American Journal of Alzheimer's Disease & Other Dementias®
2014, Vol. 29(6) 492-497

DOI: 10.1177/1533317514522539
aja.sagepub.com
SAGE

Clinical Significance in Dementia Research: A Review of the Literature

Syed H. Shabbir, BSc[1], and Amy E. Sanders, MD, MS[1]

Abstract

Clinical research traditionally relies on measures of statistical significance to assess the strength of evidence while less attention is paid to the practical import of the results. The objective of this study was to provide a critical overview of the current approaches to measuring clinical significance in dementia research and to provide suggestions for future research. A systematic search was conducted of Medline and Embase for original, English-language, peer-reviewed articles published before July 2012. A total of 18 articles met the inclusion criteria, of which 13 used multiple approaches to measure clinical significance. In all, 5 articles used expert opinion as anchors; 4 also used distribution-based approaches. In all, 8 articles used Goal Attainment Scaling; 7 of these also relied on clinician-based impressions of change. Another 3 articles used only clinical global impressions of change, 1 article used changes in symptomatology, and another used the value from literature.

Keywords

systematic review, minimal clinically significant difference, clinical significance, dementia, Alzheimer's, cognitive impairment

Introduction

Clinical research relies on measures of statistical significance to assess the strength of evidence. In contrast, relatively less attention is paid to the clinical significance and practical import of the results. The term minimal clinically important difference (MCID) was first described by Jaeschke and colleagues in 1989 as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management."[1] The definition has taken varied constructs since then but the common denominator remains that of the smallest numerical difference which has been defined in some way as being clinically important.[2,3] The change itself may be toward harm or benefit and is usually variable between outcomes of interest and patient populations.

Change serves as an important index of disease progression in dementia. However, the assessment and quantification of change can often be quite challenging for researchers and clinicians alike as dementia manifests as a variety of symptoms that present, stay the same, reoccur, and/or subside over the course of the illness.[4-6] As such, measurements of MCID are doubly important in providing researchers and physicians with a frame of reference to evaluate clinically relevant progress or deterioration of their patients. According to an annual report released by the Alzheimer's Association, 5.4 million Americans of all ages were estimated to have Alzheimer's disease (AD) in 2012, with a projected rise to 11 to 16 million people by 2050.[7] Current initiatives emphasizing patient-centered approaches and expected increases in dementia incidence and prevalence in the coming decades combine to underscore the need to define and use outcomes with clinical significance.

Unfortunately, little consensus on optimal approaches and best practices exists at present. Outcome scales like the Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog)[8] have long been used in research and clinical trials with a 4-point change at 6 months taken to be clinically meaningful.[9] However, the ADAS-Cog has been shown to have little correlation with other clinical measures and the 4-point change to have little inherent meaning for the purposes of patient assessment and/or physician decision making.[10-12] Moreover, there is incredible variability in the terminology and definitions of MCID used in the literature. Although some only have a nuanced difference in semantics and are identical operationally, others are completely different approaches altogether.[3] For instance, Molnar et al outline differences between MCID from minimally detectable difference (MDD) and Goal Attainment Scaling (GAS) among other similar approaches. Like MCID, MDD is the smallest detectable change in the outcome scale but unlike MCID, it is not necessarily a clinically meaningful one.[13] The GAS, on the other hand, is a different approach

[1] Saul R. Korey Department of Neurology, Albert Einstein College of Medicine, Bronx, NY, USA

Corresponding Author:
Amy E. Sanders, MD, MS, Albert Einstein College of Medicine, Louis and Dora Rousso Building, 1165 Morris Park Avenue, Room 311, Bronx, NY 10461, USA.
Email: amy.sanders@einstein.yu.edu

that relies on patient or caregiver assessment of meaningful changes in behaviors identified as being important.[14] Unlike MCID, neither MDD nor GAS take into account the side effects and costs associated with the intervention and as such do not mirror the cost–benefit considerations undertaken by the patients, caregivers, and physicians.[13] King presents a very concise account of some of the definitions and terminologies in common use and their evolution over time.[3]

The purpose of this study was to conduct a systematic search of the existing dementia literature for measures of clinical significance, outline their strengths and limitations, and provide recommendations for future research.

