**Supplementary Material**

**Supplementary Table 1.** ADAS-Cog 11 responsiveness to group-level between-person differences in observed level of disease severity based on exposure status.

|  |
| --- |
| [Reference #] Population (sample size) Exposure |
| **Statistically significant result for test of association between exposure and ADAS-Cog-11** |
| [36] NC (229) Fish Oil Supplement versus None\* |
| [26] NC (229) Aβ |
| [35] MCI (394) Hippocampal Volume\* |
| [35] MCI (394) Entorhinal Thickness\* |
| [35] MCI (394) Fusiform Thickness\* |
| [40] MCI (198) Aβ ≤ versus > 192 pG/mL |
| [40] MCI (198) t-Tau/Aβ > versus ≤ 0.39 |
| [41] MCI (1192) APOE ε4 allele Carrier versus Non-Carrier |
| [42] MCI (1171) APOE ε4 allele Carrier versus Non-Carrier |
| [37] early MCI (162) Aβ |
| [37] late MCI (85) Aβ |
| [37] early MCI (162)Brain Glucose Metabolism |
| [37] late MCI (85)Brain Glucose Metabolism |
| [44] MCI (201) Gait Velocity\* |
| [26] MCI (398) Future Progression to AD at 1 year versus No Progression to AD |
| [26] MCI (398) Aβ |
| [46] MCI (74) Progressive versus Stable MCI |
| [46] MCI (160) Extrapyramidal Signs versus None |
| [48] MCI (392) ChEI versus ChEI and Memantine Hydrochloride versus Neither |
| [50] MCI (516) APOE ε4 allele Carriers versus Non-Carriers |
| [56] MCI (217) Florbetapir – versus + |
| [52] NC/MCI (75) Activities of Daily Living |
| [30] NC-MCI (396) Conversion to AD\* |
| **Non-statistically significant association between exposure and ADAS-Cog-11, where no other cognitive or brain imaging outcome measure found a statistically significant association** |
| [35] NC (225)Hippocampal Volume \* |
| [35]NC (225) Parahippocampal Thickness\* |
| [35] NC (225)Entorhinal Thickness\* |
| [37] NC (126) Brain Glucose Metabolism |
| [37] NC (126) Aβ |
| [38] MCI (286) Body Mass Index\* |
| [36] MCI (397) Fish Oil Supplement versus None\* |
| [53] MCI (51) Aβ + versus – |
| [54] MCI (47) Aβ + versus – |
| [39] MCI (747) Urate Quintiles |
| [45] MCI (173) CSF Neurogranin Quartiles |
| [49] MCI (187) CSF levels Complement 3\* |
| [49] MCI (187) CSF levels Factor H\* |
| [49] MCI (187) Complement 3/Factor H\* |
| [57] MCI Amyloid – subgroup (150) Depressed versus Non-Depressed |
| [57] MCI Amyloid + subgroup (186) Depressed versus Non-Depressed |
| [51] NC/MCI (3069) Lipid Lowering Medication versus None\* |
| **Non-statistically significant association between exposure and ADAS-Cog-11, where at least one other cognitive or brain imaging outcome measure detected a statistically significant association** |
| [35]NC (225) Fusiform Thickness\* |
| [53]NC(69) Aβ + versus – |
| [54]NC (67)Aβ + versus – |
| [34]NC (191)Future Conversion to MCI or AD versus No Future Conversion |
| [35] MCI (394) Parahippocampal Thickness\* |
| [43] MCI (405) Subsyndromal Symptoms of Depression versus None |
| [44] MCI (201) Physical Activity\* |
| [44] MCI (201) Grip Strength\* |
| [55] MCI (101) Subsyndromal Symptoms of Depression versus None\* |

Exposure was treated as a continuous variable, unless otherwise specified. \* indicates adjusted for potential confounders. ADAS-Cog-11, Alzheimer’s Disease Assessment Scale – Cognitive Subscale; MCI, mild cognitive impairment; NC, normal cognition.

**Supplementary Table 2.** Responsiveness to group-level between-person differences of within-person observed change in those estimated to be different based on baseline exposure status.

