The ADAS-cog and clinically meaningful change in the VISTA clinical trial of galantamine for Alzheimer's disease

Kenneth Rockwood, Sherri Fay and Mary Gorman

Division of Geriatric Medicine, Capital District Health Authority, Dalhousie University, Halifax, NS, Canada

Background: A minimum 4-point change at 6 months on the Alzheimer's disease assessment scale-cognitive subscale (ADAS-cog) is deemed clinically important, but this cut-point has been little studied in relation to clinical meaningfulness. In an investigator-initiated, clinical trial of galantamine, we investigated the extent to which a 4-point change classifies goal attainment by individual patients.

Methods: Secondary analysis of the video imaging synthesis of treating Alzheimer's disease (VISTA) study: a 4-month, multi-centre, parallel-group, double-blind, placebo-controlled, trial of galantamine in 130 mild-moderate Alzheimer's disease patients (4-month open-label follow-up). ADAS-cog responses at 6 months were compared with outcomes on three clinical measures: clinician's interview based impression of change-plus caregiver input (CIBIC+), patient/carer-goal attainment scaling (PGAS) and clinician-GAS (CGAS).

Results: Thirty-seven of 99 patients improved by ≥4 points on the ADAS-cog at 6 months, and 16/99 showed ≥4-point worsening. ADAS-cog change scores correlated notionally to modestly with changes on the CGAS (r = −0.31), the PGAS (r = −0.29) and the CIBIC+ (r = 0.31). As a group, patients with ADAS-cog improvement were significantly more likely to improve on the clinical measures; those who worsened showed non-significant clinical decline. Individually, about half were misclassified in relation to each clinical measure; often when the ADAS-Cog detected 'no change', clinically meaningful effects could be detected. Even so, no ADAS-Cog cut-point optimally classified patients' clinical responses.

Conclusion: A 4-point ADAS-cog change at 6 months is clinically meaningful for groups. Substantial individual misclassification between the ADAS-cog and clinical measures suggests no inherent meaning to a 4-point ADAS-cog change for a given patient. Copyright © 2009 John Wiley & Sons, Ltd.

Key words: Alzheimer's disease; ADAS-cog; clinical meaningfulness; judgement-based measures; VISTA

Introduction

The clinical meaningfulness of cholinesterase inhibition in Alzheimer's disease is commonly regarded as small (Livingston and Katona, 2000; Birks, 2006; Gammanpila et al., 2007; Raina et al., 2008). A review conducted for the American College of Physicians and American Academy of Family Physicians (ACP/AAFP) used a minimum clinically meaningful difference approach (Burbank et al., 1999), arguing that a 4-point change at 6 months on the Alzheimer's disease assessment scale-cognitive subscale (ADAS-cog) (Rosen et al., 1984) is clinically meaningful. The 4-point change threshold was accepted by the
accompanying clinical practice guideline from the ACP/AAFP (Qaseem et al., 2008). The commissioned review found a mean effect of −2.83 (Raina et al., 2008)—similar to the Cochrane estimate of a mean 2.7-point difference on the ADAS-cog at 6 months (Birks, 2006)—so that the clinical meaningfulness of current treatment for Alzheimer’s disease (AD) was disparaged.

But how confident can we be that a minimum 4-point change on the ADAS-cog is a valid measure of a clinically meaningful effect in clinical trials data? The contention that a 4-point change is clinically meaningful has been assumed more than studied. In 18-month follow-up studies, differences can be seen with cut-points that show as little as 2-point improvement (cf. ±2 points, 2-7 point decline and > 7 point decline) (Vellas et al., 2007). Recently, we re-analysed data from an open-label trial of donepezil in people with mild-moderate AD (Rockwood et al., 2002) to evaluate the clinical meaningfulness of a 4-point change on the ADAS-cog at 6 months (Rockwood et al., 2007a). In relation to three self-evidently clinically meaningful measures, donepezil-treated patients were more likely to have a response of ‘no change’ (56%, mean change from baseline −0.1 ± 2.0) on the ADAS-cog at 6 months than either a 4-point improvement (20%, −6.2 ± 1.7) or a 4-point decline (24%, 8.0 ± 4.7) (Rockwood et al., 2007a). While patients with ADAS-cog improvement commonly showed improvement on judgement-based measures, many patients with ADAS-cog decline did not decline clinically. These findings suggest that changes in ADAS-cog performance detectable at the group level are not readily detectable at the level of the individual patient. Still, given that regulators commonly rely on the ADAS-cog to evaluate treatment outcomes, understanding its clinical meaningfulness at both levels (the group and the individual) is of interest.

