

RESEARCH ARTICLE

Agitation in Alzheimer's disease: Novel outcome measures reflecting the International Psychogeriatric Association (IPA) agitation criteria

Adelaide De Mauleon¹ | Zahinoor Ismail² | Paul Rosenberg³ | David Miller⁴ |
Christelle Cantet^{1,5} | Cedric O'Gorman⁶ | Bruno Vellas¹ | Constantine Lyketsos³ |
Maria Soto¹

¹ Gerontopole Alzheimer Clinical and Research Center, University Hospital of Toulouse (CHU Toulouse), URM 1295, CERPOP, Inserm, UPS, Université de Toulouse, Toulouse, France

² Psychiatry, Clinical Neurosciences, and Community Health Sciences, Hotchkiss Brain Institute and O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, Canada

³ Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

⁴ Signant Health, Blue Bell, Pennsylvania, USA

⁵ USMR, Epidemiology and Public Health Department, University Hospital of Toulouse (CHU Toulouse), Toulouse, France

⁶ Axsome Therapeutics, New York, New York, USA

Correspondence

Adelaide De Mauleon, Gerontopole Alzheimer Clinical & Research Center, Department of Geriatric Medicine, University Hospital of Toulouse, 224, avenue de Casselardit, Toulouse, Cedex 9 31059, France.
E-mail: demauleon.a@chu-toulouse.fr

Adelaide De Mauleon and Zahinoor Ismail are co-first authors.

Constantine Lyketsos and Maria Soto co-senior authors.

1 | INTRODUCTION

Agitation frequently affects patients with Alzheimer's disease (AD) and is among the most disruptive clinical features of the disease. Agitation is associated with highly impactful negative outcomes affecting patients, caregivers, and health systems.^{1,2} Given its impact, AD

agitation is an important target for effective treatment development. Although progress has been made in implementing randomized clinical trials (RCTs) for agitation,³ the choice of optimal outcome measures to demonstrate treatment effectiveness is still unresolved.⁴

Two approaches have been used to assess treatment response in RCTs targeting agitation in AD: (1) global ratings based on the judgment

Abstract

Introduction: The 2017 European Union-North American Clinical Trials in Alzheimer's Disease Task Force recommended development of clinician-rated primary outcome measures for Alzheimer's disease (AD) agitation trials, incorporating International Psychogeriatric Association (IPA) criteria.

Methods: In a modified Delphi process, Cohen-Mansfield Agitation Inventory (CMAI) and Neuropsychiatric Inventory-Clinician (NPI-C) items were mapped to IPA agitation domains generating novel instruments, CMAI-IPA and NPI-C-IPA. Validation in the Agitation and Aggression AD Cohort (A3C) assessed minimal clinically important differences (MCIDs), change sensitivity, and predictive validity.

Results: MCID was -17 (odds ratio [OR] = 14.9, 95% confidence interval [CI] = 6.8-32.6) for CMAI; -5 (OR = 9.3, 95% CI = 4.0-21.2) for CMAI-IPA; -3 (OR = 11.9, 95% CI = 4.1-34.8) for NPI-C-A+A; and -5 (OR = 7.8, 95% CI = 3.4-17.9) for NPI-C-IPA at 3 months. Areas under the curve suggested no scale better predicted global clinician ratings. Sensitivity to change for all measures was high.

Conclusion: Internal consistency and reliability analyses demonstrated better accuracy for the NPI-C-IPA than for the CMAI-IPA and can be used for agitation clinical trial inclusion, and for response to intervention.

KEYWORDS

aggression, agitation, Alzheimer's disease, dementia, efficacy, gold standard, measure, neuropsychiatric symptoms, outcome, research, scales, trials, validation

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association

HIGHLIGHTS

- Optimal clinical rating outcome measures to evaluate treatment effectiveness for agitation in Alzheimer's disease (AD) clinical trials is lacking and no International Psychogeriatric Association (IPA)-specific measures exist.
- Using a modified Delphi process, Cohen-Mansfield Agitation Inventory (CMAI)-IPA and Neuropsychiatric Inventory-Clinician (NPI-C)-IPA were derived, to better reflect the IPA agitation criteria.
- In the Agitation and Aggression AD Cohort (A3C) study, all four measures demonstrated high sensitivity to change and the ability to predict clinician ratings; NPI-C-IPA showed a better accuracy over CMAI-IPA.
- The novel measures CMAI-IPA and NPI-C-IPA, specially NPI-C-IPA, were more efficient at measuring agitation than CMAI and NPI-C-A+A.

RESEARCH IN CONTEXT

1. Systematic review: This research responds to the European Union-North American Clinical Trials in Alzheimer's Disease Task Force recommendation to develop an optimized, clinician-rated, primary outcome measure for clinical trials targeting agitation in Alzheimer's disease (AD), consistent with the International Psychogeriatric Association (IPA) agitation criteria. A literature search discovered no such validated measures, supporting the need to develop an IPA agitation-specific measure.
2. Interpretation: A modified Delphi process was implemented to derive IPA domain-specific measures from the Cohen-Mansfield Agitation Inventory (CMAI) and the Neuropsychiatric Inventory-Clinician scale (NPI-C). Original and novel measures were validated in the Agitation and Aggression AD Cohort (A3C). The IPA-informed novel scales, the CMAI-IPA and NPI-C-IPA, performed as well as the original scales in a naturalistic study of AD patients.
3. Future directions: As these novel measures accurately represent the IPA Agitation Syndrome, future studies incorporating IPA Agitation Criteria can use the novel scales for participant inclusion, and as outcome measures.

of experienced clinicians; and (2) severity and/or frequency ratings of items on scales reflecting components of the agitation syndrome. Examples of the latter approach include the Cohen-Mansfield Agitation Inventory (CMAI)⁵ and the Neuropsychiatric Inventory (NPI),⁶ the most widely used measures across trials. A shortcoming of this approach is that deriving ratings exclusively from caregiver input leads

to multiple biases.⁴ The NPI-Clinician (NPI-C)⁷ version was developed to address such limitations; however, to date there is no gold standard scale rating for these trials.

In 2015 the International Psychogeriatric Association (IPA) published provisional criteria for agitation in cognitive disorders to facilitate research in the field. Unfortunately, no data have reported on how to measure clinically meaningful change using these criteria.⁸ In 2018, the European Union-North American Clinical Trials in Alzheimer's Disease (EU-US CTAD) Task Force on Agitation encouraged the evidence-based development of a single rating scale tailored to the IPA agitation criteria, using existing datasets, and using items from existing scales.⁹ The desired instrument should reflect the syndromic agitation criteria developed by the IPA, incorporate information from both patient and caregiver, define clinically meaningful effects, demonstrate sensitivity to change, and form the basis for powering studies. In this study we respond to this EU-US CTAD Task Force recommendation by constructing novel outcome measures from existing instruments, with items mapped onto the IPA agitation criteria, and then choosing the best measure based on performance characteristics, including sensitivity to change and predictive validity, as assessed in the Agitation and Aggression AD Cohort (A3C) database.

2 | METHODS

2.1 | Design overview

In a modified Delphi process, items from the CMAI and the NPI-C were mapped by an expert panel onto the IPA agitation criteria domains to generate derivative "novel" measures, CMAI-IPA and NPI-C-IPA. Original existing measures were the CMAI and the NPI-C-Agitation and Aggression domains (NPI-C-A+A). The original and novel instruments were examined in the A3C study database to assess their performance in terms of minimal clinically important difference (MCID), sensitivity to change, and predictive validity. Internal consistency and reliability were also assessed for the novel measures. For these purposes, the modified Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change (mADCS-CGIC)¹⁰ was considered the gold standard.