|  |
| --- |
| [Reference #] Population (sample size) Exposure |
| **Statistically significant result for test of association between exposure and ADAS-Cog-11** |
| [51] NC (2578) Lipid Lowering Medications versus None\* |
| [35] NC (112) AD CSF Signature versus No Signature\* |
| [53] NC (67) Aβ Positive versus Negative\* |
| [54] NC (67) Aβ Positive versus Negative\* |
| [37] NC (76) Aβ Positive versus Negative\* |
| [58] NC (36) Aβ\* |
| [58] NC (104) Brain Glucose Metabolism\* |
| [26] NC (229) Aβ |
| [34] NC (191) Age\* |
| [34] NC (191) Male\* |
| [34] NC (191) Education\* |
| [34] NC (191) APOE ε4 allele\* |
| [34] NC (191) Category (Animal) Fluency\* |
| [34] NC (191) Whole Brain Volume\* |
| [34] NC (186) Hippocampal Volume\* |
| [34] NC (188) t-Tau\* |
| [34] NC (191) p-Tau\* |
| [34] NC (191) Aβ\* |
| [34] NC (188) t-Tau/Aβ\* |
| [34] NC (191) p-Tau/Aβ\* |
| [59] NC Females (137) Serum Uric Acid\* |
| [31] NC (92) Posterior Cortical Hypometabolism |
| [60] MCI (96) Melatonin versus None |
| [35] MCI (193) AD CSF Signature versus No Signature\* |
| [38] MCI (286) BMI\* |
| [53] MCI (46) Aβ positive versus negative\* |
| [53] MCI (46) Florbetapir SUVr\* |
| [54] MCI (46) Aβ positive versus negative\* |
| [33] MCI (231) Brain Atrophy Rates\* |
| [33] MCI (231) Ventricular Expansion\* |
| [61] MCI (50) Melatonin versus None |
| [24] MCI (94) Progressive versus Non-Progressive MCI |
| [42] MCI (1171)APOE ε4 allele Present versus Absent |
| [37] MCI (81) Aβ Positive versus Negative\* |
| [37] MCI (81) Brain Glucose Hypometabolism Positive versus Negative\* |
| [58] MCI (54) Aβ\* |
| [58] MCI (203) Brain Glucose Metabolism\* |
| [58] MCI (390) Hippocampal Volume\* |
| [26] MCI (398) Aβ |
| [45] MCI (173) CSF Neurogranin Quartiles\* |
| [48] MCI (392) ChEIs versus None\* |
| [49] MCI (160) CSF levels Complement 3\* |
| [49] MCI (160) CSF levels Factor H\* |
| [50] MCI (516) APOE ε4 Allele Present versus Absent\* |
| [59] MCI Females (244) Serum Uric Acid\* |
| [19] MCI (102) Phytotherapeutic Compound plus Phosphatidyl Serine and Vitamin E versus Placebo |
| [55] MCI (101) Subsyndromal symptoms of depression versus None\* |
| [57] MCI Amyloid + subgroup (186) Depressed versus Non-Depressed |
| [31] MCI (184) Posterior Cortical Hypometabolism |
| **Non-statistically significant association between exposure and ADAS-Cog-11, where no other cognitive or brain imaging outcome measure detected a statistically significant association** |
| [53] NC (67) Florbetapir SUVr\* |
| [37] MCI (76) Brain Hypometabolism\* |
| [58] NC (228) Hippocampal Volume\* |
| [34] NC (191) Race, white\* |
| [34] NC (191) MMSE\* |
| [34] NC (191) ANART\* |
| [34] NC (191) RAVLT trial 5\* |
| [34] NC (191) RAVLT short recall\* |
| [34] NC (191) TMT A or B\* |
| [34] NC (191) WMS Logical Memory (immediate or delayed)\* |
| [34] NC (191) Boston Naming Test\* |
| [34] NC (191) Ventricle Volume\* |
| [59] NC Males (134) Serum Uric Acid\* |
| [51] MCI (491) Lipid Lowering Medications\* |
| [39] MCI (747) Plasma Urate\* |
| [43] MCI (405) Subsyndromal Symptoms of Depression versus None\* |
| [49] MCI (160) Complement 3/Factor H\* |
| [59] MCI Males (352) Serum Uric Acid\* |
| [57] MCI Amyloid - subgroup (150) Depressed versus Non-Depressed |
| **Non-statistically significant association between exposure and ADAS-Cog-11, where at least one other cognitive or brain imaging outcome measure detected a statistically significant association** |
| [24] MCI (94) Progressive versus Non-Progressive MCI |
| [48] MCI (251) ChEI and Memantine Hydrochloride versus None\* |
| [48] MCI (177) ChEIs and Memantine Hydrochloride versus ChEIs only |

Exposure was treated as a continuous variable, unless otherwise specified. \* indicates adjusted for potential confounders. ADAS-Cog-11, Alzheimer’s Disease Assessment Scale – Cognitive Subscale; MCI, mild cognitive impairment; NC, normal cognition.

**Supplementary Table 3.** ADAS-Cog 11 Responsiveness to Treatment Effects in Pre-Dementia Clinical Trials

|  |
| --- |
| [Reference] Treatment (versus placebo unless otherwise specified) (n) |
| **Statistically significant treatment effect found with ADAS-Cog-11** |
| [63] Multicomponent Cognitive Group Intervention\* (22) |
| [64] Multicomponent Cognitive Group Intervention (24) |
| [65] Omega-3 Polyunsaturated Fatty Acids\* (23) |
| [67] Cognitive Intervention\* (21) |
| [68] Immunoglobulin (49) |
| [73] Aerobic Training (40) |
| [74] Cognitive Training+Transfer Training+Psychomotor Training versus CT+TT versus CT\* (223) |
| [79] Multidomain Training versus Usual Care (113) |
| [80] Di-Huang-Yi-Zhi Formula versus Aniracetam (100) |
| [81] MLC601 versus Placebo (70) |
| **Non-statistically significant treatment effect found with ADAS-Cog-11, where no other outcome measure found a statistically significant association** |
| [62]Donepezil (408) |
| [69] Sodium Benzoate\* (31) |
| [71] Chinese Herbal Medicine versus Donepezil (72) |
| [72] Donepezil and Vitamin E (769) |
| [75] Computerized Multidomain Cognitive Training (51) |
| [76] G Biloba Extract (3069) |
| [77] Multicomponent Exercise with Multitask Conditions versus Educational Classes (100) |
| **Non-statistically significant treatment effect found with ADAS-Cog-11, where at least one other outcome measure detected a statistically significant result** |
| [66] Pro-Cholinergic Drug\* (241) |
| [70] Metaformin\* (80) |
| [75] High Intensity Progressive Resistance Training (49) |
| [78] Rofecoxib (1457) |

