Measuring clinically relevant change in apathy symptoms in ADMET and ADMET 2

Shankar Tumati,[1] Nathan Herrmann,[1] Jaime Perin,[2] Paul B. Rosenberg,[3] (iD)
Alan J. Lerner,[4] Jacobo Mintzer,[5] Prasad R. Padala,[6] (iD) Olga Brawman-Mintzer,[5]
Christopher H. van Dyck,[7] Anton P. Porsteinsson,[8] Suzanne Craft,[9] Allan Levey,[10]
David Shade,[2] and Krista L. Lanctôt[1,11,12] (iD)

[1] *Sunnybrook Research Institute, University of Toronto, Toronto, ON, Canada*
[2] *Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*
[3] *Johns Hopkins School of Medicine, Baltimore, MD, USA*
[4] *University Hospitals Cleveland Medical Center, Case Western Reserve University School of Medicine, Cleveland, OH, USA*
[5] *Ralph H. Johnson VA Medical Center, Medical University of South Carolina, Charleston, SC, USA*
[6] *Central Arkansas Veterans Healthcare System, University of Arkansas for Medical Sciences, Little Rock, AR, USA*
[7] *Yale School of Medicine, New Haven, CT, USA*
[8] *University of Rochester School of Medicine and Dentistry, Rochester, NY, USA*
[9] *Wake Forest University, Winston-Salem, NC, USA*
[10] *Emory Goizueta Alzheimer's Disease Research Center, Atlanta, GA, USA*
[11] *Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON, Canada*
[12] *Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada*

ABSTRACT

Objectives: Among participants with Alzheimer's disease (AD) we estimated the minimal clinically important difference (MCID) in apathy symptom severity on three scales.

Design: Retrospective anchor- and distribution-based analyses of change in apathy symptom scores.

Setting: Apathy in Dementia Methylphenidate Trial (ADMET) and ADMET 2 randomized controlled trials conducted at three and ten clinics specialized in dementia care in United States and Canada, respectively.

Participants: Two hundred and sixty participants (60 ADMET, 200 ADMET 2) with clinically significant apathy in Alzheimer's disease.

Measurements: The Clinical Global Impression of Change in Apathy scale was used as the anchor measure and the MCID on the Neuropsychiatric Inventory – Apathy (NPI-A), Dementia Apathy Interview and Rating (DAIR), and Apathy Evaluation Scale-Informant (AES-I) were estimated with linear mixed models across all study visits. The estimated thresholds were evaluated with performance metrics.

Results: Among the MCID was a decrease of four points (95% CI: −4.0 to −4.8) on the NPI-A, 0.56 points (95% CI: −0.47 to −0.65) on the DAIR, and three points on the AES-I (95% CI: −0.9 to −5.4). Distribution-based analyses were largely consistent with the anchor-based analyses. The MCID across the three measures showed ~60% accuracy. Sensitivity analyses found that MMSE scores and apathy severity at baseline influenced the estimated MCID.

Conclusions: MCIDs for apathy on three scales will help evaluate treatment efficacy at the individual level. However, the modest correspondence between MCID and clinical impression of change suggests the need to consider other scales.

Key words: Apathy, Alzheimer's disease, clinical assessment, behavioral and psychological symptoms of dementia (BPSD), rating scales

Introduction

Apathy, defined as a reduction in goal-directed behavior (Miller *et al.*, 2021), is the most common neuropsychiatric symptom in Alzheimer's disease (AD) (Zhao *et al.*, 2016). It is associated with worse

outcomes including impaired quality of life, faster disease progression, need for institutional care, and increased mortality, leading to increased caregiver burden and cost of care (Kruse *et al.*, 2023). As such, apathy is an important target for treatment (Mortby *et al.*, 2022). Apathy has been shown to respond to several treatments such as methylphenidate (Mintzer *et al.*, 2021; Padala *et al.*, 2018a; Rosenberg *et al.*, 2013), repetitive transcranial magnetic stimulation (Padala *et al.*, 2018b), and possibly cholinesterase inhibitors (Lopez *et al.*, 2008). While treatments for apathy have shown potential, there is currently little evidence to determine whether the observed change in apathy symptoms on rating scales is clinically meaningful.

Clinical trials for apathy assessed outcomes with a variety of scales such as the informant-rated apathy sub-scale of the Neuropsychiatric Inventory (NPI-A) (Cummings *et al.*, 1994), Apathy Evaluation Scale (AES-I) (Marin, *et al.*, 1991), or the clinician rated Dementia Apathy Interview and Rating scale (DAIR) (Strauss and Sperry, 2002), among others. These rating scales assess apathy symptoms (including loss of interest, loss of emotional expression, and difficulty initiating activities) by administering questionnaires to the clinician, care partner, or patient. These scales were developed to assess the severity of apathy symptoms but it is not known whether they are sensitive to changes in symptom severity, and if so, what threshold of a change score is clinically meaningful. The need for such thresholds was recognized for evaluating treatments for AD (Robert *et al.*, 2010) and have been developed for outcome measures such as composite scores of cognition assessed on neuropsychological batteries and for integrated multidimensional measures (Lansdall *et al.*, 2023; Papp *et al.*, 2020; Robert *et al.*, 2010; Schneider and Goldberg, 2020; Siemers *et al.*, 2016). Given the growing interest in developing treatments for apathy (Mortby *et al.*, 2022), there is a similar need to identify thresholds of meaningful within-person change on apathy scales to evaluate treatment response.

Thresholds for meaningful change on clinical outcome measures can be estimated with anchor-based or distribution-based methods. Anchor-based methods utilize an anchoring measure of within-person change in symptom severity compared to pretreatment level such as a rating of improvement (or deterioration) by a clinician or care partner or patient. Those ratings are used to establish a threshold for meaningful change on a target measure that also assesses the same symptom concept. In contrast, distribution-based methods use the variance (standard deviation, effect sizes, etc.) in scores among participants on a target measure to determine a threshold for minimum clinically important difference at a group level. The former approach is favored by regulatory bodies as the symptom improvement due to treatment is determined for each individual rather than at the group level (Food & Drug Administration, 2019). However, distribution-based methods can provide supporting evidence for a threshold derived by an anchor-based method.

The aim of this study was to determine the minimal clinically important difference (MCID) thresholds for commonly used apathy scales (NPI-A, DAIR, and AES-I) using data from two randomized trials for treating apathy in AD.

Methods

Participants

This study analyzed data from the Apathy in Dementia Methylphenidate Trial (ADMET) (NCT01117181) and ADMET 2 (NCT02346201), which were phase 2 and 3, multicenter, randomized, double-blind, placebo-controlled, parallel group studies of methylphenidate for the treatment of clinically significant apathy in patients with mild-to-moderate AD (Mintzer *et al.*, 2021; Rosenberg *et al.*, 2013). The studies were conducted between 2010 and 2011 (ADMET) and 2016–2020 (ADMET 2). Inclusion and exclusion criteria for both studies were identical and are published elsewhere (Drye *et al.*, 2013; Scherer *et al.*, 2018); key criteria were as follows: diagnosis of possible or probable AD based on the National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association (McKhann *et al.*, 1984); Mini-Mental State Exam (MMSE) score of 10–28; clinically significant apathy as evidenced by an NPI-A score of 4 or more. Participants were excluded if they had Major Depressive Episode as defined in the Diagnostic and Statistical Manual of Mental Disorders – IV (TR); had clinically significant agitation/aggression, delusions, or hallucinations on the NPI; recent changes to AD or antidepressant medications; use of trazodone >50 mg or lorazepam >0.5 mg for indications other than insomnia; failure to respond to past methylphenidate treatment for apathy; current or recent use of amphetamines, antipsychotics, bupropion, monoamine oxidase inhibitors, tricyclic antidepressants; need for acute psychiatric hospitalization or were suicidal (Drye *et al.*, 2013; Scherer *et al.*, 2018).