2.2 | Mapping CMAI and NPI-C items onto IPA agitation criteria: modified Delphi process

The IPA definition for agitation describes three domains: excessive motor activity (EMA), verbal aggression (VA), and physical aggression (PA).⁸ The modified Delphi process included six clinicians with expertise in agitation in dementia. All were directly or indirectly involved with the CTAD Agitation Task Force. All Delphi raters are co-authors. We used a multi-step iterative process that we have previously used for scale generation.¹¹ As a first step, items from the CMAI and the NPI-C were reviewed for relevance to any of the three domains: EMA, VA, or PA. For the CMAI, all items were included, and for the NPI-C, all items from the agitation, aggression, aberrant motor activity, abnormal vocalizations, disinhibition, and irritability/lability domains were

included. As a next step, all relevant items were incorporated into an online survey and rated by the Delphi Panel as 1 (none), 2 (weak), or 3 (strong) for association to each of the three IPA definition domains. For each item, if mean score was ≥ 2.5 , the item was included and applied to the corresponding domain, and if < 1.5 , the item was discarded. Items with scores from 1.5 to 2.5 were retained for further discussion. These residual items were discussed via teleconference and assigned to a domain if 80% consensus was reached. Items that did not distinctly map onto one domain were discarded. Thus, using this Delphi process, all items from the CMAI and the items from the six domains from the original NPI-C: agitation, aggression, aberrant motor activity, disinhibition, irritability/lability, and abnormal vocalizations, were mapped onto the IPA agitation domains. The results were derivative novel measures CMAI-IPA and NPI-C-IPA. The CMAI-IPA is a scale of 19 items rated based on symptom frequency rated by caregiver; the NPI-C-IPA is a scale of 25 items rated based on frequency and severity in a clinician-rated composite score based on information from patient and caregiver (see [supporting information](#)).

2.3 | Original clinical trials outcome measures for agitation in AD

We considered the CMAI and the NPI-C agitation and aggression domains (A+A) to be original measures. The CMAI has been used more frequently across trials in AD agitation while the NPI-C-A+A has been used more recently in trials.¹² The CMAI is a caregiver-rated questionnaire that quantifies the frequency of 29 behaviors. The NPI-C-A+A measures clinician-rated severity of agitation by summing scores on the agitation (13 items) and aggression (eight items) domains of the NPI-C.

2.4 | mADCS-CGIC as a gold standard outcome measure in clinical trials for AD agitation

The Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change (ADCS-CGIC)¹³ is a global rating of change developed to assess clinically significant change in symptoms over time in AD clinical trials. In the A3C study, mADCS-CGIC was rated by experienced clinicians, during face-to-face interviews, and recorded on a standardized case record form. All rating clinicians were previously trained in a standardized training session based on the protocol. The clinicians rated the mADCS-CGIC after assessing the other two scales: CMAI and NPI-C-A+A. The mADCS-CGIC for agitation¹⁰ rates five areas: mood lability, emotional distress, physical agitation, VA, and PA. At follow-up visits in A3C, clinicians rated patient agitation on the mADCS-CGIC as: very much improved (1), much improved (2), minimally improved (3), no change (4), minimally worse (5), much worse (6), and very much worse (7) compared to baseline symptoms. Here we considered the mADCS-CGIC as the gold standard to differentiate two groups of patients, those whose agitation improved significantly (scores of 1 or 2) and those who were minimally improved or unimproved (defined as scores ≥ 3) at 1 and at 3 months of follow-up.

2.5 | Agitation and aggression in AD patients cohort (A3C) study

The A3C is a longitudinal, prospective, multicenter observational study performed at eight memory clinics in France, and their associated long-term care facilities (LTCFs). Clinical visits were scheduled at baseline, monthly during the first 3 months, at 6, at 9, and at 12 months. This frequent schedule of assessments during the first 3 months was intended to simulate a 12-week randomized control treatment trial design. The main objective of A3C was to study, over 12 months, the course of clinically significant agitation based on validated measures (NPI-C, CMAI). An additional goal was to assess relationships between agitation symptoms, and to estimate clinical meaningfulness via comparison to global ratings of agitation (i.e., the mADCS-CGIC) to optimize future trials.

A3C enrolled 262 patients: methods and preliminary results are described in de Mauleon et al.¹⁴ Briefly, participants were aged ≥ 60 years, had a diagnosis of probable AD according to 2011 National Institute on Aging-Alzheimer's Association criteria,¹⁵ with or without cerebrovascular components. Participants exhibited clinically significant agitation at baseline defined by the presence of significant symptoms on at least one of the following agitation symptoms as rated by NPI: agitation/aggression, disinhibition, aberrant motor behavior, and/or irritability with a NPI item score ≥ 4 with NPI frequency score ≥ 2 at entry. Participants also met criteria for the IPA provisional definition of agitation in cognitive disorders.⁸ To be included, patients could live at home with an identified primary caregiver or reside in a LTCF for at least 2 months before inclusion. Participants and their caregivers took part in the study voluntarily: written informed consent was obtained from all patients (or legal representatives) and caregivers (for the community dwelling population).

2.6 | Statistical analyses

All analyses were performed using SAS version 9.4 (SAS Institute Inc.). Baseline characteristics are reported as frequencies and percentages for nominal variables and mean \pm standard deviation (SD) for quantitative variables.

To estimate MCID between mADCS-CGIC (very/much improved: 1 or 2 vs. others: ≥ 3) and the agitation measures (CMAI and NPI-C-A+A as original and CMAI-IPA and NPI-C-IPA as novel) at 1 and 3 months follow-up, receiver operating characteristic (ROC) analyses¹⁶ were conducted. Youden's index (YI, calculated as sensitivity + specificity - 1) was used to calculate the optimal cut-offs in the ROC curves.¹⁷ Areas under ROC curves (AUC) were also calculated.

To assess sensitivity to change of the four measures (CMAI, CMAI-IPA, NPI-C-A+A, and NPI-C-IPA) for participants with significant improvement (mADCS-CGIC = 1 or 2), five indices were calculated at 1- and at 3-month follow-up:

1. Effect size (ES)¹⁸ = (change/SD [T0]);
2. Standardized response mean (SRM)¹⁹ = (change/SD [change]);

3. Guyatt response index (GRI)²⁰ = (change/SD [change for stable subjects]);
4. Reliable change index (RCI)²¹ = (change/SE [change]); numerically, RCI represents the number of scale points needed on a given psychometric measure to determine whether a change in score from pre- to-post-treatment is due to real change or chance variation. The RCI expresses the amount of change between pre- and post-treatment scores on the measurements that would be statistically reliable);
5. Change from baseline, estimated with linear mixed models, of the four measures converted to a Z-score (reduced centered variable). In these models, the dependent variables were CMAI or CMAI-IPA or NPI-C-A+A or NPI-C-IPA and the independent fixed effects were: baseline value of the outcome, grouped according to the mADCS-CGIC (1 or 2; = 3; = 4; ≥5), time, and interaction between group and time.

For all indices, a high absolute value meant a high sensitivity to change.