Methods

An electronic literature search of Medline (Pubmed) and Embase was conducted for articles published before July 1, 2012. The following free-text search was carried out using different possible permutations of MCID:

((“clinically meaningful change”) OR (“clinically important difference”) OR (“minimally important difference”) OR (“minimally detectable difference”) OR (“minimal important difference”) OR (“clinically important change”) OR (“clinically significant difference”) OR (“clinically relevant change”) OR (“minimal clinical difference”) OR (“Goal Attainment Scaling”) OR (“MCRC”) OR (“MCID”) OR (“MCSD”))

AND

((“cognitive impairment”) OR (“cognitive decline”) OR (“dementia”) OR (“Alzheimer’s”) OR (“memory decline”) OR (“memory impairment”) OR (“mental impairment”)).

The systematic review was restricted to studies published in English-language, peer-reviewed journals. Figure 1 details the literature search and study selection process. The initial combined search of Medline (Pubmed) and Embase retrieved 75 published articles including 28 duplicates. After removal of the duplicates, the remaining 48 studies were assessed for (1) a diagnosis of dementia in at least some of the participants, (2) a measure of clinical significance in study design or analysis, and (3) original research (ie, no editorials, reviews, comments).

Results

Eighteen studies met the inclusion criteria for the systematic review. Most studies (13 of 18; 72%) used multiple approaches to measure clinical significance. In all, 5 articles used expert opinion as anchors of which 4 used distribution-based approaches (such as effect size and standardized response means) as well. Eight articles used GAS of which most also used clinical global impressions of change. In all, 3 other articles only used clinical global impressions of change, 1 article used changes in symptomatology, and another used the value of MCID from literature.

Anchor-Based Methods

Anchor-based methods employ the relationship between the instrument of interest and any number of external indicators

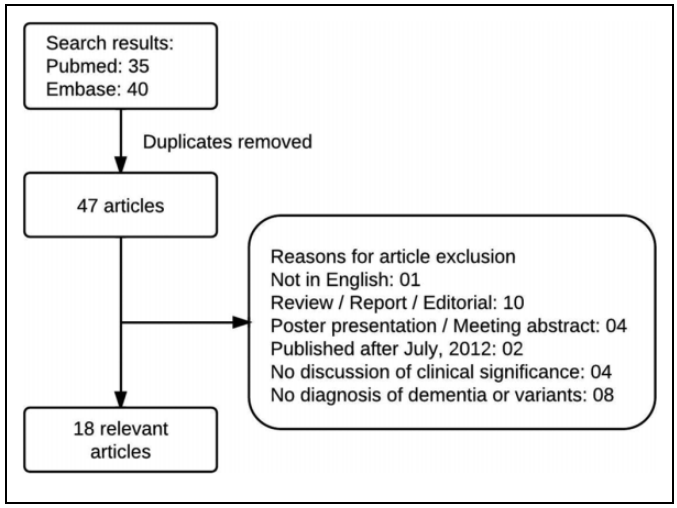

Figure 1. Study selection process.

to give meaning to the degree of change.[1] As such, these external criteria (or anchors) must have an appreciable association, both conceptually and statistically, with the instrument measure. Second, they must be interpretable as well as able to identify, to the greatest precision possible, individuals who have changed to a small but meaningful degree.[15] There are many ways in which anchor-based methods may be employed and can be broadly categorized as cross-sectional approaches and longitudinal approaches. In cross-sectional approaches, the instrument of interest is compared between groups that differ in some criteria, such as disease severity, and the difference is taken to establish the MCID. Ideally, the groups should be similar in every regard except for the differentiating criteria.