**Supplementary Table 4.** Sample Size Estimates to Detect Treatment Effects in Pre-Dementia Clinical Trials

|  |  |  |
| --- | --- | --- |
| **[Reference] Study Details** | **ADAS-Cog 11 rank versus other outcome measures** | |
| [82] Estimate n per treatment arm needed to detect 20% reduction in disease progression over 24 months, with beta=0.20, and alpha=0.05. | | |
| MCI with Aβ | | 6th (n=568) of 6; best=brain atrophy rate (n=46) |
| MCI with Hippocampal Atrophy | | 6th (n >1000) of 6; best=brain atrophy rate (n=77) |
| [83] Estimate n per treatment arm required to detect 25% treatment effect in cognitive measures over 24 and 36 months with beta=0.20 and alpha=0.05. Assessed different sample enrichment strategies. | | |
| NC with APOE ε4 allele, 36 months | | 6th (No decline) of 6; best=RAVLT (n=499) |
| MCI with APOE ε4 allele, 24 months | | 5th (n=908) of 6; best=CDR-SB (n=329) |
| NC with Aβ, 36 months | | 6th (n=420495) of 6; best=RAVLT (n=1090) |
| MCI with Aβ, 36 months | | 3rd (n=639) of 6; best=CDR-SB (n=292) |
| NC with Total CSF Tau, 36 months | | 6th (no decline) of 6; best=RAVLT (n=817) |
| MCI with Total CSF Tau, 24 months | | 4th (n=537) of 6; best=CDR-SB (n=292) |
| NC with CSF Tau phosphorylated at threonine 181, 36 months | | 6th (n=2200678) of 6; best=RAVLT total score (n=559) |
| NC with CSF Tau phosphorylated at threonine 181, 24 months | | 3rd (n=714) of 6; best=CDR-SB (n=296) |
| NC with CSF Total Tau/Aβ, 36 months | | 6th (no decline) of 6; best=RAVLT (n=559) |
| MCI with CSF Total Tau/Aβ, 24 months | | 4th (n=676) of 6; best=CDR-SB (n=258) |
| NC with CSF pTau/Aβ, 36 months | | 6th (n=214455) of 6; best=RAVLT (n=552) |
| MCI with CSF pTau/Aβ, 24 months | | 3rd (n=696) of 6; best=CDR-SB (n=313) |
| NC with Brain Glucose Hypometabolism, 36 months | | 6th (n=13136) of 6; best=CDR-SB (n=1039) |
| MCI with Brain Glucose Hypometabolism, 24 months | | 3rd (n=357) of 6; best=MMSE (n=314) |
| NC with Hippocampal Volume, 36 months | | 6th (n=21359) of 6; best=CDR-SB (n=1057) |
| MCI with Hippocampal Volume, 24 months | | 5th (n=754) of six; best=CDR-SB (n=300) |
| NC with Lateral Ventricle Volume, 36 months | | Tied for 6th (no decline) of 6; best=RAVLT delayed recall (n=1039) |
| MCI with Lateral Ventricle Volume, 24 months | | 3rd (n=666) of 6; best=CDR-SB (n=381) |
| [84] Estimate n per treatment arm to measure 25% reduction in rate of change over 12 months, beta=0.20 and alpha=0.05. | | |
| MCI | | 4th (n=1183) of 5; best=Rate of Annual Brain Volume Loss (n=108) |
| [85] Estimate n required to detect 25% reduction in rate of decline over 12 months with beta=0.20 or 0.10 and alpha=0.05. | | |
| MCI, 80% power | | 6th (n=6797) of 6; best=Atrophy using symmetric Kullback-Leibler S9L5 distance (n=85) |
| MCI, 90% power | | 6th (n=9092) of 6; best=Atrophy using symmetric Kullback-Leibler S9L5 distance (n=114) |

CDR-SB, Clinical Dementia Rating Scale-Sum of Boxes; CSF, cerebrospinal fluid; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; n, sample size; NC, normal cognition; RAVLT, Rey Auditory Verbal Learning Test. Note: Red indicated the ADAS-Cog-11 required the largest sample size of all outcome measures assessed (worst performance).