Intervention

Both ADMET and ADMET 2 consisted of a treatment and a placebo arm. Participants receiving methylphenidate were titrated to a target dose of 20 mg/day. The placebo arm received identical appearing and tasting over-encapsulated pills. Both groups also received a psychosocial intervention consisting of counseling, provision of education material on AD, its course, behavioral symptoms and treatment expectations, and 24-hours crisis support for the care partner. The intervention was given for 6 weeks (ADMET) or 6 months (ADMET 2). Both ADMET and ADMET 2 found methylphenidate to be a safe and efficacious treatment for apathy in AD (Mintzer *et al.*, 2021; Rosenberg *et al.*, 2013).

Measures

Clinical Global Impression of Change in Apathy symptoms (CGIC-A): The CGIC-A, administered in both studies, was used as the anchor measure. It is a 7-level ordinal scale where each level indicates a clinically meaningful change in symptoms compared to baseline and includes "Marked Improvement," "Moderate Improvement," "Minimal Improvement," "No Change," "Minimal Worsening," "Moderate Worsening," and "Marked Worsening." It is rated by a clinician. The CGIC-A was a co-primary outcome in the ADMET 2 study.

Neuropsychiatric Inventory – apathy subscale (NPI-A): The NPI consists of twelve items that assess behavioral symptoms in the preceding four weeks including apathy and is administered to the care partner (Cummings *et al.*, 1994). Each item is first rated for presence, and then for frequency on a 4-point scale ["Rarely," "Sometimes," "Often," "Very Often"] and severity on a 3-point scale ["Mild," "Moderate," "Severe"]. The frequency × severity score yields the symptom score [range 0–12] with higher scores indicating more apathy. The NPI-A was a co-primary outcome in the ADMET 2 study.

Dementia Apathy Interview and Rating (DAIR): The DAIR is a 16-item questionnaire that is administered to the care partner by a clinician. It assesses apathy symptoms on a 4-point scale [0–3] to rate how often a specific behavior was observed in the last month, and whether the behavior was changed from its pre-illness level. The final score is a mean calculated as the sum of only those item scores where a change in symptoms from their pre-illness level is indicated, divided by the number of included items [range 0–3] (Strauss and Sperry, 2002).

Apathy Evaluation Scale – Informant (AES-I): The AES-I is an 18-item questionnaire with each item rated on a 4-point scale by the care partner. Individual item scores are summed to yield a total score [range: 18–72] with higher scores indicating more apathy (Clarke *et al.*, 2007; Marin *et al.*, 1991). The AES-I was the primary outcome in the ADMET study.

Mini Mental State Examination (MMSE): The MMSE is a screening instrument administered by a clinician for rapid assessment of cognitive functions. It consists of 11 items summed for a total score ranging from 0 to 30. Higher scores indicate better cognition (Folstein *et al.*, 1975). Participants were categorized into "Mild" or "Moderate" AD if their MMSE score was 20 or more or between 10 and 19, respectively (Wattmo *et al.*, 2013).

Timepoints: In ADMET, the NPI-A was completed at baseline and at the 6 week endpoint. The AES-I was collected at baseline and at the 2-, 4-, and 6-week visits. In ADMET 2, the NPI-A and DAIR were administered at baseline and each month (months 1 to 6). In both studies, the CGIC-A was administered at each visit beginning with the first follow-up visit. For this study, all visits were analyzed together as the aim was to determine the change scores on the target measures corresponding to the clinical ratings. Visit-level data are provided in table S1 published as supplementary material online attached to the electronic version of this paper at https://www.cambridge.org/core/journals/international-psychogeriatrics

Statistical analyses

The CGIC-A served as the anchor measure and the NPI-A, DAIR, and AES-I were the target measures. Thresholds of meaningful within-person change in target measures were estimated by anchor-based analyses and supported by distribution-based analyses. Analyses were conducted for all timepoints with available data from both studies. Characteristics of measures of apathy were assessed by descriptive statistics (mean [standard deviation (SD)] or median values, proportions of categorical variables including floor and ceiling scores for target measures). Spearman correlations between CGIC-A ratings and change scores were assessed to determine appropriateness of anchor-based analyses, for which a correlation strength of ≥ 0.3 between the anchor and target measures is suggested (Lansdall *et al.*, 2023). To estimate thresholds of change, the following linear mixed model was used: change in target measure $\sim 1 + $ CGIC-A $+ (1 | $Subj$)$ to estimate the least square means (LS means) and [standard deviation (SD)] of change on the target measures for each change level of the CGIC-A using data from all visits. In addition, the mean and SD of change on the target measures at each visit for each rating level on the CGIC-A were determined.

Supporting the anchor-based estimates of within-person change on target measures, the empirical cumulative distribution function (eCDF) for each CGIC-A rating level across all visits was plotted. For levels that contained ≥ 10 participants, 95% confidence intervals for the eCDF were calculated based on the Kolmogorov–Smirnov's D distribution, which produces a continuous distribution band (Bickel and Doksum, 2015), and was implemented with the function "*ecdf.ksCI*" in the "*sfsmisc*" package in R. A greater separation between the eCDF for CGIC-A levels at a change score on a target measure suggests a better ability to indicate a clinically meaningful benefit at that score, which determined the within-person thresholds for improvement. As both ADMET and ADMET 2 showed positive results and provided a psychosocial intervention for the drug and placebo groups, thresholds were only determined for improvement.

For distribution-based analyses, the test-retest reliability of the target measures was first determined by the intra-class correlation coefficient (ICC) with a two way random model, single measurement and consistency definition between a target measure assessed at baseline and at the first follow-up visit among participants who were rated as "No Change" on the CGIC-A. A correlation of ≥ 0.7 is considered to provide adequate reliability. Distribution-based measures utilize the between-subject variance in a target measure to determine the minimum clinically important difference. These include half-SD [$SD * 0.5$], standard error of measurement (SEM) [$SD * \sqrt{1 - ICC}$], standard error of difference (S_{diff}) [$SEM * \sqrt{2}$], and reliable change index (RCI) [$1.96 * SEM\sqrt{2}$] (Charter, 1997; Hays *et al.*, 2005; Norman *et al.*, 2003; Wyrwich *et al.*, 1999). In progressive order, the obtained values provide cutoff values with increasing confidence of meaningful change. For instance, a half-SD change indicates a moderate effect size whereas a change score greater than the RCI has <5% of lying within the SEM. The distribution-based values are only intended to aid interpretation of the target measures.

The threshold scores determined from anchor-based methods provide within-person assessment of meaningful change. Performance of the thresholds for each scale were evaluated by defining "Responders" and "Non-responders" in ADMET and ADMET 2 and metrics including sensitivity (probability of positive test result among true positives), specificity (probability of negative test result among true negatives), accuracy (ratio of correct predictions [true positives and true negatives] to all cases examined), precision (ratio of number of true positives to all positive results), and F1 score (harmonic mean of precision and sensitivity or $\frac{2 * true\ positives}{2 * true\ positives + false\ positives + false\ negatives}$) were determined.