None of the studies retrieved in our systematic search employ a cross-sectional approach as defined earlier but several used expert opinion as anchors. Burback et al surveyed 161 geriatricians and neurologists about the “the smallest changes in the Folstein Mini-Mental State Examination (MMSE) scores that are compatible with a noticeable change in the patient’s overall condition.”(p 535) They then used the mean of the responses (3.72) to interpret the clinical significance of the results of published randomized controlled trials (RCTs) assessing the efficacy of tacrine in the treatment of AD.[16] On the other hand, Carpenter et al used a 1-point change in the Minimum Data Set (MDS) Activities of Daily Living Scale to denote a clinically meaningful change in the functional abilities and cost of care when examining change in physical function in nursing home residents with moderate to severe dementia.[17] The MCID was determined during the original development of the MDS-Resident Assessment Instrument.

Anchors derived from expert opinions (via conferences, surveys, committees, or other means) have several limitations. They are highly prone to error or bias[18] and do not directly incorporate patient or caregiver perspectives. Moreover, as Burback et al[16] point out, the MCID varies with the severity of dementia. As the question in their survey did not take that into account, the mean of the responses is unlikely to have been

an accurate representation of the MCID. Studies employing this method should specify the degree of illness in the survey but the obvious difficulty in taking the gradient of severity into account poses a serious limitation to the use of expert opinion alone in arriving at an estimate of the MCID and its generalizability to other studies.

Longitudinal approaches to anchor-based methods typically involve comparing the instrument of interest in a group that has changed in some meaningful way, such as disease severity. One commonly employed approach is to use global assessment scales. These are cumulative measures that take into account changes in cognitive, behavioral, and psychological functions in conjunction with caregiver reports to arrive at an overall numerical assessment of improvement or deterioration. Our systematic search retrieved 7 articles that used global assessment scales[10,11,19-23] of which 5 used Clinician's Interview-Based Impression of Change Plus Caregiver Input (CIBIC-Plus) secondary to GAS or ADS-Cog[10,11,19,21,22] for comparison and/or correlation. The CIBIC-Plus[24] is a widely used metric in drug trials and measures overall global change relative to baseline on a 7-point Likert-type scale ranging from 1 (*markedly improved*) to 7 (*markedly worse*). There is also systemic evaluation of patient's cognition, behavior, and function based on patient and caregiver interviews. The CIBIC-plus represents the clinician's perspective of change, and while patient or partner-administered global ratings are available and have been described,[25] they are not feasible in patient populations having mild or more severe AD. Global assessment scales are particularly useful in cases where changes in different domains are not individually significant but altogether amount to a readily observable difference. Conversely, it is difficult to relate the global assessment of change back to the individual domains and determine their MCID.

The ability of the participants to make decisions is particularly important for GAS. The GAS is a global outcome measure in which problem areas for each individual patient are identified and goals for improvement are set by the patient, caregiver, and/or the clinician. The goals can be medical, behavioral, or psychological in nature. Following some intervention, each patient is assessed at a previously defined follow-up time for improvement or deterioration. The goals are scored at a "0" while somewhat better outcomes and much better outcomes are scored at a +1 and a + 2, respectively. Similarly, somewhat worse outcomes and much worse outcomes are scored at a −1 and a −2, respectively. A standardized formula is used to sum achievement of weighted goals and come to a measure of overall goal attainment.[14] Since the goals represent the needs of the patient, GAS scores represent meaningful differences in the patients' functional abilities. However, individuals with cognitive impairment often have a limited understanding of their own capabilities and as such may be unable to set goals for themselves. In these cases, caregiver- and/or clinician-assigned goals are a good substitute. Additionally, GAS is concerned with significant, observable changes, not necessarily minimally significant changes, and as such may lead to larger measures than the MCID.

Our systematic review found 8 articles that used GAS in dementia research. Gordon et al assessed feasibility of GAS in long-term care of nursing home patients. Because 77% of the patients had dementia, the goals were set by 2 geriatricians in consultation with the nursing staff instead of the patients themselves.[26] Rockwood et al and Asp et al used a modified GAS approach by setting 2 groups of goals—one by the physicians and nurses and another by the patients and caregivers with the help of the field researchers.[10,11,19,21,22] They noticed that patients and caregivers were more likely to set goals related to behavior, leisure, and social interaction while physicians were more likely to set goals in domains that are testable, such as memory or cognition. This shows that patient collaboration invites a perspective that is important to the patients and not necessarily accounted for by their physicians.

