Size of 2mg IMM-AD04 vs group		42%	53%	36%	48%	
	P-Value vs. 2mg IMM-AD04		0.1899	0.0220	0.2543	0.0553	
aADCS-ADL	LSMEAN	-9.2	-16.0	-14.3	-16.5	-15.1	21.4*
	MSDR (MEAN/SD)	0.4487	0.7453	0.6739	0.7106	0.6755	0.7409
	Effect Size of 2mg IMM-AD04 vs group		43%	36%	44%	39%	
	P-Value vs. 2mg IMM-AD04		0.0850	0.1512	0.0414	0.0885	
Composite (Sum of unscaled aADAS-cog and aADCS-ADL; 82% cog & 18% ADLs)	LSMEAN	3.8	8.1	8.4	6.7	8.7	
	MSDR (MEAN/SD)	0.2673	0.6800	0.6059	0.4140	0.5465	--
	Effect Size of 2mg IMM-AD04 vs group		53%	55%	43%	56%	
	P-Value vs. 2mg IMM-AD04		0.0996	0.0484	0.2266	0.0315	
Post-Hoc Scales	Statistic	2mg IMM-AD04	1mg IMM-AD04	25ug 1mg	25ug 2mg	75ug 2mg	Pooled Historic ¹
	LSMEAN	-0.1	5.3	7.5	6	7.8	15.8*
	MSDR (MEAN/SD)	0.1114	0.3430	0.5474	0.4265	0.4090	1.1334
	Effect Size of 2mg IMM-AD04 vs group		102%	101%	102%	101%	
Composite 2 (76% cog & 24% ADLs)	P-Value vs. 2mg IMM-AD04		0.1343	0.0192	0.0680	0.0131	
	LSMEAN	3.4	8.1	7.9	6.8	9.2	14.9*
	MSDR (MEAN/SD)	0.2587	0.8482	0.6820	0.5680	0.7034	1.4405
	Effect Size of 2mg IMM-AD04 vs group		58%	57%	50%	63%	
Balanced Composite (50% cog & 50% ADLs)	P-Value vs. 2mg IMM-AD04		0.0421	0.0317	0.1084	0.0045	
	LSMEAN	6.5	11.4	11.2	11.1	11.8	
	MSDR (MEAN/SD)	0.3901	0.8021	0.7991	0.6509	0.7264	--
	Effect Size of 2mg IMM-AD04 vs group		43%	42%	41%	45%	

Pre-Specified Scales	Statistic	2mg IMM-AD04	1mg IMM-AD04	25ug 1mg	25ug 2mg	75ug 2mg	Pooled Historic ¹
Balanced Composite 2 (50% cog & 50% ADLs)	P-Value vs. 2mg IMM-AD04		0.0715	0.0504	0.0627	0.0271	
	LSMEAN	6.8	12.7	12.2	11.9	13.9	20.6*
	MSDR (MEAN/SD)	0.3974	0.8738	0.7911	0.7127	0.7967	1.2490
Empirical Composite (69% cog & 31% ADLs)	Effect Size of 2mg IMM-AD04 vs group		46%	44%	43%	51%	
	P-Value vs. 2mg IMM-AD04		0.0472	0.0401	0.0570	0.0064	
	LSMEAN	0.06838	0.14894	0.14394	0.12909	0.16662	0.2641*
	MSDR (MEAN/SD)	0.2967	0.8786	0.7211	0.6167	0.7381	1.4296
	Effect Size of 2mg IMM-AD04 vs group		54%	52%	47%	59%	
	P-Value vs. 2mg IMM-AD04		0.0411	0.0313	0.0898	0.0046	

* Unadjusted Mean is used instead of LSMean;

¹ For functional and global scales ADNI was excluded.