Sensitivity analyses were performed to assess whether the severity of AD or apathy influenced the estimation of clinically significant change scores. AD severity was based on MMSE scores as described above. Apathy severity was based on the NPI – apathy subitem assessing severity as "mild," "moderate," or "severe." As the severity score is multiplied with the frequency score, the change scores on the NPI-A were expected to differ by the level of apathy severity at baseline. The categorical variable for MMSE or apathy severity was included as an interaction term with the CGIC-A in linear mixed models as above, and the estimates of change scores were derived with subsets of the data with participants in each category for both variables. These analyses were done only in the larger ADMET 2 sample so that adequate data were available to assess interaction effects.

All analyses were carried out in R and significance on statistical tests was set to two-sided *p*-value of <0.05.

Results

Participants

The analyses included participants enrolled in ADMET and ADMET 2 whose characteristics are provided in Table 1. Participants in both studies were of similar age, had similar neuropsychiatric symptom burdens and levels of cognition as measured by the MMSE. Males comprised 38.3% of ADMET participants and 65.5% of ADMET 2 participants. In both studies, the majority of participants were on cholinesterase inhibitors; in ADMET, nearly two-thirds of participants were on memantine as compared to four in ten participants in ADMET 2.

Distribution of anchor and target measures

The distribution of ratings on the CGIC-A and the scores on the NPI-A, DAIR, and AES-I are shown in Table 2. On the CGIC-A in ADMET and ADMET 2 across follow-up visits, 54.4% and 45.3% were rated as having no change in symptoms, and 41.1% and 39.3% were rated as having at least minimal improvement at the endpoint visits. On the NPI-A in ADMET and ADMET 2 across follow-up visits ($v = 57$ and $v = 1084$), the change score showed improvement in 73.7% and 72.7%, no change in 21.1% and 19.2%, and worsening in 5.2% and 8.1%. These numbers on the DAIR ($v = 1051$) were 72.6%, 3.7%, and 23.7%, and on the AES-I ($v = 165$) were 50.3%, 6.7%, and 43.0%. The NPI-A showed floor and ceiling values in a maximum of 17.5% and 15.0% participants in ADMET and

Table 1. Characteristics of participants in the ADMET and ADMET 2 studies

	ADMET (N = 60)	ADMET 2 (N = 200)
Treatment (MPH)	29 (48.3%)	99 (49.5%)
Age (median)	78	76
Sex (male)	23 (38.3%)	131 (65.5%)
Education		
HS or less	24 (40.0%)	49 (24.5%)
Some College	13 (21.7%)	40 (20.0%)
Bachelor's degree	13 (21.7%)	58 (29.0%)
Grad/Professional	9 (15.0%)	52 (26.0%)
MMSE	20.1 (6.0)	18.9 (4.8)
Alzheimer's medications		
None	–	42 (21.0%)
ChEI	43 (71.7%)	145 (72.5%)
Memantine	37 (61.7%)	75 (37.5%)
Others^	–	3 (1.5%)
NPI-A	7.5 (2.3)	7.8 (2.4)
NPI-A Severity		
Mild	8 (13.3%)	24 (12.0%)
Moderate	43 (71.7%)	132 (66.3%)
Severe	9 (15.0%)	43 (21.6%)
AES-I	49.7 (11.7)	–
DAIR	–	1.92 (0.46)
NPI-total	16.5 (7.9)	16.4 (9.8)
Agitation/aggression	0.8 (1.3)	0.7 (1.3)
Anxiety	0.7 (1.7)	1.1 (2.2)
Depression	1.6 (2.5)	0.9 (1.4)
Irritability	1.1 (2.1)	1.2 (2.0)
Disinhibition	0.4 (0.9)	0.6 (1.5)
Aberrant motor behavior	1.2 (2.4)	1.2 (2.5)
Elation	0.1 (0.7)	0.2 (1.1)
Delusions	0.2 (0.6)	0.2 (0.7)
Hallucinations	0.1 (0.5)	0.1 (0.4)
Sleep disturbances	1.7 (2.9)	1.4 (2.8)
Appetite and eating disorder	1.0 (2.3)	1.2 (2.5)

Abbreviations: ADMET, Apathy in Dementia Methylphenidate Trial; AES-I, Apathy Evaluation Scale – Informant version; ChEI, Cholinesterase Inhibitor; DAIR, Dementia Apathy Interview and Rating; MMSE, Mini-Mental State Exam; MPH, methylphenidate; NPI, Neuropsychiatric Inventory; NPI-A, Neuropsychiatric Inventory Apathy subscale; SD, standard deviation.
^Combination pill of donepezil and memantine.

Table 2. Description of clinical outcome measures (MPH and placebo groups combined)

	ADMET (visits = 240)		ADMET 2 (visits = 1294)	
	NPI-A	AES-I	NPI-A	DAIR
Score range	0–12	18–72	0–12	0–3
Mean (SD)	5.79 (3.28)	49.7 (11.7)	4.75 (3.34)	1.52 (0.70)
Floor (%)	10 (4.2%)	0 (0%)	198 (14.9%)	35 (2.7%)
Ceiling (%)	11 (4.6%)	0 (0%)	73 (5.6%)	0 (0%)

Abbreviations: ADMET, Apathy in Dementia Methylphenidate Trial; AES-I, Apathy Evaluation Scale – Informant version; DAIR, Dementia Apathy Interview and Rating; NPI, Neuropsychiatric Inventory; SD, standard deviation.

first follow-up visit (week 2 for AES-I and week 6 for NPI-A in ADMET or Month 1 in ADMET 2) among participants rated as having "No Change" on the CGIC-A. In ADMET, the ICC for NPI-A was 0.32 (95% CI = −0.05 to 0.59; $n = 31$; $F_{(30,30)} = 1.93$, $p = 0.04$) and for AES-I was 0.88 (95% CI = 0.75 to 0.94; $n = 27$; $F_{(26,26)} = 15.1$, $p < 0.001$). In ADMET 2, the ICC for NPI-A was 0.31 (95% CI = 0.12 to 0.48; $n = 94$; $F_{(93,93)} = 1.91$, $p = 0.001$) and for DAIR was 0.48 (95% CI = 0.30 to 0.62; $n = 91$; $F_{(90,90)} = 2.81$, $p < 0.001$).

The responsiveness of target measures to assess change was determined by Spearman correlations between CGIC-A and the change score on target measures. The correlation between CGIC-A and (a) NPI-A (ADMET) was 0.24 ($p = 0.07$), (b) AES-I was 0.39 ($p < 0.001$), (c) NPI-A (ADMET 2) was 0.51 ($p < 0.001$), and (d) DAIR - 0.38 ($p < 0.001$). The thresholds for within-person change score were estimated for each level on the CGIC-A across all visits (Fig. 1, Supplementary table S2). For the NPI-A, the threshold change score for minimal improvement was −4.4 points (95% CI: −4.0 to −4.8, visits = 303) in ADMET 2 and −3.5 points (95% CI: −2.0 to −5.0, visits = 48) in ADMET. The corresponding score for DAIR was −0.56 points (95% CI: −0.47 to −0.65, visits = 303), and AES-I was −3.2 points (95% CI: −0.9 to −5.4, visits = 48).