Tractenberg et al describe an approach, not very dissimilar from the GAS, in which changes in behaviors of patients with AD were assessed by their physicians over a period of time. They used 37 items from the Behavior Rating Scale for Dementia and recorded clinically significant changes such as cessation, emergence, improvement, and intensification. An overall assessment of improvement or deterioration was then made by simply counting the number of behaviors that cease or show improvement as opposed to behaviors that emerge or show deterioration.[27]

Symptomology or accomplishments of goals are generally more applicable to patients and participants having later stages of dementia when observable behaviors and symptoms start manifesting themselves. Additionally, symptomology does not take into account the fact that many behaviors or symptoms in dementia are associated. So it is possible for 2 patients to show the same change overall even when one may have a certain number of changes in associated behaviors and the other has the same number of changes in unassociated behaviors even though clinically they are not the same and present a very different clinical picture.

In general, longitudinal approaches to anchor-based methods more directly measure change and are as such better suited for MCID measurements compared to cross-sectional approaches. However, none of the methods retrieved in our search took the patients' perspectives into account or the risk–benefit analysis necessary for MCID measurements. Although patient or caregiver input is highly desired in measures of clinical significance, it is often affected by recall bias, expectations, or comparison to counterparts and can therefore result in smaller assessments of change.[28] Patients may also be unable to report their condition due to a lack of insight, impaired memory, or simply an inability to communicate. These conditions are particularly relevant in moderate and late stages of AD and in this setting an exception has to be made in the original definition posited by Jaeschke and colleagues.[1] Similarly, caregivers are affected by many of the same biases and their reports may additionally be affected by their emotional state and their relationship with the patient.

Distribution-Based Methods

Several studies in our systematic review used distribution-based approaches in addition to expert opinions. Distribution-

based methods examine the relationship between the change in the instrument of interest and a measure of variability. This variation may be in the sample (effect size), change in outcome measures (standardized response mean), or the measurement precision of the instrument (standard error of the mean [SEM]). The effect size is the change in outcome measures for every pretest standard deviation (SD).[29,30] It is therefore dependent on the initial sample distribution with more heterogeneous cohorts producing a smaller effect size for any given outcome change. In contrast, standardized response mean is the change in outcome measures for every SD of that change.[31] It is therefore dependent on the variability in the changes with more variability in otherwise comparable changes resulting in a smaller standardized response mean. Finally, SEM is the change in outcome measures for every standard error of measurement.[32] This method relies on the assumption that measurement error is constant across the range of changes in outcome measures. But in practical application, the SEM is smaller at the extremes of the changes than it is in the middle.

Howard et al assessed the functional and cognitive effects of donepezil and memantine in community-living patients with Alzheimer's disease using MCIDs they had published a year earlier. They sought expert opinion from psychiatrists and geriatricians through e-mail discussions prior to a meeting and during face-to-face discussions. They were unable to reach agreement on a single value for the standardized MMSE (sMMSE) that would serve as the MCID but decided that it was between 1.0 point and 2.0 scale points. They also reviewed the SDs for both the baseline and the change from baseline scores on the sMMSE, Bristol Activities of Daily Living Scale (BADLS), and Neuropsychiatric Inventory (NPI). An SD of 0.4 in the sMMSE gave a value of 1.4 on the outcome measure which met the opinion-based range. They then used 0.4 SD of the change in score from baseline to give MCIDs of 3.5 points for the BADLS and 8.0 points for the NPI.[33,34]

The process whereby multiple values from different approaches are used to converge to a single value or a narrower range of values is known as triangulation. Anchor-based and distribution-based methods, while limited when used alone, can provide very useful estimates of MCID when used in conjunction with each other.[35] Howard et al used expert opinions to guide their selection of MCID for sMMSE but only extrapolated from their distribution estimate to arrive at the MCIDs for BADLS and NPI. The authors do point out that the MCIDs were very close to those chosen for the AD2000 study by expert opinion.[36] As discussed earlier, expert opinions lack the perspective of the patients and the consideration of risks and benefits that they undertake. As such, even with triangulation, measures of MCID do not fully conform to the original definition put forth by Jaeschke et al.