To assess predictive validity of the four measures (CMAI, CMAI-IPA, NPI-C-A+A, and NPI-C-IPA), linear mixed models were estimated to assess whether an improvement at 1 or 3 months on each measure, using cut-offs from the ROC analysis, predicted subsequent change, at 11 or 9 months, respectively, in cognitive function (Mini-Mental State Examination [MMSE] scale²²) or functional autonomy (Activities of Daily Living [ADL] scale²³). Predictive validity in the mADCS-CGIC gold standard was similarly assessed. In each model, we included subject-specific random effects to consider intra-subject correlations, a random intercept to consider the heterogeneity of outcomes at the first timepoint (here 1 or 3 months), and a random slope to take into account heterogeneity of slopes between participants. In the models, we included the following fixed effects: value of the outcome at the first timepoint, grouped according to improvement in the four agitation measures (improvement = yes vs. no) or mADCS-CGIC scale (1 or 2 vs. ≥3), time, and interaction between group and time. In all models the following potential confounders were tested: sex, age, agitation level at baseline, and the change from the first time of agitation level. None of these variables were associated with both the outcomes and with the variable of interest (improvement of agitation).

To check the accuracy of the novel scales, Cronbach's alpha was estimated to examine the internal consistency of items for each IPA domain (EMA, VA, PA). To assess test-retest reliability (in terms of stability) we estimated intraclass correlation coefficients (ICC) and their 95% confidence intervals (CIs) at 1 and 3 months for stable participants (mADCS-CGIC = 4).

In accordance with the guidelines of Koo and Li,²⁴ ICC estimates and their 95% CIs were calculated using the SAS ICC macro developed by Robert Hamer (<https://support.sas.com/kb/25/031.html>) based on a single rater/measurement, absolute-agreement, two-way mixed-effects model.²⁵ ICC results can be interpreted as follows:

- 5- <0.5 are indicative of poor reliability
- 5- 0.5 to 0.75 moderate reliability

TABLE 1 Characteristics at baseline of A3C population

Characteristics at baseline	N = 262
Sex (female), n (%)	153 (58.4%)
Age, years ^a	82.4 (±7.2)
Level of education, n (%):	
No diploma	15 (6.1%)
Primary school without certificate	75 (30.4%)
Primary school certificate	59 (23.9%)
Secondary education, without high-school diploma	54 (21.9%)
High-school diploma (Baccalaureat) or higher	44 (17.8%)
Living at home, n (%)	183 (69.9%)
MMSE, total ^a	10.0 (±8.0)
Neuropsychiatric symptoms measurements	
CMAI, total ^a	62.0 (±15.8)
CMAI-IPA ^a	38.5 (±12.2)
NPI-C-A+A ^a	15.8 (±10.8)
NPI-C-IPA ^a	15.2 (±10.8)
IPA, definition of agitation, n (%):	
Excessive motor activity	199 (76.3%)
Verbal aggression	199 (76.3%)
Physical aggression	115 (44.1%)
CGI-S, severity n (%):	
Normal	0 (0.0%)
Borderline mentally ill	0 (0.0%)
Mildly ill	32 (12.2%)
Moderately ill	89 (34.0%)
Markedly ill	93 (35.5%)
Severely ill	39 (14.9%)
Among the most extremely ill	9 (3.4%)

Abbreviations: A+A, Agitation+Aggression; CGI-S, Clinical Global Impression of Severity; CMAI, Cohen-Mansfield Agitation Inventory; IPA, International Psychogeriatric Association; MMSE, Mini-Mental State Examination; NPI-C, Neuropsychiatric Inventory Clinician Rating.

^aMean (± standard deviation).

- 5- 0.75 to 0.9 good reliability
- 5- >0.90 excellent reliability.

Finally, we examined the association between the novel scales and the inclusion criteria with binary responses for each IPA domain to study the internal validity of novel measures.

3 | RESULTS

3.1 | A3C study participants

Table 1 reports baseline characteristics of the A3C population. Mean age was 82.4 (±7.2) years and 153 (58.4%) were female. Mean MMSE score was 10.0 (±8.0).

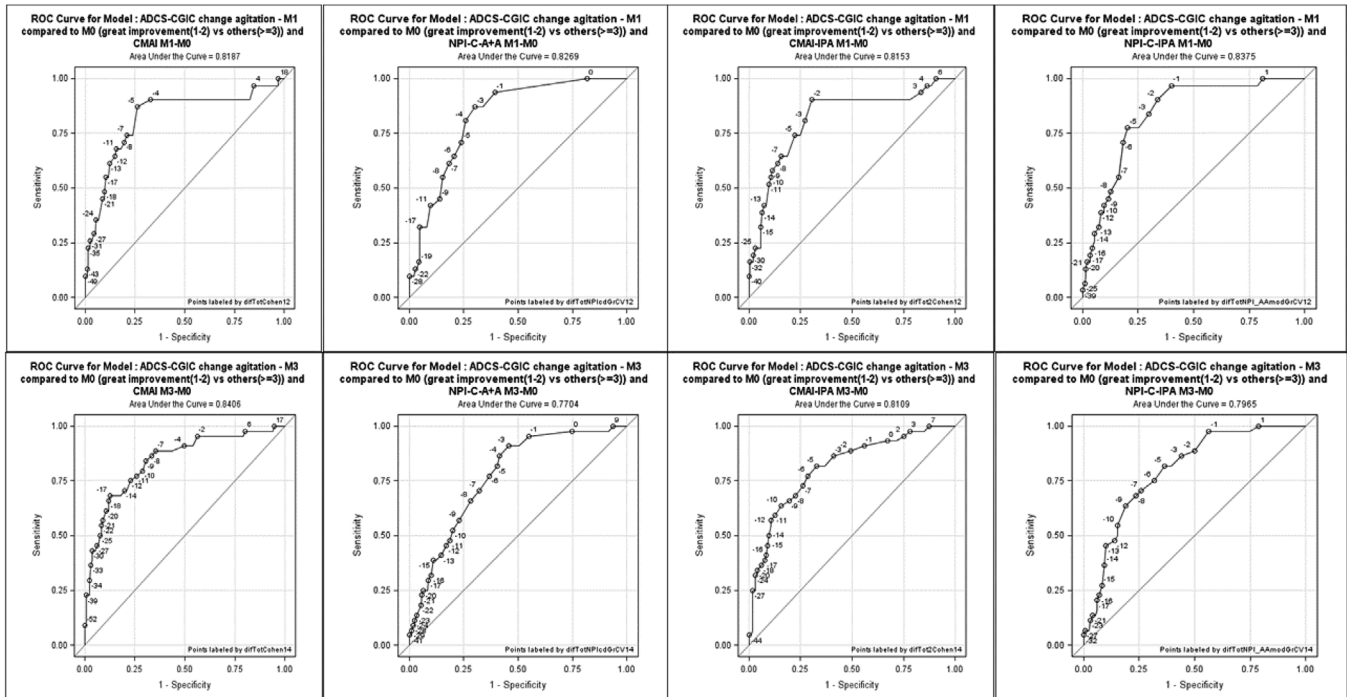


FIGURE 1 Receiver operating characteristic (ROC) curve for model: modified Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change (mADCS-CGIC) 1–2 versus ≥ 3 and Cohen-Mansfield Agitation Inventory (CMAI), Neuropsychiatric Inventory Clinician Rating (NPI-C)-Agitation+Aggression (A+A), CMAI-International Psychogeriatric Association (IPA), and NPI-C-IPA at 1 and at 3 months

3.2 | Minimal clinically important differences of CMAI, CMAI-IPA, NPI-C-A+A, and NPI-C-IPA

At 1 month, 31 patients (13.0%) were much or very much improved by mADCS-CGIC. For the whole cohort, CMAI mean score was 56.7 (SD \pm 16.2), CMAI-IPA was 35.0 (SD \pm 10.9), NPI-C-A+A was 12.6 (SD \pm 10.9), and NPI-C-IPA was 12.4 (SD \pm 10.9).