Figure 2 shows the empirical cumulative distribution frequencies (eCDF) of target measures for each level of the CGIC-A across all visits along with confidence intervals bounds. A greater separation between the eCDF of CGIC-A level suggests better discrimination between clinical impressions of change at that corresponding change score on the target measure. Figure 2C shows that a change score
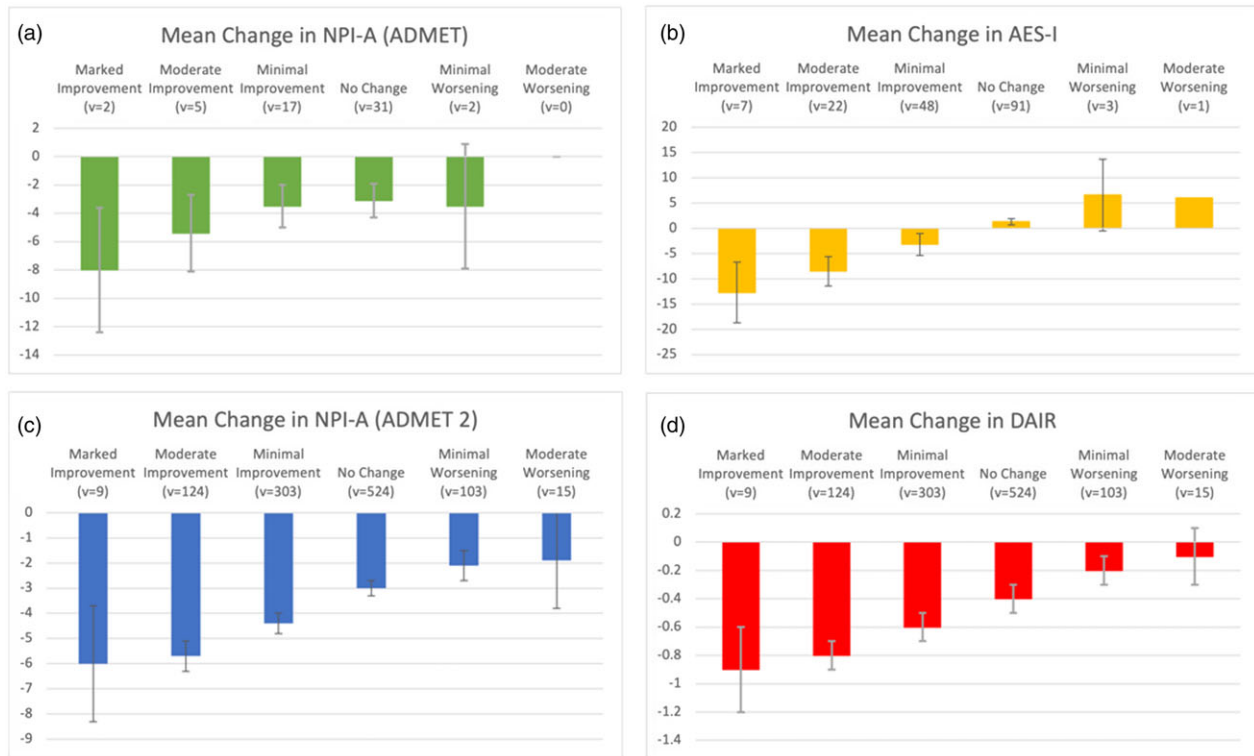
20.4% and 17.0% in ADMET 2. A maximum of 4.7% participants were rated at the lowest score on the DAIR, whereas none of the participants were rated at the lower or higher end of the score range on the AES-I. The visit level distributions of scores are shown in Supplementary table S1.

Thresholds for clinically meaningful change on the NPI-A, DAIR, and AES-I

The test-retest reliability of target measures was first ±*/ with ICC (2,1) with consistency definition, calculated between the scores at baseline and at the

Figure 1. Clinically meaningful within-person change scores on the NPI-A, AES-I, and DAIR. The figure shows the least squares estimated mean change score on the target measures at any visit (v) from baseline for each level of change on the CGIC-A. The estimates were derived by including scores from all visits in a linear mixed model with CGIC-A as the predictor and subjects as the random factor. Abbreviations: ADMET, Apathy in Dementia Methylphenidate Trial; DAIR, Dementia Apathy Interview and Rating; AES-I, Apathy Evaluation Scale – Informant; NPI-A, Neuropsychiatric Inventory – Apathy; SD, standard deviation.

(*x*-axis) of -5 points on the NPI-A was observed in 54% of ADMET 2 visits that were considered to show "Minimal improvement" and 30% of the visits where "No change" was indicated on the CGIC-A. Similarly, a 0.56 points improvement on the DAIR was found in 53 and 30% visits indicated as "Minimal improvement" and "No change," respectively (Fig. 2D).

Distribution-based analysis

The variance in target measures at baseline among participants was used to determine metrics of minimal clinically important change; these included the half-SD, SEM, Sdiff, and RCI. The estimated values for each metric for the NPI-A (ADMET) were – 1.2, 1.2, 1.8, and 3.5 points; for the AES-I – 6.0, 4.2, 5.9, and 11.6 points, and for the NPI-A with ADMET 2 study data were 1.2, 1.2, 1.8, and 3.6 points; and for the DAIR – 0.2, 0.2, 0.3, and 0.5 points.

Application of proposed thresholds

The performance of the estimated thresholds were evaluated in the empirical data from ADMET and ADMET 2. In the evaluation, the threshold scores derived from the model were rounded to the closest clinically obtainable scores on the scales – NPI-A: −4, DAIR: −0.56, AES-I: −3. The number of participants who improved or not as per the CGIC-A ratings and as per the estimated MCID on each target measure (same or larger change than the estimated threshold) are shown in Supplementary table S3. Figure 3 shows metrics of classification performance. The accuracy (or positive predictive value) and precision of the NPI-A (ADMET 2) and DAIR thresholds were found to be around 66% and ~60%.