|  |  |  |  |
| --- | --- | --- | --- |
| **Supplementary Table 5.** Responsiveness of ADAS-Cog-11 Modifications | | | |
| **Outcome Measure** | **Baseline Discrimination** | **Disease Progression** | **Treatment Effect** |
| **ADAS-Cog 13 [90]** | Mean score lower for 229 subjects with NC (mean=9.5, SD=4.1) than 394 subjects with MCI (mean=18.6, SD=6.2) and 187 subjects with AD (mean=28.9, SD=7.6) [25]. In a separate analysis, mean ADAS-Cog 13 score for 382 subjects with MCI (mean=15.23, SD=6.68) was lower than that of 97 subjects with mild AD (mean=29.91, SD=7.44) [2]. Statistical significance of the above differences was not tested. | There was little change on the ADAS-Cog 13 for 382 subjects with MCI over 24 months (mean change=1.34 points) or for 168 subjects with MCI over 36 months (mean change=2.59 points) [2]. There was slightly more change detected in an enriched MCI subgroup (mean 24 month change=2.63 points; mean 36 month change=5.02 points), and no meaningful change on the ADAS-Cog 13 in a non-enriched MCI subgroup (mean 24 month change=-0.18 points, mean 36 month change=-0.15 points) [2]. Among 97 subjects with mild AD there was a modest change in mean ADAS-Cog 13 score over 12 months (mean change=4.35 points) and among 38 subjects with AD over 24 months (mean change=9.46 points) [2]. The SRM for change over 24 months in 382 subjects with MCI, adjusting for baseline age, baseline MMSE score, sex, and APOE ε4 allele, was 0.39 (95% CI 0.16, 0.60) for the ADAS-Cog 13 compared to 0.37 (95% CI 0.15, 0.57) for the ADAS-Cog 11 [2]. The SRM for change over 12 months in 97 subjects with AD was 0.98 (95% CI 0.58, 1.26) for the ADAS-Cog 13 and 0.87 (95% CI 0.46, 1.13) for the ADAS-Cog 11 [2]. Skinner et al. (2012) found the Z-statistic for change over time in 394 subjects with MCI was slightly larger for the ADAS-Cog 13 (Z=10.70) than for the ADAS-Cog 11 (Z=9.44), adjusting for age, education, gender, and APOE ε4 allele [25].Raghavan et al. (2013) also found the ADAS-Cog 13 had larger standardized two-year change than the ADAS-Cog 11 in an MCI sample [15]. | The estimated sample size per group to detect a 25% decrease over 12 months in subjects with MCI with 80% power and an alpha of 0.05 was smaller for the ADAS-Cog 13 (n=900) than for the ADAS-Cog 11 (n=1230) [25].  In a separate study the estimated sample size to detect a hypothetical 25% treatment effect over 2 years in subjects with MCI with 80% power was also smaller for the ADAS-Cog 13 (n=582) than for the ADAS-Cog 11 (n=772) [15]. |
| **VaDAS [91,92]** |  |  |  |
| **ADAS-Cog-12 [9,28,93]** | Sano et al. (2011) showed that 111 subjects with AD had significantly higher mean scores (p<0.001) on the ADAS-Cog 12 (mean=33.27 points, SD=10.3) than 259 subjects with MCI (mean=17.22 points, SD=5.9) [28]. | 12 month unadjusted change scores were significantly different between MCI and AD groups for the ADAS-Cog 11 (*t*=4.26, p<0.001) and ADAS-Cog 12 (*t*=3.89, p<0.001), but the Delayed Word Recall task on its own was not (*t*=-0.45, p=0.654) [28]. Among the MCI group, the 12 month SRM was lower for the ADAS-Cog 11 (0.142) than for the ADAS-Cog 12 (0.160) [28]. The ratio of the SRM for the ADAS-Cog 12 divided by the SRM for the ADAS-Cog 11 was used to show that including Delayed Word Recall with the ADAS-Cog 11 increased the SRM by 12% (more responsive) [28].For the AD group, the 12 month SRM was similar between the ADAS-Cog 11 (0.589) and ADAS-Cog 12 (0.569) [28]. | The estimated sample size required to detect a 33% treatment effect in MCI with 80% power was over 600 subjects lower for the ADAS-Cog 12 than the ADAS-Cog 11 [28]. In contrast, the ADAS-Cog 12 did not outperform the ADAS-Cog 11 for estimations of sample size needed for a trial of AD [28]. |
| **TE4D-Cog [27]** | Scores were significantly better for 25 subjects with NC than for 178 subjects with AD both in terms of overall score (Mann-Whitney *U* test (*U*)=24.0, p<0.001), and each of the seven subscales (p<0.001) [27]. | For 148 subjects with AD, baseline (mean=16.2, SD=11.1) and six month follow-up scores (mean=14.2, SD=10.8) were correlated (*r*=0.90, I<0.001) and there was a statistically significant worsening in scores over time (Wilcoxon signed rank test; Z=-4.9, p<0.001) m[27]. |  |
| **Pooled Index [95]** |  |  | Effect sizes were calculated for secondary analysis of a clinical trial for each individual subscale measure as well as the full Pooled Index for 3, 6, and 12 months of follow-up (Effect size=linear regression coefficient/SE of linear regression coefficient). None of the individual subscale measures demonstrated a statistically significant treatment effect at more than one time point [95]. The Pooled Index found a statistically significant treatment effect at the 3 month and 12 month, but not 6 month, follow-up assessments [95].  AUC analyses of individual scores plotted against time were performed for both the standardized ADAS-Cog 11 and the Pooled Index. The standardized ADAS-Cog 11 showed a statistically significant difference between treatment and placebo groups at the finite time period of 6 months, but not when assessing the 12 month time period as a whole [95]. The Pooled Index showed statistically significant difference between placebo and treatment groups over the entire 12 month period, and at the individual time points of 3 and 12 months, but not at 6 months [95]. |
| **ADAS-Rasch [86]** | Skinner et al. (2012) found that mean scores on the ADAS-Rasch were lower for 229 subjects with NC (mean=4.8, SD=3.5) than for 394 subjects with MCI (mean=11.8, SD=5.5), or 187 subjects with AD (mean=19.5, SD=7.4)[25].  Crane et al. (2012) found mean scores were lower for 225 subjects with NC (mean=4.8, SD=3.5) than for 394 subjects with MCI (mean=11.8, SD=5.5) and 184 subjects with AD (mean=19.5, SD=7.4)[35]. | The Z-score for change over time in 394 subjects with MCI, adjusted for age, education, gender, and APOE ε4 allele, was smaller for the ADAS-Rasch (Z=8.50) than the ADAS-Cog 11 (Z=9.44) [25].  Similar analyses in a separate study found adjusted Z-scores for time were smaller for the ADAS-Rasch than ADAS-Cog 11 in NC (ADAS-Rasch=3.10, ADAS-Cog 11=3.20), MCI (ADAS-Rasch=-10.51, ADAS-Cog 11=-10.78), and AD (ADAS-Rasch=-11.28, ADAS-Cog 11=-12.25) samples [35]. | The estimated sample size per group to detect a 25% decrease over 12 months in MCI with 80% power and an alpha of 0.05 was larger for the ADAS-Rasch (n=1409) than for the ADAS-Cog 11 (n=1230) [25]. Crane et al. (2012) found that the ADAS-Rasch required a larger estimated sample size than the ADAS-Cog 11 to detect a 25% decrease over 12 months, with 80% power and an alpha of 0.05, for NC (41,295 versus 37,971), MCI (1692 versus 1651), and AD (346 versus 242) [35]. |