At 3 months, 44 patients (20.5%) were much or very much improved by mADCS-CGIC. CMAI mean score was 52.5 (SD \pm 15.9), CMAI-IPA was 32.6 (SD \pm 11.0), NPI-C-A+A was 9.9 (SD \pm 9.0), and NPI-C-IPA was 10.1 (SD \pm 10.0).

Based on ROC analysis and the YI, at 1 month, the estimated MCID (defined as 1 or 2 on the mADCS-CGIC vs. ≥ 3) for the original CMAI was -5 (odds ratio [OR] = 18.9, 95% CI = 6.3–56.4). For the derived CMAI-IPA the MCID was -2 (OR = 21.2; 95% CI = 6.2–72.3). For the NPI-C-A+A the estimated MCID was -3 (OR = 15.5, 95% CI = 5.2–46.2). For the novel NPI-C-IPA the MCID was -5 (OR = 13.5, 95% CI = 5.4–33.4). At 3 months, the estimated MCID for CMAI was -17 (OR = 14.9, 95% CI = 6.8–32.6); for CMAI-IPA was -5 (OR = 9.3, 95% CI = 4.0–21.2); for NPI-C-A+A was -3 (OR = 11.9, 95% CI = 4.1–34.8); and for NPI-C-IPA was -5 (OR = 7.8, 95% CI = 3.4–17.9). The estimated MCIDs were similar at 1 and 3 months for both NPI-C instruments. AUCs of all four scales were comparable (Figure 1).

Table 2 shows the different characteristics of MCID for CMAI, CMAI-IPA, NPI-C-A+A, and NPI-C-IPA at 1- and 3-month follow-up, respectively.

3.3 | Sensitivity to change of CMAI, CMAI-IPA, NPI-C-A+A, and NPI-C-IPA for subjects very much improved

Figure 2 displays sensitivity to change properties of the four measures between baseline and 3 months and between baseline and 1 month according to five indices. All four measures had high sensitivity to change at 1 and at 3 months according to ES and SRM indices. All measures also showed a very high GRI.

According to the RCI, CMAI should vary by -9.48 points and the CMAI-IPA by -8.04 points between baseline and 3 months to ensure that the change between pre-treatment and post-treatment is statistically reliable. Similarly, the NPI-C-A+A should vary by -8.08 points and the NPI-C-IPA by -8.96 points, respectively, to ensure statistically reliable change.

According to the mixed models the four scales had high sensitivity to change. At 1 month, for CMAI the estimated change was -19.11 (95% CI = $[-22.32$ to $-15.90]$, $P < .0001$), for CMAI-IPA -12.54 (95% CI = $[-14.85$ to $-10.23]$, $P < .0001$), for NPI-C-A+A -11.68 (95% CI = $[-13.76$ to $-9.60]$, $P < .0001$), and for NPI-C-IPA -11.17 (95% CI = $[-13.14$ to $-9.20]$, $P < .0001$). At 3 months, for CMAI the change was -20.96 (95% CI = $[-23.68$ to $-18.25]$, $P < .0001$), for CMAI-IPA -13.31 (95% CI = $[-15.26$ to $-11.37]$, $P < .0001$), for NPI-C-A+A -12.73 (95% CI = $[-14.37$ to $-11.09]$, $P < .0001$), and for NPI-C-IPA -11.65 (95% CI = $[-13.26$ to $-10.05]$, $P < .0001$). To compare the changes observed, the four measures were converted to a Z-score.

TABLE 2 Estimate minimal clinically important differences (MCID) of CMAI, CMAI-IPA, NPI-C-A+A, and NPI-C-IPA by comparing the mADCS-CGIC change agitation in two groups (very much improved and much improved vs. the others) at 1 and at 3 months

mADCS-CGIC change agitation M1 compared to M0 great improvement : 1-2 (N = 31, 13.14%) versus others : > = 3 (N = 205, 86.86%)														
	N	Unit	OR	Lower OR	Upper OR	P	AUC	Lower AUC	Upper AUC	Se (%)	Sp (%)	PLR (%)	PPV (%)	NPV (%)
CMAI M1-M0	236	1.00	0.92	0.90	0.95	<.0001	0.82	0.73	0.91	-	-	-	-	-
CMAI M1-M0 (< = -5 vs > -5)	236	-	18.87	6.31	56.40	<.0001	-	-	-	87.10	73.66	3.31	33.33	97.42
mADCS-CGIC change agitation M3 compared to M0 great improvement : 1-2 (N = 44, 20.85%) versus others : > = 3 (N = 167, 79.15%)														
	N	Unit	OR	Lower OR	Upper OR	P	AUC	Lower AUC	Upper AUC	Se (%)	Sp (%)	PLR (%)	PPV (%)	NPV (%)
CMAI M3-M0	211	1.00	0.91	0.89	0.94	<.0001	0.84	0.77	0.91	-	-	-	-	-
CMAI M3-M0 (< = -17 vs > -17)	211	-	14.90	6.82	32.57	<.0001	-	-	-	68.18	87.43	5.42	58.82	91.25
mADCS-CGIC change agitation M1 compared to M0 great improvement : 1-2 (N = 31, 13.14%) versus others : > = 3 (N = 205, 86.86%)														
	N	Unit	OR	Lower OR	Upper OR	P	AUC	Lower AUC	Upper AUC	Se (%)	Sp (%)	PLR (%)	PPV (%)	NPV (%)
CMAI-IPA M1-M0	237	1.00	0.91	0.87	0.94	<.0001	0.82	0.73	0.90	-	-	-	-	-
CMAI-IPA M1-M0 (< = -2 vs > -2)	237	-	21.19	6.21	72.26	<.0001	-	-	-	90.32	69.42	2.95	30.77	97.95
mADCS-CGIC change agitation M3 compared to M0 great improvement : 1-2 (N = 44, 20.75%) versus others : > = 3 (N = 168, 79.25%)														
	N	Unit	OR	Lower OR	Upper OR	P	AUC	Lower AUC	Upper AUC	Se (%)	Sp (%)	PLR (%)	PPV (%)	NPV (%)
CMAI-IPA M3-M0	212	1.00	0.90	0.87	0.93	<.0001	0.81	0.74	0.88	-	-	-	-	-
CMAI-IPA M3-M0 (< = -5 vs > -5)	212	-	9.25	4.03	21.23	<.0001	-	-	-	81.82	67.26	2.50	39.56	93.39
mADCS-CGIC change agitation M1 compared to M0 great improvement : 1-2 (N = 31, 12.97%) versus others : > = 3 (N = 208, 87.03%)														
	N	Unit	OR	Lower OR	Upper OR	P	AUC	Lower AUC	Upper AUC	Se (%)	Sp (%)	PLR (%)	PPV (%)	NPV (%)
NPI-C-A+A M1-M0	239	1.00	0.88	0.84	0.93	<.0001	0.83	0.76	0.89	-	-	-	-	-
NPI-C-A+A M1-M0 (< = -3 vs > -3)	239	-	15.53	5.22	46.24	<.0001	-	-	-	87.10	69.71	2.88	30.00	97.32
mADCS-CGIC change agitation M3 compared to M0 great improvement : 1-2 (N = 44, 20.47%) versus others : > = 3 (N = 171, 79.53%)														
	N	Unit	OR	Lower OR	Upper OR	P	AUC	Lower AUC	Upper AUC	Se (%)	Sp (%)	PLR (%)	PPV (%)	NPV (%)
NPI-C-A+A M3-M0	215	1.00	0.91	0.88	0.94	<.0001	0.77	0.70	0.84	-	-	-	-	-
NPI-C-A+A M3-M0 (< = -3 vs > -3)	215	-	11.92	4.09	34.79	<.0001	-	-	-	90.91	54.39	1.99	33.90	95.88