Influence of disease and symptom severity on clinical significance thresholds

In sensitivity analyses, MMSE categories showed an interaction effect with CGIC-A levels with NPI-A but not DAIR as the dependent variable (Supplementary table S4). Minimal improvement on the CGIC-A was associated with larger change score (least square mean) among those with MMSE >= 20 compared to those with MMSE score of 10–19 (B = 0.95, t = −2.3, p = 0.02) (Supplementary table S4). The thresholds for each CGIC-A level on both scales are shown
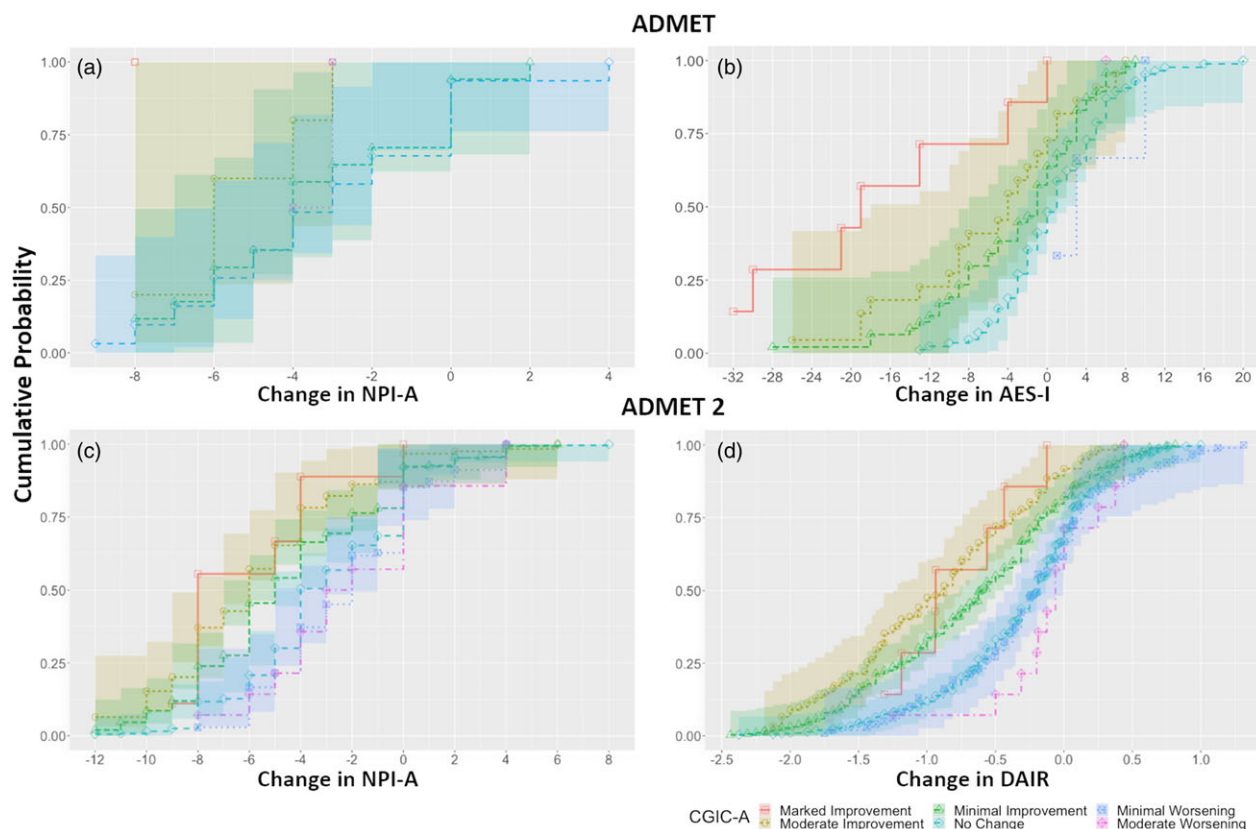
ADMET

ADMET 2

Figure 2. Empirical CDF of change on apathy scales as per clinically meaningful change levels. The figure shows cumulative proportion and 95% confidence interval bands of participants in each level of the CGIC-A across the range of change scores (*x*-axis). The range of change scores where the levels of CGIC-A show a greater separation indicates thresholds that are sensitive to detecting clinically meaningful change in symptoms. Abbreviations: AES-I, Apathy Evaluation Scale – Informant; CDF, cumulative distribution function; DAIR, Dementia Apathy Interview and Rating ; NPI-A, Neuropsychiatric Inventory – Apathy.
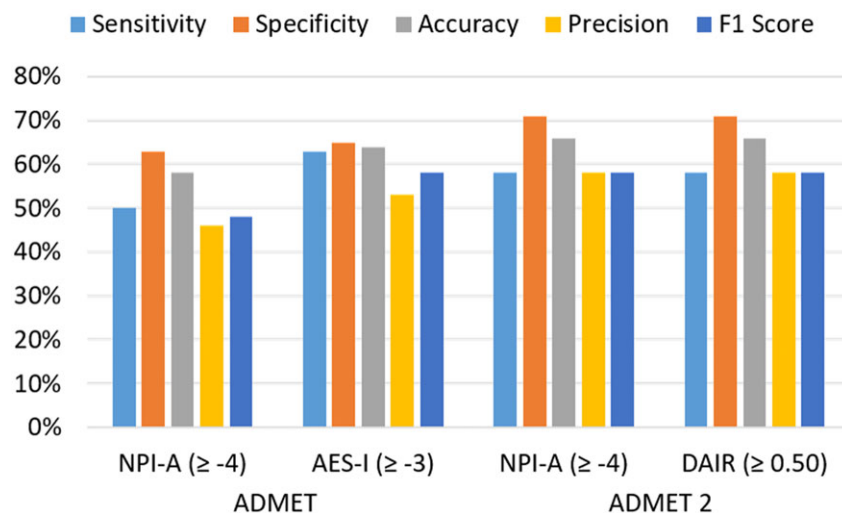
Figure 3. Performance metrics of proposed thresholds for target measures. The figure shows sensitivity (true positives identified among actual positives), specificity (true negatives identified among actual negatives), accuracy (true positives and true negatives among all observations), precision (positive predictive value; true positives among all positives predicted by threshold), and F1 score (equally weighs precision and sensitivity).

in Supplementary table S5 and Supplementary figure S1. The threshold score for minimal improvement among participants with MMSE >= 20 was −4.8 (NPI-A) and −0.66 (DAIR), and for MMSE of 10–19 was −3.6 (NPI-A) and −0.60 (DAIR). Similar analyses assessing NPI-A severity categories showed an interaction effect with CGIC-A for both NPI-A and DAIR (supplementary table S6). Compared to those with mild apathy, those with severe apathy showed a larger decrease than those with mild apathy for minimal improvement on the NPI-A (B = -1.59, t = −2.13, $p = 0.03$) and the DAIR (B = −4.22, t = −3.18, $p = 0.001$). On the DAIR, minimal worsening was also associated with a larger decrease in change scores among those with moderate (B = −0.33, t = −2.18, $p = 0.03$) and severe (B = −3.94, t = −2.35, $p = 0.02$) apathy compared to those with mild apathy. The estimated thresholds increased on both scales with an increase in baseline severity of apathy Supplementary table S7 and Supplementary figure S2).

Discussion

This study estimated clinically meaningful within-person change scores on three apathy scales with anchor- and distribution-based methods utilizing data from two multisite, randomized double-blind placebo controlled trials of methylphenidate for treating apathy in AD. Using anchor-based methods, the MCIDs were at least a 4 point decrease on the NPI-A, 0.5 point decrease on the DAIR, and 3 point decrease on the AES-I. These thresholds were supported by distribution-based methods that yielded similar values on the reliable change index. Additional sensitivity analyses suggest that baseline AD severity and apathy severity may affect the estimated thresholds. Specifically, minimal improvement on the CGIC-A among those with greater severity of AD or apathy was associated with a larger change score on both the NPI-A and DAIR than those with mild AD or mild apathy. While the proposed thresholds will help evaluate treatments for apathy using apathy scales and may improve design of clinical trials for apathy, the metrics evaluating the performance of the thresholds indicate a need to consider additional scales that may be better correlated with the anchor measure.

This study included three target measures of apathy that are rated by either a trained rater following an interview with the care partner (NPI-A, AES-I) or by a clinician following an interview with the care partner (DAIR), which provides thresholds for several target outcomes for future studies. The estimated thresholds were consistent between the two trials (NPI-A) and with distribution-based

measures (NPI-A and DAIR). These thresholds are expected to help assess within-person clinically meaningful improvement in apathy and evaluate efficacy of treatments (Assunção *et al.*, 2022; Webster *et al.*, 2017). As apathy is a predictor of several poor outcomes in people with AD, these thresholds may also aid assessment of apathy in trials where it is not a primary outcome. Neuropsychiatric symptoms are now recognized as a key feature of AD (Cummings, 2021), with a negative impact on affected persons and care partners (Naglie *et al.*, 2011). As disease-modifying therapies show clinical benefits, these thresholds can help determine whether burdensome symptoms like apathy also improve with treatment at the individual level.