| **ADAS-Tree [1]** | Statistically significant difference in scores across NC, MCI, and AD diagnostic categories (p<0.0001) [1]. The Kruskal-Wallis test statistic used to assess the magnitude of difference between these categories was larger for the ADAS-Tree (401.1) than the ADAS-Cog 13 (393.3), ADAS-Cog 11 (378.9), and MMSE (368.8) [1].  A separate study found ADAS-Tree scores were lower for 229 subjects with NC (mean=7.9 points, SD=3.5) than 394 subjects with MCI (mean=15.9 points, SD=5.1), and 187 subjects with AD (mean=24.2 points, SD=5.6) [25].  Crane et al. (2012) found mean scores were lower for 225 subjects with NC (mean=7.9, SD=3.5) than for 394 subjects with MCI (mean=15.9, SD=5.1) and 184 subjects with AD (mean=24.2, SD=5.6) [35]. | For 394 subjects with MCI the ADAS-Tree had a larger Z-score for time (Z=12.04) than the ADAS-Cog 11 (Z=9.44), adjusted for age, education, gender, and APOE ε4 allele [25].  Similar analyses in a separate study the ADAS-Tree also had a larger adjusted Z-score for time than the ADAS-Cog 11 in MCI (ADAS-Tree: Z=-13.67, ADAS-Cog 11: Z=-10.78) and AD (ADAS-Tree: Z=-14.05, ADAS-Cog 11: Z=-12.25), but not NC (ADAS-Tree: Z=0.73, ADAS-Cog 11: Z=3.20) samples [35]. | The estimated sample size per group to detect a 25% decrease over 12 months in MCI with 80% power and an alpha of 0.05 was smaller for the ADAS-Tree (n=733) than for the ADAS-Cog 11 (n=1230) [25].  In a separate study the ADAS-Tree required a larger estimated sample size than the ADAS-Cog 11 to detect a 25% decrease over 12 months, with 80% power and an alpha of 0.05, for subjects with NC (573,996 versus 37,971), and a smaller estimated sample size than the ADAS-Cog 11 for subjects with MCI (981 versus 1651) or AD (214 versus 242) hypothetical clinical trials [35]. |
| **cADAS-Cog [94]** |  |  |  |
| **TDAS [96]** |  |  |  |
| **CAMCOG-Plus [97]** |  |  |  |
| **ADAS-Cog -5 Subset [87]** |  |  | The ADAS-Cog-5-Subset found statistically significant differences in the proportion of responders in the treatment compared to control groups for the overall study population (p=0.0001), as well as subgroups of subjects with mild AD (p=0.01), and moderate AD (p=0.01) [87]. The ADAS-Cog 11 found no statistically significant difference between the proportion of responders in the treatment versus control group [87]. |
| **ADAS-Cog-6 Subset [87]** |  |  | The ADAS-Cog-6-Subset found statistically significant treatment effects for the overall study population (p=0.0016) and the moderate AD subgroup (p=0.0002), but not among the subgroup of mild AD subjects (p=0.53) [87]. The ADAS-Cog 11 found no statistically significant difference between the proportion of responders in the treatment versus control group [87]. |
| **ADAS-bifactor [25]** |  | Z-score for change over time adjusting for age, education, gender, and APOE ε4 allele status was larger for the ADAS-bifactor (Z=10.26) than for the ADAS-Cog-11 (Z=9.44) [25]. | Estimated sample sizes to detect a 25% change in cognition over 12 months with 80% power and alpha of 0.05 were calculated [25]. The ADAS-Plus-EF&FA required a smaller sample size (n=547) than the ADAS-Plus EF (n=883), ADAS-Bifactor (n=1103), and ADAS-Cog 11 (n=1230) [25]. |
| **ADAS-Cog-Plus-EF [25]** |  | Z-score for change over time adjusting for age, education, gender, and APOE ε4 allele status was larger for the ADAS-Cog-Plus-EF (Z=10.61) than for the ADAS-Cog-11 (Z=9.44) [25]. | Estimated sample sizes to detect a 25% change in cognition over 12 months with 80% power and alpha of 0.05 were calculated [25]. The ADAS-Plus-EF&FA required a smaller sample size (n=547) than the ADAS-Plus EF (n=883), ADAS-Bifactor (n=1103), and ADAS-Cog 11 (n=1230) [25]. |
| **ADAS-Plus-EF&FA bifactor model [25]** |  | Z-score for change over time adjusting for age, education, gender, and APOE ε4 allele status was larger for the ADAS-Cog-Plus-EF&FA (Z=11.81) than for the ADAS-Cog-11 (Z=9.44) [25]. | Estimated sample sizes to detect a 25% change in cognition over 12 months with 80% power and alpha of 0.05 were calculated [25]. The ADAS-Plus-EF&FA required a smaller sample size (n=547) than the ADAS-Plus EF (n=883), ADAS-Bifactor (n=1103), and ADAS-Cog 11 (n=1230) [25]. |
| **Common Item Pooling [20]** |  |  |  |
| **ADNI Memory Composite [35]** | Scores were higher for subjects with NC (mean=1.0 points, SD=0.5) than subjects with MCI (mean=-0.1 points, SD=0.6) and subjects with AD (mean=-0.8 points, SD=0.5) [35]. | Standardized regression coefficients for time, controlling for age, education, and sex, and presence of at least one APOE ε4 allele were statistically significant for the ADNI Memory Composite in NC (3.02), MCI (-9.43), and AD (-11.59) subgroups (all p<0.05) [35]. In comparison, coefficients for the ADAS-Cog 11 were larger in the NC (3.20), MCI (-10.78), and AD (-12.25) subgroups (all p<0.05) [35]. | Standardized coefficients and adjusted SD were used to estimate the sample size needed to detect a 25% reduction in rate of cognitive decline over 12 months with 80% power in a hypothetical two-arm clinical trial [35]. The ADNI Memory Composite required a smaller sample size than the ADAS-Cog 11 for a hypothetical trial of NC (28,512 versus 37,971), but required a larger sample size than the ADAS-Cog 11 for MCI (2,167 versus 1,651) and AD trials (568 versus 242) [35]. |
| **ADAS-Cog IRT [88,89]** |  |  | Verma et al. (2015) used clinical trial simulations to compare the ADAS-Cog 11 and ADAS-Cog IRT in terms of the power needed to detect a pre-specified treatment effect for various sample sizes (n=200 to 1,000) over 24 months, and for various lengths of follow-up with the sample size set at 400 [89]. Both ADAS-Cog IRT scoring methodology and original ADAS-Cog 11 scoring with an Analysis of Covariance test for a treatment effect showed low power (< 80%) for detecting a mild treatment effect regardless of the sample size or trial duration [89]. For a moderate treatment effect, ADAS-Cog IRT methodology reached 80% power with a smaller sample size and shorter trial duration compared to original ADAS-Cog 11 methods [89]. Sensitivity analysis in a real clinical trial was also performed where the ADAS-Cog IRT scoring methodology detected a larger treatment effect than original ADAS-Cog 11 methods [89]. |
| **ADAS-3b [15]** |  | The standardized two-year change of the ADAS-3b was larger than that of the ADAS-Cog 11 and all individual tasks of the ADAS-Cog 11[15]. |  |
| **CC1 [15]** | Scores for 377 subjects with MCI were worse (mean=0.15 points, SD=1.64) than for 192 subjects with AD (mean=3.15 points, SD=1.68), or 142 subjects with MCI and Aβ pathology (mean=0.49 points, SD=1.55) [15]. | The CC1 demonstrated greater standardized two-year mean change than the ADAS-Cog 11 and all individual items of the ADAS-Cog 11[15]. | The estimated sample size required to detect a hypothetical 25% treatment effect with 80% power in a two-arm clinical trial of subjects with MCI was smaller for the CC2 (n=300) than the CC1 (n=477), the ADAS-Cog 11 (n=772), and the CDR-SB (n=375) [15].  The estimated sample size required to detect a hypothetical 25% treatment effect with 80% power in a two-arm clinical trial of subjects with AD was smaller for the CC2 (n=160) than the CC1 (n=189), the ADAS-Cog 11 (n=256), and the CDR-SB (n=193) [15]. |
| **CC2 [15]** | Scores for 377 subjects with MCI scored worse (mean=0.07 points, SD=0.94) than for 192 subjects with AD (mean=2.38 points, SD=1.28), or 142 subjects with MCI and Aβ pathology (mean=0.22 points, SD=0.95) [15]. | The CC2 demonstrated greater standardized two-year mean change than the ADAS-Cog 11 and all individual items of the ADAS-Cog 11[15]. | The estimated sample size required to detect a hypothetical 25% treatment effect with 80% power in a two-arm clinical trial of subjects with MCI was smaller for the CC2 (n=300) than the CC1 (n=477), the ADAS-Cog 11 (n=772), and the CDR-SB (n=375) [15].  The estimated sample size required to detect a hypothetical 25% treatment effect with 80% power in a two-arm clinical trial of subjects with AD was smaller for the CC2 (n=160) than the CC1 (n=189), the ADAS-Cog 11 (n=256), and the CDR-SB (n=193) [15]. |