(Continues)

TABLE 2 (Continued)

mADCS-CGIC change agitation M3 compared to M0 great improvement : 1-2 (N = 44, 20.47%) versus others : >= 3 (N = 171, 79.53%)															
N	Unit	OR	Lower OR	Upper OR	P	AUC	Lower AUC	Upper AUC	Se (%)	Sp (%)	PLR (%)	PPV (%)	NPV (%)		
mADCS-CGIC change agitation M1 compared to M0 great improvement : 1-2 (N = 31, 13.03%) versus others : >= 3 (N = 207, 86.97%)															
N	Unit	OR	Lower OR	Upper OR	P	AUC	Lower AUC	Upper AUC	Se (%)	Sp (%)	PLR (%)	PPV (%)	NPV (%)		
NPI-C-IPA M1-M0	238	1.00	0.87	0.82	0.91	<.0001	0.84	0.77	0.90	-	-	-	-	-	-
NPI-C-IPA M1-M0 (<= -5 vs >-5)	238	-	13.47	5.44	33.38	<.0001	-	-	-	77.42	79.71	3.82	36.36	95.93	
mADCS-CGIC change agitation M3 compared to M0 great improvement : 1-2 (N = 44, 20.56%) versus others : >= 3 (N = 170, 79.44%)															
N	Unit	OR	Lower OR	Upper OR	P	AUC	Lower AUC	Upper AUC	Se (%)	Sp (%)	PLR (%)	PPV (%)	NPV (%)		
NPI-C-IPA M3-M0	214	1.00	0.89	0.85	0.93	<.0001	0.80	0.73	0.86	-	-	-	-	-	-
NPI-C-IPA M3-M0 (<= -5 vs >-5)	214	-	7.84	3.43	17.93	<.0001	-	-	-	81.82	63.53	2.24	36.73	93.10	

Abbreviations: A+A, Agitation+Aggression; AUC, area under the curve; CMAI, Cohen-Mansfield Agitation Inventory; IPA, International Psychogeriatric Association; M, month; mADCS-CGIC, modified Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change; NPI-C, Neuropsychiatric Inventory Clinician Rating; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; SD, standard deviation; Se, sensitivity; Sp, specificity.

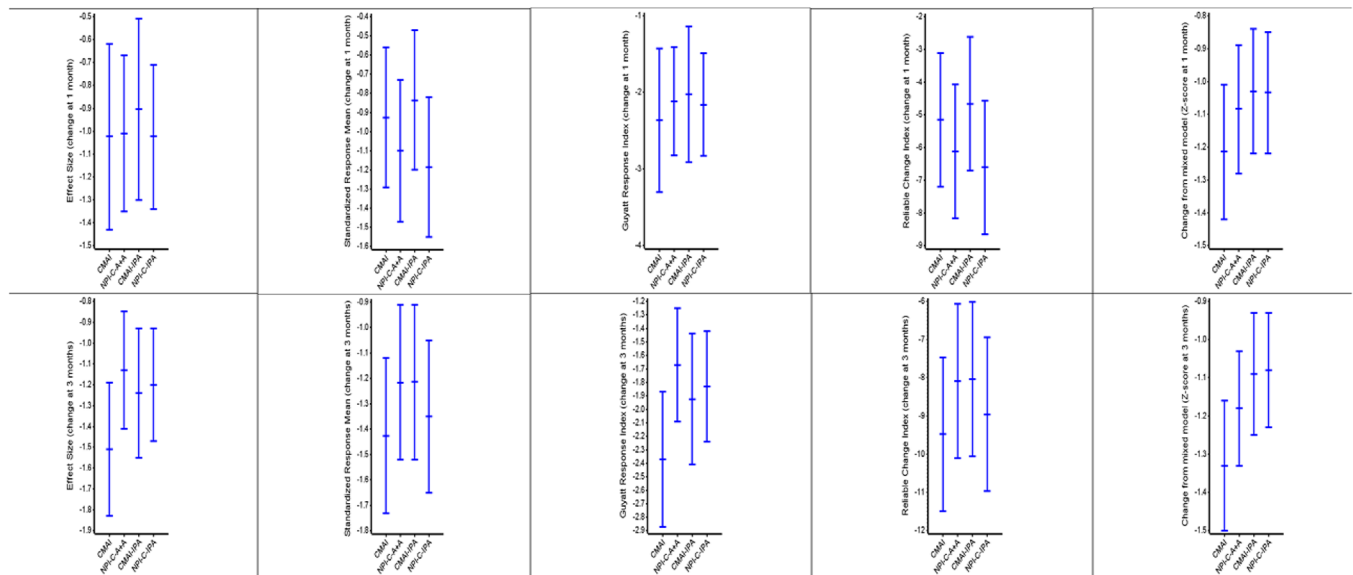


FIGURE 2 The sensitivity to change of Cohen-Mansfield Agitation Inventory (CMAI), Neuropsychiatric Inventory Clinician rating (NPI-C)-Agitation+Aggression (A+A), CMAI-International Psychogeriatric Association (IPA), and NPI-C-IPA between M0 and M3 and between M0 and M1 according to five indices

All four measures demonstrated comparable high sensitivity to change, especially comparing the original and novel. Figure 2 shows a slightly better performance in sensitivity to change for CMAI than NPI-C-A+A at 3 months, but not at 1 month, considering that CMAI also

lute values of different indices were higher than NPI-C-A+A. However, no significant differences in terms of performance were found between CMAI-IPA and NPI-C-IPA at 1 and at 3 months. (see Table S1 in supporting information).

3.4 | Predictive validity

Table 3 shows results from linear mixed models assessing whether a significant improvement, at 1 or 3 months, on the four measures and on the mADCS-CGIC, was predictive of subsequent changes, at 11 or 9 months, in cognitive function (MMSE) or functional autonomy (ADL). Improvement at 3 months on NPI-C-A+A was significantly predictive of less ADL and MMSE change at 9 months of follow-up. Improvement at 1 month on CMAI-IPA was predictive of less ADL change at 11 months of follow-up. Change in the other agitation measures were not predictive of MMSE or ADL at any timepoint.