The implementation of these thresholds in prospective studies may need to consider the severity of apathy and stage of the disease. The inclusion criteria for the two trials required the presence of clinically significant apathy (minimum NPI-A score of 4 points at baseline) that can be expected to affect the range of change scores that were included in the analysis. However, 94% of participants in ADMET 2 met the diagnostic criteria for apathy, suggesting that the thresholds were estimated in those with clinically relevant apathy (Lanctôt *et al.*, 2021). Similarly, inclusion criteria in both studies aimed to include participants with mild to moderate AD. Although MMSE is a relatively crude measure of AD severity, the MCID on the NPI-A differed among those stratified into higher and lower MMSE. Moreover, as the prodromal stages of AD were not included in both studies, the MCID in those individuals may differ. As the frequency and severity of apathy increases with disease progression, the distribution of change scores included in the analyses may be influenced by disease stage of the sample. For example, the MCID for participants with higher MMSE scores was greater than the MCID for the complete sample. In other words, for a change to be considered clinically relevant, a larger improvement is needed among individuals with better cognition than those with worse cognition. Finally, while the current study included change scores at 2 weeks until 6 months (only for the NPI-A), providing a wide time range for the change scores, the duration of the intervention needs to be considered before applying these thresholds (Mintzer *et al.*, 2021; Padala *et al.*, 2018a; Rosenberg *et al.*, 2013).

The results also show a limited correlation between within-person change in clinical impression and change on apathy rating scales. The recommended value for correlation between an anchor and a target measure is greater than or equal to 0.3 (Cohen, 1992; Hays *et al.*, 2005; Revicki *et al.*, 2008). In addition, the test-retest correlations of the scales

were relatively weak, especially for the NPI-A, suggesting that informants considered that apathy symptoms had changed among those participants who were not considered to have clinically meaningful change in symptoms by clinicians (who were blind to the ratings on the apathy scales). Although both scales (NPI-A and CGIC-A) showed group-level improvement in ADMET 2, the weak correlations indicate a discrepancy between informant and clinician rating. Considering that both types of correlations obtained in this study were on the lower end and that the metrics for performance of the thresholds ranged from 60 to 65%, the apathy rating scales included in this study may be limited in assessing within-person change in apathy symptoms. The NPI-A in particular also had substantial proportion of values at lower or upper end of the scoring range, suggesting that the scoring range may limit its ability to detect change in apathy symptoms. Continuous measures such as the DAIR may provide a better estimate of change in symptoms (as suggested by Fig. 2D); however, the performance metrics of the DAIR were similar to the NPI-A (Figure 3) and no treatment effect was detected with the DAIR in ADMET 2, unlike the NPI-A. Finally, while these rating scales are widely used, they do not account for the multidimensional nature of apathy (Le Heron *et al.*, 2018). The new consensus-based diagnostic criteria for apathy in neurocognitive disorders require that diminished function be present in at least two of the three dimensions of apathy (interest, initiative, and emotional expression/response) (Miller *et al.*, 2021). Thus, other rating scales such as the 12-item apathy sub-scale NPI-Clinician (de Medeiros *et al.*, 2010) may better assess apathy symptoms in line with the diagnostic criteria and may need to be considered to measure within-person change in apathy symptoms.

As noted above, the limitations of this study include, first, the design factors of the parent studies including participant inclusion criteria and study duration. However, those studies included participants with clinically significant apathy who were assessed at multiple visits over a maximum of six months, providing a broad time duration to assess within-person change in apathy. Second, the assessments were performed across several sites by various raters, which may affect the estimated thresholds as inter-rater reliability could not be assessed; however, on the other hand, the thresholds may be generalizable to real-world use. Analysis to assess inter-rater reliability was not possible due to lack of data. However, the large cohort that was collected from multiple sites and representative of patient populations is a strength of this study. Third, the replication of results was only possible for the NPI-A, which alone was measured in both trials.

Fourth, the estimates for the AES-I and for the following CGIC-A levels "Moderate worsening," "Minimal worsening," and "Marked improvement" were derived from a limited number of participants. As such, those estimates must be interpreted with caution. Fifth, while anchor-based methods for assessing meaningful change are favored, the CGIC-A may itself be influenced by the subjectivity of clinical raters; care partners who spend more time with the person may have better insight to identify symptomatic changes and meaningful improvement at the individual level. Such a measure was not available for this study and should be considered in future trials.

Conclusions

The current study estimated threshold change scores for within-person meaningful change on the NPI-A, DAIR, and AES-I using anchor- and distribution-based measures. The estimated thresholds were consistent across studies and methods and will allow for assessing the benefit of interventions for individual participants in clinical trials for apathy in AD. Given the recently published consensus criteria for apathy in AD (Miller *et al.*, 2021), symptoms from the revised criteria could be better "mapped" to apathy scales. This has been recently done for agitation in AD resulting in scales that are more sensitive to change (De Mauleon *et al.*, 2021). Improved scales may lead to better clinical trial design and thus help us develop and validate improved treatments for apathy in AD.

Conflicts of interest

Dr Mintzer reported being an advisor for Praxis Bioresearch and Cerevel Therapeutics outside the submitted work. Dr Lanctôt reported grants from the National Institutes of Health during the conduct of the ADMET 2 study, the Bernick Chair in Geriatric Psychopharmacology from Sunnybrook Research Institute and the Universisty of Toronto, and personal fees for serving on the advisory boards of BioXcel Therapeutics, Cerevel Therapeutics, Eisai, Exciva, Kondor Pharma, Lundbeck Otsuka, Novo Nordisk, and Sumitomo outside the submitted work. Dr Scherer reported grants from Johns Hopkins University during the conduct of the ADMET 2 study. Dr Rosenberg has received research grants from the National Institutes of Aging, Alzheimer's Clinical Trials Consortium, Richman Family Precision Medicine Center of Excellence on Alzheimer's Disease, Eisai, Functional Neuromodulation, and Lilly; honoraria from

Funding

Description of author(s)′ roles

ST and KLL conceived the study; NH, JP, PBR, AJL, JM, PRP, OBM, CHvD, APP, SC, AL, DS, and KLL acquired the data; ST analyzed the data; ST and KLL drafted the manuscript; NH, JP, PBR, AJL, JM, PRP, OBM, CHvD, APP, SC, AL, DS, and KLL were major contributors to critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements

Availability of data and materials

The data that support the findings of this study are available from the authors but restrictions apply to the availability of these data. The data are available from the corresponding author on request after approval of a proposal with a signed data sharing agreement.

Ethics approval and consent to participate

The ADMET and ADMET 2 studies' participants were recruited at 3 (2 US clinics and 1 Canadian clinic) and 10 (9 US clinics and 1 Canadian clinic) centers, respectively, specializing in dementia care where they or their legally authorized representatives and the primary caregiver for the participant provided informed consent. The study adhered to the Declaration of Helsinki and was approved by the ethical review boards of each site.