Van Iersel et al had a geriatrician, a neurologist, and a physiotherapist assess video recordings of elderly patients for clinically relevant change in gait, defined as change in the expected risk of falling, after 2 weeks of multidisciplinary treatment. They then carried out several measures of responsiveness at both group and individual levels.[37] Schrag et al also used half baseline SD and SEM in addition to clinical assessments of significant change to establish the MCID for the ADAS-Cog for patients with AD.[12]

Distribution-based approaches are purely statistical methods that allow us to express the change in a standardized metric, which in turn makes them easier to execute compared to anchor-based approaches and provides a means for comparison across different tests and samples.[15] They are independent of the sample size and therefore not bound by the limitations of methods based on statistical significance. However, distribution-based approaches by themselves do not provide information regarding the clinical relevance of the observed change. Interpretation of the resulting distribution-based indices is neither intuitive nor well defined. Researchers often use certain rules of thumb for some methods like the effect size and standardized response mean, where a magnitude of 0.2 is considered a small difference, a magnitude of 0.5 is considered a medium difference, and a magnitude of 0.8 is considered a large difference.[29] But even when these methods allow us to establish change as probably being clinically significant and meaningful to patients, there is little information as to whether the change is truly minimal. There is evidence in the literature that an effect size of 0.50 tends to equal the MCID arrived at by anchor-based determinations but that is not always the case.[15] As a result, distribution-based methods must always be qualified and given meaning with anchor-based estimates.

Discussion

Neely et al demonstrate that even small numerical differences in outcome measures can become statistically significant, given a large enough sample size.[38] Therefore, the MCID and the sample size calculations are particularly important in evaluating and interpreting the study results.[39] However, very few studies place due import to the clinical and practical significance of their results. An extensive review of quality of life (QOL) reporting standards in 82 RCTs using the European Organization for Research and Treatment of Cancer QOL questionnaire core 30 (EORTC QLQ-C30) found that clinical significance was only addressed in 38% of the articles. This percentage was only marginally higher (44%) for the 70% of articles that met the criteria for high-quality reporting. Most studies relied on statistical significance or lack of it to interpret QOL outcomes, while articles that addressed clinical significance relied on simple anchors like a >10-point change.[40] The RCTs in dementia and Alzheimer's research do not fare any better. A 2009 study of 4 systematic reviews on the comparative efficacy of antidepressants, corticosteroids, Alzheimer's drugs, and targeted immune modulators found that 42% of trials did not provide enough information to determine whether they were powered to assess MCID.[41] Another recent systematic review reported that 54% of the reviewed dementia drug RCTs did not mention clinical significance in their reports, and none of the studies employed measures of clinical significance that incorporated patient or caregiver considerations of risk and benefit of the drug's use.[13]

Measures of clinical significance have a particularly important role to play in dementia as change serves as an important index of disease progression. However, the assessment and quantification of change can often be quite challenging as dementia manifests as a variety of symptoms that present, stay the same, reoccur, and/or subside over the course of the illness. We have attempted to outline several approaches of determining and applying MCID measurements in dementia research. Their use remains relatively rare in published dementia literature and little consensus on best practices exists at present. Additional research is needed to define optimal approaches for measurements of clinical significance in dementia research.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article. Mr Shabbir has nothing to disclose. Dr Sanders receives loan repayment support from LRP/NIA; has received pilot funds from the Resnick Gerontology Center; funding for travel from the American Academy of Neurology, and the Albert Einstein College of Medicine; has reviewed for the NIH/NIA, the Center for Medicare and Medicaid Innovation (CMMI), the Patient-Centered Outcomes Research Institute (PCORI), and the Alzheimer's Association; has received honoraria for serving on peer-review panels from the CMMI and PCORI, and is a member of a federal advisory committee (MEDCAC).