| **CFC1[15]** | Scores for 377 subjects with MCI were worse (mean=-0.11 points, SD=1.02) than for 192 subjects with AD (mean=2.4 points, SD=1.42), or 142 subjects with MCI and Aβ pathology (mean=0.06 points, SD=0.98) [15]. | The CFC1 demonstrated a larger standardized mean change for MCI participants than the ADAS-Cog 11 and all individual ADAS-Cog 11 tasks [15]. | The estimated sample size required to detect a hypothetical 25% treatment effect with 80% power in a two-arm clinical trial of subjects with MCI was smaller for the CFC2 (n=302) than the CFC1 (n=348), the ADAS-Cog 11 (n=772), and the CDR-SB (n=375) [15].  For subjects with AD the estimated sample size was also smaller for the CFC2 (n=120) than the CFC1 (n=125), the ADAS-Cog 11 (n=256), and the CDR-SB (n=193) [15]. |
| **CFC2[15]** | Scores for 377 subjects with MCI were worse (mean=-0.13 points, SD=1.0) than for 192 subjects with AD (mean=2.48 points, SD=1.51), or for 142 subjects with MCI and Aβ pathology (mean=0 points, SD=1.01) [15]. | The CFC2 demonstrated a larger standardized mean change for MCI participants than the ADAS-Cog 11 and all individual ADAS-Cog 11 tasks [15]. | The estimated sample size required to detect a hypothetical 25% treatment effect with 80% power in a two-arm clinical trial of subjects with MCI was smaller for the CFC2 (n=302) than the CFC1 (n=348), the ADAS-Cog 11 (n=772), and the CDR-SB (n=375) [15].  For subjects with AD the estimated sample size was also smaller for the CFC2 (n=120) than the CFC1 (n=125), the ADAS-Cog 11 (n=256), and the CDR-SB (n=193) [15]. |
| **Parmacometric ADAS-Cog 13 [7]** |  |  | The longitudinal IRT model was used to simulate 20-month two-arm clinical trials with a 20% treatment effect for 100, 200, 400, or 800 subjects with mild to moderate AD [7]. Five hundred simulations were run for each sample size [7]. The IRT based pharmacometric model required 71% fewer subjects than the Least-square mean analysis, and 23% fewer subjects than the pharmacometric model, to detect a treatment effect with 80% power and no inflation of Type I error [7]. |
| **IRT& Pharmacometric ADAS-Cog 13 [7]** |  |  | The longitudinal IRT model was used to simulate 20-month two-arm clinical trials with a 20% treatment effect for 100, 200, 400, or 800 subjects with mild to moderate AD [7]. Five hundred simulations were run for each sample size [7]. The IRT based pharmacometric model required 71% fewer subjects than the Least-square mean analysis, and 23% fewer subjects than the pharmacometric model, to detect a treatment effect with 80% power and no inflation of Type I error [7]. |
| **iADRS [98]** |  | The iADRS had a larger SRM, based on visual inspection of forest plots, for MCI and mild and moderate AD compared to the ADAS-Cog 11, ADAS-Cog 13, MMSE, FAQ, CDR-SB, and several other measures of cognition [98]. | For several clinical trials including subjects with MCI or mild AD the iADRS was able to detect a statistically significant treatment effect, however the magnitude of this effect was not consistently better than that detected by the ADAS-Cog 14 [98]. |
| **Straightforward Sensitive Scale [99]** |  | The SRM of the SSS in subjects with MCI was greater than that of the CDR-SB alone or the ADAS-Cog 13 over 1 year (SRM: SSS=0.62, CDR-SB=0.55, ADAS-Cog 13=0.28), two years (SRM: SSS=0.82, CDR-SB=0.74, ADAS-Cog 13=0.56), three years (SRM: SSS=0.93, CDR-SB=0.76, ADAS-Cog 13=0.65), and when assuming a hypothetical treatment effect delayed disease progression by one year (SRM: SSS=0.37, CDR-SB=0.35, ADAS-Cog 13=0.29)[99]. The SSS maintained the highest SRMs for subgroups of subjects with MCI and biomarkers indicating increased risk of disease progression [99]. | The SSS was estimated to require a smaller sample size (n=189) to detect a hypothetical treatment effect that slows disease progression by 50% in a two-year MCI trial compared to the CDR-SB (n=231) and ADAS-Cog 13 (n=402) [99]. |
| **ADAS-Cog 3 [2]** | Mean scores for 382 subjects with MCI (mean=8.23 points, SD=3.76) were lower than scores for 97 subjects with mild AD (mean=15.95 points, SD=4.15) [2]. | There were small changes on the ADAS-Cog 3 for 382 subjects with MCI over 24 months (mean=0.71 points, SD=3.56) and for 169 subjects with MCI over 36 months (mean=1.23 points, SD=4.00) [2]. There was also little change in an enriched MCI subgroup (mean 24 month change=1.48 points, SD=3.78; mean 36 month change=2.55 points, SD=4.40), and almost no change in a non-enriched subgroup (mean 24 month change=-0.19 points, SD=3.06; mean 36 month change=-0.25 points, SD=5.12) [2]. Among 97 subjects with mild AD there was also little change in scores over 12 months (mean=1.82 points, SD=3.91) and 24 months (mean=3.81 points, SD=5.12) [2]. The SRM for change over 24 months in 382 subjects with MCI was 0.42 (95% CI 0.20, 0.61) for the ADAS-Cog 3 and 0.37 (95% CI 0.15, 0.57) for the ADAS-Cog 11, adjusting for baseline MMSE, age, sex, APOE ε4 allele [2].The SRM for change over 12 months in 97 subjects with mild AD was 0.81 (95% CI 0.43, 1.09) for the ADAS-Cog 3 and 0.87 (95% CI 0.46, 1.13) for the ADAS-Cog 11, adjusting for baseline MMSE, age, sex, APOE ε4 allele [2]. SRMs were not statistically different from each other (all p>0.10) [2]. |  |
| **ADAS-Cog 5 [2]** | Scores for 382 subjects with MCI (mean=13.96 points, SD=6.17) were lower than those of 97 subjects with mild AD (mean=26.20 points, SD=5.31) [2]. | There were small changes in ADAS-Cog 5 scores for 382 subjects with MCI over 24 months (mean change=1.13 points, SD=4.87) and for 168 subjects with MCI over 36 months (mean change=1.95 points, SD=5.58) [2]. There was very little difference in an enriched MCI subgroup (mean 24 month change=2.21 points, SD=5.58; mean 36 month change=3.82 points, SD=6.03), and in an non-enriched subgroup scores there was no meaningful change (mean 24 month change=-0.11 points, SD=4.12; mean 36 month change=-0.16 points, SD=4.15) [2]. Among 97 subjects with mild AD there also was very little change on the ADAS-Cog 5 score over 12 months (mean change=2.64 points, SD=4.39) and 24 months (mean change=5.48 points, SD=6.13) [2]. The SRM for the ADAS-Cog 5 for 382 subjects with MCI over 24 months was 0.42 (95% CI 0.19, 0.63), adjusting for baseline MMSE, age, sex, APOE ε4 allele [2].The SRM for change on the ADAS-Cog 5 over 12 months in 97 subjects with mild AD was 0.93 (95% CI 0.52, 1.22), adjusting for baseline MMSE, age, sex, APOE ε4 allele [2]. SRMs for the ADAS-Cog 5 were not significantly different than SRMs for the ADAS-Cog 11, ADAS-Cog 13, or ADAS-Cog 3 (all p>0.10) [2]. |  |
| **ADAS-13RW [100]** |  |  |  |
| **ADCOMS [101]** |  | The 12-month SRM of the ADCOMS (0.419) was larger than that of the ADAS-Cog 12 (0.196), MMSE (0.221), and CDR-SB (0.353) for subjects with aMCI, as well as subgroups of aMCI subjects with genetic or CSF AD biomarkers present [101]. | The ADCOMS was able to detect a statistically significant treatment effect for donepezil compared to placebo for aMCI participants (p=0.02), which was also found by the MMSE (p=0.02), but not by the ADAS-Cog 12 (p=0.12) or CDR-SB (p=0.11) [101]. The ADCOMS did not find a statistically significant effect for vitamin E in subjects with aMCI (p=0.89), nor did the ADAS-Cog 12 (p=0.76), MMSE (p=0.59), or CDR-SB (p=0.42) [101]. The ADCOMS was also able to detect a statistically significant treatment effect for donepezil in subjects with mild AD (p<0.0001) as did the ADAS-Cog 12 (p=0.0008), MMSE (p=0.001), and CDR-SB (p=0.02) [101]. |