Consent for publication

Not applicable.

Supplementary material

For supplementary material accompanying this paper visit https://doi.org/10.1017/S1041610224000711

References

Assunção, S. S., Sperling, R. A., Ritchie, C., Kerwin, D. R., Aisen, P. S., Lansdall, C., Atri, A., Cummings, J. (2022). Meaningful benefits: a framework to assess disease-modifying therapies in preclinical and early Alzheimers disease. *Alzheimer's Research & Therapy*, 14(1), 54. https://doi.org/10.1186/s13195-022-00984-y

Bickel, P. J., & Doksum, K. A. (2015). Inference in semiparametric models. In *Mathematical statistics: Basic ideas and selected topics*. CRC Press (Chapman & Hall/CRC

Texts in Statistical Science, v. 2). https://books.google.ca/books?id=MgXeDwAAQBAJ

Charter, R. A. (1997). Effect on measurement error on tests of statistical significance. *Journal of Clinical and Experimental Neuropsychology*, 19(3), 458–462. https://doi.org/10.1080/01688639708403872

Clarke, D. E., Rvan Reekum, Ph., Simard, M., Streiner, D. L., & Freedman, M. (2007). Apathy in dementia: an examination of the psychometric properties of the apathy evaluation scale. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 19(1), 57–64. https://doi.org/10.1176/jnp.2007.19.1.57

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Cummings, J. (2021). The role of neuropsychiatric symptoms in research diagnostic criteria for neurodegenerative diseases. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, 29(4), 375–383. https://doi.org/10.1016/j.jagp.2020.07.011

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., Gornbein, J. (1994). The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology*, 44(12), 2308–2314. https://doi.org/10.1212/wnl.44.12.2308

De Mauleon, A., Ismail, Z., Rosenberg, P., Miller, D., Cantet, C., O'Gorman, C., Vellas, B., Lyketsos, C., Soto, M. (2021). Agitation in Alzheimers disease: novel outcome measures reflecting the International Psychogeriatric Association (IPA) agitation criteria. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 17(10), 1687–1697. https://doi.org/10.1002/alz.12335

Drye, L. T., Scherer, R. W., Lanctôt, K. L., Rosenberg, P. B., Herrmann, N., Bachman, D., Mintzer, J. E. (2013). Designing a trial to evaluate potential treatments for apathy in dementia: the apathy in dementia methylphenidate trial (ADMET). *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, 21(6), 549–559. https://doi.org/10.1016/j.jagp.2012.12.018

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Food & Drug Administration. Incorporating clinical outcome assessments into endpoints for regulatory decision-making 2019. https://www.fda.gov/media/132505/download.

Hays, R. D., Brodsky, M., Johnston, M. F., Spritzer, K. L., Hui, K.-K. (2005). Evaluating the statistical significance of health-related quality-of-life change in individual patients. *Evaluation & the Health Professions*, 28(2), 160–171. https://doi.org/10.1177/0163278705275339

Kruse, C., Maier, F., Spottke, A., Bach, J-Philipp, Bartels, C., Buerger, K., Fellgiebel, A., Fliessbach, K., Frölich, L., Hausner, L., Hellmich, M., Klöppel, S., Klostermann, A., Kornhuber, J., Laske, C., Peters, O., Priller, J., Richter-Schmidinger, T., Schneider, A.,

Shah-Hosseini, K., Teipel, S., von Arnim, C. A. F., Wiltfang, J., van der Wurp, H., Dodel, R., Jessen, F. (2023). Apathy in patients with Alzheimers disease is a cost-driving factor. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 19(7), 2853–2864. https://doi.org/10.1002/alz.12915

Lanctôt, K. L., Scherer, R. W., Li, A., Vieira, D., Coulibaly, H., Rosenberg, P. B., Herrmann, N., Lerner, A. J., Padala, P. R., Brawman-Mintzer, O., van Dyck, C. H., Porsteinsson, A. P., Craft, S., Levey, A., Burke, W. J., Mintzer, J. E. (2021). Measuring apathy in Alzheimers disease in the apathy in dementia methylphenidate trial 2 (ADMET 2): a comparison of instruments. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, 29(1), 81–89. https://doi.org/10.1016/j.jagp.2020.05.020

Lansdall, C. J., McDougall, F., Butler, L. M., Delmar, P., Pross, N., Qin, S., McLeod, L., Zhou, X., Kerchner, G. A., Doody, R. S. (2023). Establishing clinically meaningful change on outcome assessments frequently used in trials of mild cognitive impairment due to Alzheimers disease. *The Journal of Prevention of Alzheimer's Disease*, 10(1), 9–18. https://doi.org/10.14283/jpad.2022.102

Le Heron, C., Apps, M.a J., & Husain, M. (2018). The anatomy of apathy: a neurocognitive framework for amotivated behaviour. *Neuropsychologia*, 118, 54–67. https://doi.org/10.1016/j.neuropsychologia.2017.07.003

Lopez, O. L., Mackell, J. A., Sun, Y., Kassalow, L. M., Xu, Y., McRae, T., Li, H. (2008). Effectiveness and safety of donepezil in Hispanic patients with Alzheimers disease: a 12-week open-label study. *Journal of the National Medical Association*, 100(11), 1350–1358. https://doi.org/10.1016/s0027-9684(15)31515-7

Marin, R. S., Biedrzycki, R. C., & Firinciogullari, S. (1991). Reliability and validity of the Apathy Evaluation Scale. *Psychiatry Research*, 38(2), 143–162. https://doi.org/10.1016/0165-1781(91)90040-v

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E. M. (1984). Clinical diagnosis of Alzheimers disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7), 939–944. https://doi.org/10.1212/wnl.34.7.939

de Medeiros, K., Robert, P., Gauthier, S., Stella, F., Politis, A., Leoutsakos, J., Taragano, F., Kremer, J., Brugnolo, A., Porsteinsson, A. P., Geda, Y. E., Brodaty, H., Gazdag, G., Cummings, J., Lyketsos, C. (2010). The Neuropsychiatric Inventory-Clinician rating scale (NPI-C): reliability and validity of a revised assessment of neuropsychiatric symptoms in dementia. *International Psychogeriatrics*, 22(6), 984–994. https://doi.org/10.1017/S1041610210000876

Miller, D. S., Robert, P., Ereshefsky, L., Adler, L., Bateman, D., Cummings, J., DeKosky, S. T., Fischer, C. E., Husain, M., Ismail, Z., Jaeger, J., Lerner, A. J., Li, A., Lyketsos, C. G., Manera, V., Mintzer, J., Moebius, H. J., Mortby, M., Meulien, D., Pollentier, S., Porsteinsson, A., Rasmussen, J., Rosenberg, P. B., Ruthirakuhan, M. T., Sano, M., Zucchero Sarracini,

C., Lanctôt, K. L. (2021). Diagnostic criteria for apathy in neurocognitive disorders. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 17(12), 1892–1904. https://doi.org/10.1002/alz.12358