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: Funding by the Medical Student Training in Aging Research (MSTAR) Program.

References

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415.
2. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol*. 2002;14(2): 109-114.
3. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171-184. doi:10.1586/erp.11.9.
4. Aalten P, de Vugt ME, Jaspers N, Jolles J, Verhey FRJ. The course of neuropsychiatric symptoms in dementia. Part I: findings from the two-year longitudinal maasbed study. *Int J Geriatr Psychiatry*. 2005;20(6):523-530. doi:10.1002/gps.1316.
5. Lee HB, Lyketsos CG. Depression in Alzheimer's disease: heterogeneity and related issues. *Biol Psychiatry*. 2003;54(3):353-362.
6. Levy ML, Cummings JL, Fairbanks LA, Bravi D, Calvani M, Carta A. Longitudinal assessment of symptoms of depression, agitation, and psychosis in 181 patients with Alzheimer's disease. *Am J Psychiatry*. 1996;153(11):1438-1443.
7. 2012 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2012;8(2):131-168. doi:10.1016/j.jalz.2012.02.001.
8. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;141(11):1356-1364.
9. CDER. *Peripheral and Central Nervous System Drugs Advisory Committee Meeting*. Rockville, MD: Food and Drug Administration; 1989:227.
10. Rockwood K, Fay S, Gorman M, Carver D, Graham JE. The clinical meaningfulness of ADAS-Cog changes in Alzheimer's disease patients treated with donepezil in an open-label trial. *BMC Neurol*. 2007;7:26. doi:10.1186/1471-2377-7-26.
11. Rockwood K, Fay S, Gorman M. The ADAS-cog and clinically meaningful change in the VISTA clinical trial of galantamine for Alzheimer's disease. *Int J Geriatr Psychiatry*. 2010;25(2): 191-201. doi:10.1002/gps.2319.
12. Schrag A, Schott JM. What is the clinically relevant change on the ADAS-Cog? *J Neurol Neurosurg Psychiatr*. 2012;83(2):171-173. doi:10.1136/jnnp-2011-300881.
13. Molnar FJ, Man-Son-Hing M, Fergusson D. Systematic review of measures of clinical significance employed in randomized controlled trials of drugs for dementia. *J Am Geriatr Soc*. 2009; 57(3):536-546. doi:10.1111/j.1532-5415.2008.02122.x.
14. Bouwens SFM, van Heugten CM, Verhey FRJ. Review of goal attainment scaling as a useful outcome measure in psychogeriatric patients with cognitive disorders. *Dement Geriatr Cogn Disord*. 2008;26(6):528-540. doi:10.1159/000178757.
15. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*. 2003;56(5):395-407.
16. Burback D, Molnar FJ, St John P, Man-Son-Hing M. Key methodological features of randomized controlled trials of Alzheimer's disease therapy. minimal clinically important difference, sample size and trial duration. *Dement Geriatr Cogn Disord*. 1999;10(6):534-540.
17. Carpenter GI, Hastie CL, Morris JN, Fries BE, Ankri J. Measuring change in activities of daily living in nursing home residents with moderate to severe cognitive impairment. *BMC Geriatr*. 2006;6:7. doi:10.1186/1471-2318-6-7.
18. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs*. 2003; 12(1):77-84. doi:10.1046/j.1365-2702.2003.00662.x.
19. Asp E, Cloutier F, Fay S, et al. Verbal repetition in patients with Alzheimer's disease who receive donepezil. *Int J Geriatr Psychiatry*. 2006;21(5):426-431. doi:10.1002/gps.1486.
20. Knopman DS, Knapp MJ, Gracon SI, Davis CS. The Clinician Interview-Based Impression (CIBI): a clinician's global change rating scale in Alzheimer's disease. *Neurology*. 1994;44(12): 2315-2321.
21. Rockwood K, Graham JE, Fay S. Goal setting and attainment in Alzheimer's disease patients treated with donepezil. *J Neurol Neurosurg Psychiatr*. 2002;73(5):500-507.
22. Rockwood K, Fay S, Song X, MacKnight C, Gorman M. Attainment of treatment goals by people with Alzheimer's disease receiving galantamine: a randomized controlled trial. *CMAJ*. 2006;174(8):1099-1105. doi:10.1503/cmaj.051432.
23. Schneider LS, Olin JT, Doody RS, et al. Validity and reliability of the Alzheimer's disease cooperative study-clinical global impression of change. the Alzheimer's disease cooperative study. *Alzheimer Dis Assoc Disord*. 1997;11 suppl 2:S22-S32.