3.5 | The accuracy of CMAI-IPA and NPI-C-IPA

The analysis of the internal consistency of the NPI-C-IPA showed a Cronbach's alpha at 0.8 for EMA, 0.7 for the VA, and 0.7 for the PA. Concerning the CMAI-IPA, the Cronbach's alpha was 0.8 for PA, 0.6 for EMA, and 0.3 for VA (see Table S2 in supporting information). Concerning test-retest reliability, the mean of ICC for the CMAI-IPA was 0.8 at 1 month and was 0.7 at 3 months, the ICC for the NPI-C-IPA was 0.9 at 1 month, and was 0.8 at 3 months (see Table S3 in supporting information).

Table S4 in supporting information showed statistically significant OR of the association between the CMAI-IPA and NPI-C-IPA and each of the IPA domain (EMA, VA, and PA; $P < .0001$).

4 | DISCUSSION

Accurate and meaningful measurement of agitation is central to developing and testing the efficacy of treatments for agitation in AD. However, no gold standard exists. Little is known about the longitudinal course of clinically significant agitation in AD, about the variability in different outcome measures over time, or the definition of a clinically meaningful improvement. In this study of A3C participants we assessed the most widely used measures of agitation in clinical trials, the CMAI and NPI-C-A+A, as well as Delphi panel-derived measures informed by the IPA agitation criteria, the CMAI-IPA and the NPI-C-IPA.

At 3-month follow-up, one fifth of patients with clinically significant agitation were much or very much improved globally. Estimated MCIDs were the same at 1 and 3 months for both NPI-C instruments, indicating that to differentiate trial participants who are very much improved, a period of 1 month may be sufficient. Similar findings were reported for the placebo arm of a recent RCT.²⁶ AUCs of all four scales were similar, suggesting no scale had an advantage in predicting clinician ratings. The sensitivity to change properties of the four measurements at 1 and at 3 months follow-up according to five different indices were high, and comparable among the scales, especially between original measures and novel versions. Cronbach's alpha and ICC results showed a better accuracy for the NPI-C-IPA than for the CMAI-IPA.

In 2015, the field significantly progressed with publication of the IPA agitation provisional criteria in cognitive disorders, helping to define

agitation better as a syndrome in both clinical and research setting. This was an important step forward, given that agitation was interchangeably or confusedly described as a symptom or a syndrome, with unclear definitions of which behaviors constituted agitation, debates about the differences between agitation and aggression, and the role of other emergent dementia-related behaviors.²⁷ The EU-US CTAD Task Force focused its 2017 meeting on finding the best outcome measure for agitation in dementia trials. The Task Force advocated the development of a single rating scale that reflects agitation as a unitary phenomenon, and that best reflects the IPA criteria and the situations in which agitation occurs.

We explored the MCID of the four different measures. MCID is crucial to calculate sample size in clinical trials. Information about MCID is very important for newer scales; less widely used measures, such as the NPI-C-A+A; as well as for the novel, IPA criteria-informed measures.⁴ Regarding MCID of NPI-C-A+A and NPI-C-IPA, follow-up of 1 month seemed sufficient to detect the group of patients who would be very much improved by 3 months. This finding has important implications for trial design; while symptom response in clinical trials may have a different timeframe than in observational studies such as ours, what is important is that natural variation in symptoms can be observed in a short time frame (1 month) and predicts later response (3 months). This suggests that the widely used trial durations of 6 to 12 weeks is long enough to capture initial clinical response, and that we do not need substantially longer timeframes. Moreover, this may have important implications for clinical practice. In addition, these results, combined with the report by Rosenberg et al.,²⁶ are helpful in considering the use of a "run-in period" of 3 to 4 weeks before active treatment to reduce the placebo effect often observed in such trials. Findings from this study also provide a better estimation of placebo group variability in trials, thus allowing for more precise power estimates.

We used global ratings as a gold standard indicator of meaningful clinical change against which we compared scale performance. Notably CMAI-IPA and CMAI had high scores at follow-up, even for participants who were "much" better clinically. This was not the case for the NPI-C measures on which participants who were clinically "much" better had scores approaching zero at follow-up. The NPI-C scales better approximate meaningful clinical improvement. This may be because CMAI contains items less clinically relevant to this population (e.g., the verbal nonaggressive items²⁸). In fact, CMAI has many items that are rarely rated considering that it was originally developed in nursing homes on residents with advanced dementia. This assertion is corroborated by the fact that mapping the original CMAI onto the IPA criteria based on the expert Delphi process resulted in a shorter scale suggesting that the original one was unnecessarily broad. Another explanation could be that the CMAI is a frequency scale, while NPI-C captures frequency and severity resulting in a more robust assessment of the clinically relevant behaviors. Further, CMAI ratings rely solely on subjective caregiver input resulting in potential bias, in contrast to NPI-C in which trained clinicians use input from caregiver and patient to produce clinician severity scores. With respect to the novel NPI-C-IPA, not every one of the 25 items have been assessed for their "utility" in clinical and research settings, although this is the work for future studies.

TABLE 3 Predictive validity of CMAI, NPI-C-A+A, CMAI-IPA, NPI-C-IPA, and mADCS-CGIC between M0 and M3 and between M0 and M1 according to MMSE and ADL