Mintzer, J., Lanctôt, K. L., Scherer, R. W., Rosenberg, P. B., Herrmann, N., van Dyck, C. H., Padala, P. R., Brawman-Mintzer, O., Porsteinsson, A. P., Lerner, A. J., Craft, S., Levey, A. I., Burke, W., Perin, J., Shade, D., Michalak, H., Ni, G., Good, S., Mecca, A., Salem-Spencer, S., Keltz, M., Morales, E., Clark, E. D., Williams, A., Kindy, A., Freeman, R., Jamil, N., Schultz, M., Sami, S., Padala, K. P., Parkes, C., Lah, J., Vaughn, P., Hales, C., Rapoport, M., Gallagher, D., Li, A., Sandra, B., Vieira, D., Ruthirakuhan, M., Babani, P. (2021). Effect of methylphenidate on apathy in patients with Alzheimer disease: the ADMET 2 randomized clinical trial. *JAMA Neurology*, 78(11), 1324–1332. https://doi.org/10.1001/jamaneurol.2021.3356

Mortby, M. E., Adler, L., Agüera-Ortiz, L., Bateman, D. R., Brodaty, H., Cantillon, M., Geda, Y. E., Ismail, Z., Lanctôt, K. L., Marshall, G. A., Padala, P. R., Politis, A., Rosenberg, P. B., Siarkos, K., Sultzer, D. L., Theleritis, C. (2022). Apathy as a treatment target in Alzheimers disease: implications for clinical trials. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, 30(2), 119–147. https://doi.org/10.1016/j.jagp.2021.06.016

Naglie, G., Hogan, D. B., Krahn, M., Black, S. E., Beattie, B. L., Patterson, C., MacKnight, C., Freedman, M., Borrie, M., Byszewski, A., Bergman, H., Streiner, D., Irvine, J., Ritvo, P., Comrie, J., Kowgier, M., Tomlinson, G. (2011). Predictors of family caregiver ratings of patient quality of life in Alzheimer disease: cross-sectional results from the Canadian Alzheimers Disease Quality of Life Study. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, 19(10), 891–901. https://doi.org/10.1097/JGP.0b013e3182006a7f

Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care*, 41(5), 582–592. https://doi.org/10.1097/01.MLR.0000062554.74615.4C

Padala, P. R., Padala, K. P., Lensing, S. Y., Ramirez, D., Monga, V., Bopp, M. M., Roberson, P. K., Dennis, R. A., Petty, F., Sullivan, D. H., Burke, W. J. (2018a). Methylphenidate for apathy in community-dwelling older veterans with mild Alzheimers disease: a double-blind, randomized, placebo-controlled trial. *The American Journal of Psychiatry*, 175(2), 159–168. https://doi.org/10.1176/appi.ajp.2017.17030316

Padala, P. R., Padala, K. P., Lensing, S. Y., Jackson, A. N., Hunter, C. R., Parkes, C. M., Dennis, R. A., Bopp, M. M., Caceda, R., Mennemeier, M. S., Roberson, P. K., Sullivan, D. H. (2018b). Repetitive transcranial magnetic stimulation for apathy in mild cognitive impairment: a double-blind, randomized, sham-controlled, cross-over pilot study. *Psychiatry Research*, 261, 312–318. https://doi.org/10.1016/j.psychres.2017.12.063

Papp, K. V., Buckley, R., Mormino, E., Maruff, P., Villemagne, V. L., Masters, C. L., Johnson, K. A., Rentz, D. M., Sperling, R. A., Amariglio, R. E., on behalf of Collaborators from the Harvard Aging Brain Study, the Alzheimer's Disease Neuroimaging Initiative and the Australian Imaging, Biomarker and Lifestyle Study of Aging (2020). Clinical meaningfulness of subtle cognitive decline on longitudinal testing in preclinical AD. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 16(3), 552–560. https://doi.org/10.1016/j.jalz.2019.09.074

Revicki, D., Hays, R. D., Cella, D., Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102–109. https://doi.org/10.1016/j.jclinepi.2007.03.012

Robert, P., Ferris, S., Gauthier, S., Ihl, R., Winblad, B., Tennigkeit, F. (2010). Review of Alzheimers disease scales: is there a need for a new multi-domain scale for therapy evaluation in medical practice? *Alzheimer's Research & Therapy*, 2(4), 24. https://doi.org/10.1186/alzrt48

Rosenberg, P. B., Lanctôt, K. L., Drye, L. T., Herrmann, N., Scherer, R. W., Bachman, D. L., Mintzer, J. E., ADMET investigators (2013). Safety and efficacy of methylphenidate for apathy in Alzheimers disease: a randomized, placebo-controlled trial. *The Journal of Clinical Psychiatry*, 74(8), 810–816. https://doi.org/10.4088/JCP.12m08099

Scherer, R. W., Drye, L., Mintzer, J., Lanctôt, K., Rosenberg, P., Herrmann, N., Padala, P., Brawman-Mintzer, O., Burke, W., Craft, S., Lerner, A. J., Levey, A., Porsteinsson, A., van Dyck, C. H. (2018). The Apathy in Dementia Methylphenidate Trial 2 (ADMET 2): study protocol for a randomized controlled trial. *Trials*, 19(1), 46. https://doi.org/10.1186/s13063-017-2406-5

Schneider, L. S., & Goldberg, T. E. (2020). Composite cognitive and functional measures for early stage Alzheimers disease trials. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, 12(1), e12017. https://doi.org/10.1002/dad2.12017

Siemers, E., Holdridge, K. C., Sundell, K. L., Liu-Seifert, H. (2016). Function and clinical meaningfulness of treatments for mild Alzheimers disease. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, 2(1), 105–112. https://doi.org/10.1016/j.dadm.2016.02.006

Strauss, M. E., & Sperry, S. D. (2002). An informant-based assessment of apathy in Alzheimer disease. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 15(3), 176–183.

Wattmo, C., Wallin, A. K., & Minthon, L. (2013). Progression of mild Alzheimers disease: knowledge and prediction models required for future treatment strategies. *Alzheimer's Research & Therapy*, 5(5), 44. https://doi.org/10.1186/alzrt210

Webster, L., Groskreutz, D., Grinbergs-Saull, A., Howard, R., O'Brien, J. T., Mountain, G., Banerjee, S., Woods, B., Perneczky, R., Lafortune, L., Roberts, C., McCleery, J., Pickett, J., Bunn, F., Challis, D., Charlesworth, G., Featherstone, K., Fox, C., Goodman, C., Jones, R., Lamb, S., Moniz-Cook, E., Schneider, J., Shepperd, S., Surr, C., Thompson-Coon, J., Ballard, C., Brayne, C., Burns, A., Clare, L., Garrard, P., Kehoe, P., Passmore, P.,

Holmes, C., Maidment, I., Robinson, L., Livingston, G., Quinn, T. J. (2017). Core outcome measures for interventions to prevent or slow the progress of dementia for people living with mild to moderate dementia: systematic review and consensus recommendations. *PloS One*, 12(6), e0179521. https://doi.org/10.1371/journal.pone.0179521

Wyrwich, K. W., Tierney, W. M., & Wolinsky, F. D. (1999). Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *Journal of Clinical Epidemiology*, 52(9), 861–873. https://doi.org/10.1016/s0895-4356(99)00071-2

Zhao, Q.-F., Tan, L., Wang, H.-F., Jiang, T., Tan, M.-S., Tan, L., Xu, W., Li, J.-Q., Wang, J., Lai, T.-J., Yu, J.-T. (2016). The prevalence of neuropsychiatric symptoms in Alzheimers disease: systematic review and meta-analysis. *Journal of Affective Disorders*, 190, 264–271. https://doi.org/10.1016/j.jad.2015.09.069