24. Reisberg B, Schneider L, Doody R, et al. Clinical global measures of dementia. position paper from the international working group on harmonization of dementia drug guidelines. *Alzheimer Dis Assoc Disord*. 1997;11 suppl 3:8-18.
25. Schneider LS, Clark CM, Doody R, et al. ADCS Prevention Instrument Project: ADCS-clinicians' global impression of change scales (ADCS-CGIC), self-rated and study partner-rated versions. *Alzheimer Dis Assoc Disord*. 2006;20(4 suppl 3):S124-S138. doi:10.1097/01.wad.0000213878.47924.44.
26. Gordon JE, Powell C, Rockwood K. Goal attainment scaling as a measure of clinically important change in nursing-home patients. *Age Ageing*. 1999;28(3):275-281.
27. Tractenberg RE, Jin S, Patterson M, et al. Qualifying change: a method for defining clinically meaningful outcomes of change score computation. *J Am Geriatr Soc*. 2000;48(11):1478-1482.
28. Schwarz N, Sudman S, Brewer WF, et al. *Autobiographical Memory and the Validity of Retrospective Reports*. Springer-Verlag; 1994. http://deepblue.lib.umich.edu/handle/2027.42/64018. Accessed July 21, 2013.
29. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Mahwah, NJ: Psychology Press; 1988.
30. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. 1989;27(3 suppl):S178-S189.
31. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *J Clin Epidemiol*. 1995;48(11):1369-1378.
32. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol*. 1999;52(9):861-873.
33. Howard R, Phillips P, Johnson T, et al. Determining the minimum clinically important differences for outcomes in the DOMINO trial. *Int J Geriatr Psychiatry*. 2011;26(8):812-817. doi:10.1002/gps.2607.
34. Howard R, McShane R, Lindesay J, et al. Donepezil and memantine for moderate-to-severe Alzheimer's disease. *N Engl J Med*. 2012;366(10):893-903. doi:10.1056/NEJMoa1106668.
35. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102-109. doi:10.1016/j.jclinepi.2007.03.012.
36. Courtney C, Farrell D, Gray R, et al. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomised double-blind trial. *Lancet*. 2004;363(9427):2105-2115. doi:10.1016/S0140-6736(04)16499-4.
37. Van Iersel MB, Munneke M, Esselink RAJ, Benraad CEM, Olde Rikkert MGM. Gait velocity and the timed-up-and-go test were sensitive to changes in mobility in frail elderly patients. *J Clin Epidemiol*. 2008;61(2):186-191. doi:10.1016/j.jclinepi.2007.04.016.
38. Neely JG, Karni RJ, Engel SH, Fraley PL, Nussenbaum B, Paniello RC. Practical guides to understanding sample size and minimal clinically important difference (MCID). *Otolaryngol Head Neck Surg*. 2007;136(1):14-18. doi:10.1016/j.otohns.2006.11.001.
39. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol*. 1998;16(1):139-144.
40. Cocks K, King MT, Velikova G, Fayers PM, Brown JM. Quality, interpretation and presentation of European organisation for research and treatment of cancer quality of life questionnaire core 30 data in randomised controlled trials. *Eur J Cancer*. 2008;44(13):1793-1798. doi:10.1016/j.ejca.2008.05.008.
41. Gartlehner G, Thieda P, Hansen RA, Morgan LC, Shumate JA, Nissman DB. Inadequate reporting of trials compromises the applicability of systematic reviews. *Int J Technol Assess Health Care*. 2009;25(3):323-330. doi:10.1017/S0266462309990122.