CV, Coefficient of Variation (Standard Deviation/Mean); NC=Normal Cognition; MCI, mild cognitive impairment; ADAS-Cog, Alzheimer’s Disease Assessment Scale-Cognitive Subscale; VaDAS, Vascular Dementia Assessment Scale; N/A, not available; GDS, Geriatric Depression Scale; DBRI, Dysfunctional Behavior Rating Instrument; MMSE, Mini-Mental State Examination, ADL, Activities of Daily Living; IRT, Item Response Theory; OPLM, One Parameter Logistic Model; cADAS-Cog, Computerized ADAS-Cog; TDAS, Touch Panel Type Dementia Assessment Scale ; CAMCOG, Cambridge Cognitive Examination; EF, Executive Function; FA, Functional Assessment; TMT, Trail Making Test; DSST, Digit Symbol Substitution Test; FAQ, Functional Assessment Questionnaire; ADNI, Alzheimer’s Disease Neuroimaging Initiative; RAVLT, Rey Auditory Visual Learning Test; CC, Cognitive Composite; AVLT-Immed, Auditory Visual Learning Test–Immediate; CDR-SB, Clinical Dementia Rating Sum of Boxes; CFC, Cognitive Functional Composite; iADRS, Integrated Alzheimer’s Disease Rating Scale; iADL, Instrumental Activities of Daily Living; ADAS-13RW, ADAS-Cog 13 Re-weighted; ADCOMS, Alzheimer’s Disease Composite Score. Superscripts refer to reference numbers at the end of Chapter 3. When compared to the ADAS-Cog, results for both the modified and original ADAS-Cog 11 are presented. When a modified version was compared to another modified version or another standard global outcome measure such as the MMSE or CDR-SB, then results from those modified or alternative measures are also presented. If comparison was made with the ADAS-Cog 11 as well as other standard cognitive outcome measures, then only the results fostering comparison with the ADAS-Cog 11 are presented.