Change MMSE/ADL at 11 months (M12-M1)				
Outcome	Improvement agitation at 1 month	Yes mean [95% CI]	No mean [95% CI]	Yes_versus_no mean [95% CI] P
1. Total MMSE	1. CMAI M1-M0 (≤ -5 vs > -5)	-2.32 [-3.57 to -1.08]	-2.07 [-2.93 to -1.20]	-0.26 [-1.77 to 1.26] P = .7375
1. Total MMSE	2. NPI-C-A+A M1-M0 (≤ -3 vs > -3)	-2.05 [-3.23 to -0.88]	-2.20 [-3.07 to -1.33]	0.15 [-1.32 to 1.61] P = .8419
1. Total MMSE	3. CMAI-IPA M1-M0 (≤ -2 vs > -2)	-2.34 [-3.49 to -1.18]	-2.02 [-2.92 to -1.12]	-0.31 [-1.78 to 1.15] P = .6728
1. Total MMSE	4. NPI-C-IPA M1-M0 (≤ -5 vs > -5)	-2.20 [-3.54 to -0.86]	-2.13 [-2.96 to -1.31]	-0.07 [-1.64 to 1.51] P = .9336
1. Total MMSE	5. mADCS-CGIC M1-M0 (1-2 vs ≥ 3)	-3.21 [-5.16 to -1.26]	-2.00 [-2.75 to -1.25]	-1.21 [-3.30 to 0.88] P = .2546
2. Total ADL	1. CMAI M1-M0 (≤ -5 vs > -5)	-0.76 [-1.10 to -0.43]	-0.93 [-1.16 to -0.70]	0.17 [-0.24 to 0.57] P = .4154
2. Total ADL	2. NPI-C-A+A M1-M0 (≤ -3 vs > -3)	-0.75 [-1.06 to -0.44]	-0.94 [-1.17 to -0.71]	0.19 [-0.19 to 0.58] P = .3273
2. Total ADL	3. CMAI-IPA M1-M0 (≤ -2 vs > -2)	-0.61 [-0.92 to -0.30]	-1.02 [-1.25 to -0.79]	0.41 [0.03 to 0.80] P = .0360
2. Total ADL	4. NPI-C-IPA M1-M0 (≤ -5 vs > -5)	-0.66 [-1.02 to -0.30]	-0.95 [-1.17 to -0.73]	0.29 [-0.13 to 0.72] P = .1706
2. Total ADL	5. mADCS-CGIC M1-M0 (1-2 vs ≥ 3)	-1.03 [-1.56 to -0.50]	-0.86 [-1.06 to -0.66]	-0.18 [-0.74 to 0.39] P = .5381
Change MMSE/ADL at 9 months (M12-M3)				
Outcome	Improvement agitation at 3 months	Yes mean [95% CI]	No mean [95%CI]	Yes_versus_no mean [95% CI] P
1. Total MMSE	1. CMAI M3-M0 (≤ -17 vs > -17)	-1.59 [-2.96 to -0.22]	-1.71 [-2.52 to -0.89]	0.12 [-1.48 to 1.71] p = 0.8863
1. Total MMSE	2. NPI-C-A+A M3-M0 (≤ -3 vs > -3)	-1.06 [-1.97 to -0.15]	-2.45 [-3.48 to -1.42]	1.38 [0.01 to 2.76] P = .0483
1. Total MMSE	3. CMAI-IPA M3-M0 (≤ -5 vs > -5)	-1.64 [-2.70 to -0.58]	-1.68 [-2.61 to -0.76]	0.04 [-1.37 to 1.45] P = .9534
1. Total MMSE	4. NPI-C-IPA M3-M0 (≤ -5 vs > -5)	-1.04 [-2.06 to -0.03]	-2.21 [-3.14 to -1.27]	1.17 [-0.21 to 2.55] P = .0973
1. Total MMSE	5. mADCS-CGIC M3-M0 (1-2 vs ≥ 3)	-1.62 [-3.09 to -0.16]	-1.69 [-2.47 to -0.91]	0.07 [-1.60 to 1.73] P = .9366
2. Total ADL	1. CMAI M3-M0 (≤ -17 vs > -17)	-0.56 [-0.94 to -0.18]	-0.82 [-1.04 to -0.60]	0.27 [-0.17 to 0.71] P = .2315
2. Total ADL	2. NPI-C-A+A M3-M0 (≤ -3 vs > -3)	-0.58 [-0.83 to -0.33]	-0.96 [-1.24 to -0.68]	0.38 [0.00 to 0.75] P = .0480
2. Total ADL	3. CMAI-IPA M3-M0 (≤ -5 vs > -5)	-0.59 [-0.88 to -0.30]	-0.87 [-1.12 to -0.62]	0.29 [-0.10 to 0.67] P = .1427
2. Total ADL	4. NPI-C-IPA M3-M0 (≤ -5 vs > -5)	-0.63 [-0.91 to -0.35]	-0.85 [-1.10 to -0.59]	0.21 [-0.17 to 0.59] P = .2681
2. Total ADL	5. mADCS-CGIC M3-M0 (1-2 vs ≥ 3)	-0.74 [-1.15 to -0.33]	-0.75 [-0.97 to -0.54]	0.02 [-0.44 to 0.48] P = .9377

Abbreviations: A+A, Agitation+Aggression; ADL, Activities of Daily Living; CI, confidence interval; CMAI, Cohen-Mansfield Agitation Inventory; IPA, International Psychogeriatric Association; M, month; mADCS-CGIC, modified Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change; MMSE, Mini-Mental State Examination; NPI-C, Neuropsychiatric Inventory Clinician Rating; SD, standard deviation.

Note: Results from mixed models.

Comparing the original to novel measures, all AUCs were similar, suggesting that no measure had a clear advantage in predicting clinician ratings. In addition, the sensitivity to change properties of all four measures at 1 and at 3 months were similar, supporting the conclusion that the four measures have comparable psychometric properties. Thus, for the field to incorporate IPA agitation definition into measurement, the novel CMAI-IPA and NPI-C-IPA are preferred as they are more efficient. Furthermore, the original NPI-C-A+A only included the agitation and the aggression domains, while the new NPI-C-IPA included six domains from the original NPI-C related to agitation symptoms: agitation, aggression, aberrant motor activity, abnormal vocalizations, disinhibition, and irritability/lability domains, therefore better capturing the breadth of the IPA agitation domains. Thus, the novel scales CMAI-IPA and NPI-C-IPA perform at least as well as the original scales but in addition, they reflect the IPA agitation criteria. The Food and Drug Administration recognized "agitation as a syndrome for clinical trials targeting," and IPA criteria aim to define this syndrome given the historical heterogeneity in definition used.

Concerning the accuracy of the novel scales, the statistical metrics showed that the internal consistency of the NPI-C-IPA was good concerning the three IPA domains (EMA, VA, and PA). However, the internal coherence of CMAI is more questionable; Cronbach's alpha was good for EMA but it was low for VA and very low for PA. One explanation could be that when calculating the Cronbach's alpha, the number of items is taken into account. Indeed, in the EMA and VA domains of CMAI-IPA, there are only four items in each domain. This could explain the slightly lower coefficient in the EMA domain but not the very low coefficient for VA. To estimate the test-retest reliability, ICC statistical analysis of the NPI-C-IPA highlighted excellent reliability at 1 month and good reliability at 3 months. In contrast, the ICC of the CMAI-IPA showed good reliability at 1 month and moderate reliability at 3 months. In summary, internal consistency and reliability analyses demonstrated a better accuracy for the NPI-C-IPA than for the CMAI-IPA. The internal validity of the novel scales seemed to be good based on the analyses of the association between the CMAI-IPA and the NPI-C-IPA items with each IPA domain (EMA, VA, and PA).

The major strength of this study is that it uses data from A3C that mimics a clinical trial in terms of methodology and measurement in a naturalistic setting in both community and nursing home settings. However, there are notable limitations to this work. A3C was an observational, usual care study, whose population received close follow-up, and was frequently treated with psychotropics or non-pharmacological approaches for agitation.¹⁴ However, we believe it is an appropriate dataset for the validation of outcome measures—because no trial can ethically eliminate other efforts at treating agitation, development of novel treatments for agitation will by necessity occur against this background. Concerning other statistical metrics to assess the novel scales, unfortunately, we were unable to assess inter-rater reliability because these data were not available. In fact, during the A3C study, two different raters for each patient did not administer the same scale. However, the novel scales are derived from existing scales; therefore, authors hypothesize that the inter-rater variability is comparable to the existing ones (CMAI and NPI-C). Nevertheless, a deeper statistical vali-

ation of novel measures will be the subject of further studies. Another limitation was the substantial attrition observed in the number of participants: 30% at 1 year. Although the attrition rate is high, our result is consistent with studies in the literature. Indeed, in a 2013 study conducted in Norway in subjects with dementia living in nursing homes, the attrition rate in the first year of follow-up was $\approx 32\%$.²⁹ Steinberg et al. followed patients with neuropsychiatric symptoms in dementia at home for 5 years. The attrition rate was $\approx 42\%$ at 1.5 years of follow-up.³⁰ Attrition is common in cohorts of older adults with dementia.³¹ The study by Burke et al.³² highlighted that the presence of neuropsychiatric symptoms is a factor influencing attrition. This is why A3C sample size was calculated considering 15% attrition.¹⁴ Finally, this attrition is consistent with that seen in "real-world clinical settings" for this vulnerable population, as A3C is a naturalistic study in a usual care setting. Nevertheless, attrition during the first 3 months, the critical period of A3C study, was much lower (13%).¹⁴ For all these reasons, we believe that our data can be generalizable to the population of patients with AD presenting with agitation symptoms. The attrition data we report will also help in power calculations for sample sizes in future trials.