**GLOSSARY**

**Responsiveness** is the ability to detect change, and is population and context specific. The population specific aspect of responsiveness means that even though the ADAS-Cog has been shown to detect important changes in dementia, it may not be able to detect important changes at pre-dementia levels of impairment. The context specific aspect of responsiveness means that an outcome measure may be highly responsive to particular treatment effects in a clinical trial, but not very responsive to changes in an observational study or a clinical setting where a physician is assessing a single patient. Thus, when talking about responsiveness it is important to specify the type of change one is interested in detecting. This can be done using three axes of classification described by de Beaton et al. (2001):

1) The “Who” axis differentiates between individual level and group level of analysis and interpretation. A group-level analysis and interpretation of change is often used for research studies, but outcome measures will require adequate levels of responsiveness to individual-level change if they are intended to also be used for one-on-one assessments, such as in a clinical setting.

2) The “Which” axis describes whether the scores being contrasted are measuring between-person differences at one point in time, within-person changes over time, or between-person differences of within-person change over time.

3) The “What” axis specifies the type of change being quantified in the study, such as minimum potentially detectable change by the instrument, observed change measured by an instrument in a population, or observed change in a population deemed to have improved by a clinician.

**Baseline Discrimination**: Responsiveness to group-level between-person differences in stage of disease progression at one point in time. An outcome measure demonstrating responsiveness to baseline discrimination will be able to distinguish between groups of study participants that have been grouped according to different diagnostic categories where the categories are considered to follow a somewhat stepwise increase in severity, such as normal cognition, mild cognitive impairment, and dementia.

**Disease Progression:** Responsiveness to group-level within-person observed change measured by an outcome measure in a given population. An outcome measure responsive to disease progression will be able to detect changes over time in a cohort study that does not administer an intervention.

**Treatment Effect:** Responsiveness to group-level between-person differences of within-person observed change over time. An outcome measure demonstrating responsiveness to a treatment effect would be able to tell if a group of participants receiving donepezil improved over the course of a randomized controlled trial compared to a group of participants receiving placebo.

**Glossary reference**

De Beaton E, Bombardier C, Katz JN, Wright JG (2001) A taxonomy for responsiveness. *J Clin Epidemiol* **54**, 1204-1217.