These results also have relevance to clinical practice. IPA agitation domain-specific measures are an important advance in measurement but also in management of agitation in routine care. Use of these scales in clinical settings will allow for better definition of agitation symptoms, optimization of non-pharmacological or medication options, and better assessment of efficacy.²⁷

5 | CONCLUSION

In summary, in a naturalistic study of AD patients with agitation, the IPA-informed novel CMAI-IPA and NPI-C-IPA scales, both designed to reflect the IPA Agitation Criteria, performed at least as well as the original scales. We found better statistical accuracy and clinical relevance for the NPI-C-IPA over the CMAI-IPA. As these novel measures accurately represent the IPA Agitation Syndrome, we propose that future agitation clinical trials using IPA criteria for inclusion use these novel scales, notably the NPI-C-IPA, to capture clinical effects of treatments. In the current absence of optimal outcome measures to demonstrate treatment effectiveness, these novel scales represent a step forward in the field of treatment development of agitation in AD.

ACKNOWLEDGMENTS

This work was supported in part by Ethypharm, by grant P30AG066507 to the Johns Hopkins Alzheimer's Disease Research Center, the Richman Family Precision Medicine Center of Excellence in Alzheimer's Disease at Johns Hopkins University, and grants R01AG050515 and R01AG054771 to Dr. Rosenberg.

CONFLICTS OF INTEREST

David Miller is a full-time employee of Signant Health. Cedric O'Gorman is a full-time employee of Axsome Therapeutics. Constantine Lyketsos declares: (1) grant support (research or CME) from NIH,

Functional Neuromodulation, Bright Focus Foundation and (2) payment as consultant or advisor from Avanir, Astellas, Roche, Karuna, SVB Leerink, Maplight, Axsome, Global Institute on Addictions. Maria Soto declares: payment as consultant or advisor from Avanir, Acadia. Zahinoor Ismail declares: (1) payment as consultant or advisor from Lundbeck/Otuska outside the submitted work; (2) payment to his institution from Acadia, Biogen, Roche, and Sunovion outside the submitted work. No conflicts are declared for the other authors.

REFERENCES

- Lancôt KL, Amatniek J, Ancoli-Israel S, et al. Neuropsychiatric signs and symptoms of Alzheimer's disease: new treatment paradigms. *Alzheimers Dement*. 2017;3:440-449.
- Lyketsos CG, Carrillo MC, Ryan JM, et al. Neuropsychiatric symptoms in Alzheimer's disease. *Alzheimers Dement*. 2011;7:532-539.
- Porsteinsson AP, Drye LT, Pollock BG, et al. Effect of citalopram on agitation in Alzheimer disease: the CitAD randomized clinical trial. *JAMA*. 2014;311:682-691.
- Soto M, Abushakra S, Cummings J, et al. Progress in treatment development for neuropsychiatric symptoms in Alzheimer's disease: focus on agitation and aggression. A report from the EU/US/CTAD task force. *J Prev Alzheimers Dis*. 2015;2:184.
- Cohen-Mansfield J. Conceptualization of agitation: results based on the Cohen-Mansfield agitation inventory and the agitation behavior mapping instrument. *Int Psychogeriatr*. 1997;8:309-315.
- Cummings JL. The Neuropsychiatric inventory assessing psychopathology in dementia patients. *Neurology*. 1997;48:10S-6S.
- De Medeiros K, Robert P, Gauthier S, et al. The Neuropsychiatric Inventory-Clinician rating scale (NPI-C): reliability and validity of a revised assessment of neuropsychiatric symptoms in dementia. *Int Psychogeriatr*. 2010;22:984-994.
- Cummings J, Mintzer J, Brodaty H, et al. Agitation in cognitive disorders: International Psychogeriatric Association provisional consensus clinical and research definition. *Int Psychogeriatr*. 2015;27:7-17.
- Sano M, Soto M, Carrillo M, et al. Identifying better outcome measures to improve treatment of agitation in dementia: a report from the EU/US/CTAD task force. *J Prev Alzheimers Dis*. 2018;5:98-102.
- Drye LT, Ismail Z, Porsteinsson AP, et al. Citalopram for agitation in Alzheimer's disease: design and methods. *Alzheimers Dement*. 2012;8:121-130.
- Ismail Z, Aguera-Ortiz L, Brodaty H, et al. The Mild Behavioral Impairment Checklist (MBI-C): a rating scale for neuropsychiatric symptoms in pre-dementia populations. *J Alzheimers Dis*. 2017;56:929-938.
- Soto M, Andrieu S, Nourhashemi F, et al. Medication development for agitation and aggression in Alzheimer disease: review and discussion of recent randomized clinical trial design. *Int Psychogeriatr*. 2014;16:1-17.
- Schneider LS, Olin JT, Doody RS, et al. Validity and reliability of the Alzheimer's disease cooperative study-clinical global impression of change. *Alzheimer Dis Assoc Disord*. 1997;11(Suppl 2):S22-32.
- de Mauleon A, Delrieu J, Cantet C, et al. Longitudinal course of agitation and aggression in patients with Alzheimer's disease in a COHORT Study: methods and baseline results of the A3C study. *J Prev Alzheimers Dis*. 2021;8(2):199-209.
- McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:263-269.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
- Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biom J*. 2008;50:419-430.
- Cohen J. Statistical power analysis for the behavioral science. *J Am Stat Assoc*. 1988;84:19-74.
- Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care*. 1990;632-642.
- Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40:171-178.
- Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59(1):12-19.
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.
- Katz S. The index of ADL: a standardized measure of biological and psychosocial function. *JAMA*. 1963;185:914-919.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163.
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30.
- Rosenberg PB, Drye LT, Porsteinsson AP, et al. Change in agitation in Alzheimer's disease in the placebo arm of a nine-week controlled trial. *Int Psychogeriatr*. 2015;27(12):1-9.
- Ismail Z, Goodarzi Z. Neuropsychiatric aspects of Alzheimer's disease clinically significant neuropsychiatric symptoms need evidence-based treatment. *Pract Neurol*. 2019;78-83.
- Cohen-Mansfield J. Agitated behavior in persons with dementia: the relationship between type of behavior, its frequency, and its disruptiveness. *J Psychiatr Res*. 2008;43:64-69.
- Selbaek G, Engedal K, Benth JS, Bergh S. The course of neuropsychiatric symptoms in nursing-home patients with dementia over 53-month follow-up period. *Int Psychogeriatr Assoc*. 2014;26(1):81-91.
- Steinberg M, Shao H, Zandi P, et al. Point and 5-year period prevalence of neuropsychiatric symptoms in dementia: the Cache County Study. *Int J Geriatr Psychiatry*. 2010;23(2):170-177.
- Coley N, Gardette V, Toulza O, et al. Predictive factors of attrition in a cohort of Alzheimer disease patients. The REAL.FR study. *Neuroepidemiology*. 2008;31:69-79.
- Burke SL, Hu T, Naseh M, et al. Factors influencing attrition in 35 Alzheimer's Disease centers across the United States. A longitudinal examination of the National Alzheimer's coordinating center's uniform data set. *Aging Clin Exp Res*. 2019;31(9):1283-1297.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: De Mauleon A, Ismail Z, Rosenberg P, et al. Agitation in Alzheimer's disease: Novel outcome measures reflecting the International Psychogeriatric Association (IPA) agitation criteria. *Alzheimer's Dement*. 2021;17:1687-1697. <https://doi.org/10.1002/alz.12335>