Here we considered the clinical meaningfulness of a 4-point ADAS-cog change at 6 months in patients who entered a randomized, placebo-controlled trial of galantamine for treatment of mild-moderate AD. Individualized goal attainment is recognized as being clinically meaningful—i.e. if people meet their own goals, that is likely to be meaningful to them; and if patients meet goals set by clinicians, such goals are likely to be clinically meaningful (Burns, 2002; Burns et al., 2006; Hirsch, 2006; Frank, 2007). In consequence, we looked at an investigator-initiated, clinical trial of galantamine to investigate the extent to which a 4-point change classifies goal attainment by individual patients. In addition, we compared whether patients with varying ADAS-cog performance (improved/no change/worse) collectively did better/worse on judgement-based clinical measures. Finally, we presented data to see whether 3- or 2-point changes (more in keeping with what is commonly seen) might better classify clinical responses.

Methods

Study design, sample and trial overview

This is a secondary analysis of the video imaging synthesis of treating Alzheimer’s disease (VISTA) study, a 32 week, Canadian, investigator initiated, multi-centre, trial of 130 community-dwelling patients with mild-to-moderate AD (mean age = 77 SD = 7.7; 63% women; 67% mild [MMSE ≥ 20]) (Rockwood et al., 2006). The study was conducted in two phases (Figure 1): a 16 week, parallel group, double-blind, randomized, placebo-controlled trial of flexibly dosed galantamine (12–24 mg/day), followed by a 16 week open-label phase. Diagnoses were made using standard criteria (McKann et al., 1984), as detailed elsewhere (Rockwood et al., 2006; Rockwood et al., 2007b). At the end of the placebo-controlled phase (16 weeks), galantamine patients improved on both the ADAS-cog (Cohen’s d = −0.36, p = 0.04) and the treatment goals set by clinicians (Cohen’s d = −0.30, p = 0.02). Eight weeks into the open label phase (24 weeks), when all patients received galantamine, the judgement-based, clinical measures—the clinician-set treatment goals, the patient/carer-set treatment goals and the clinician’s interview-based clinical impression of change-plus caregiver input (CIBIC+) (Schneider and Olin, 1996)—still showed significant differences by initial treatment assignment, making comparison with the ADAS-cog scores of some interest.

Outcome measures

Cognition was assessed using the 70-point ADAS-cog (Rosen et al., 1984), which tests patients’ abilities in memory, language, orientation, and praxis. Studies of natural progression on this scale, where higher scores (sum of errors) indicate declining cognition, estimate that untreated patients will worsen by (gain) up to 8 points per year (Stern et al., 1994).

The judgment-based goal attainment scaling (GAS) (Kiresuk and Sherman, 1968) was the primary outcome for the VISTA study. Prior to randomization, individualized treatment goals targeting important, dementia-related symptoms were set and monitored by

patients and carers (patient/carer-GAS or PGAS, n = 445 goals), and by treating physicians (clinician-GAS or CGAS, n = 382 goals) (Rockwood et al., 2006). Goals were set on a 5-point scale corresponding to descriptive accounts of the pre-treatment status (baseline, anchored on the scale at 0) and potential better and worse outcomes (+1 = somewhat better than baseline, +2 = much better, −1 = somewhat worse, −2 = much worse). Patient/carer goals were weighted, with the highest rank assigned to the most important goal. Clinician goals were not weighted. At follow-up, the current status for each goal was described in detail and then scaled against the levels established at baseline: goal attainment was ‘0’ (no change) when the outcome was neither better nor worse than baseline, but scaled between +0.5 and +2 when the outcome was better or between −0.5 and −2 when worse. Patients/carers and physicians were each blinded to the goals set by the other. Qualitative analysts categorized the goals into five domains: cognition, function, behaviour, social activities and leisure. Goal setters were neither required to set goals in each domain, nor limited in the number of goals that could be set. Goal setters were required to adhere to the following minimum criteria: (1) that problem areas be evidently related to the patient’s dementia; (2) that the posted better and worse outcomes (i.e. the goals) represent meaningful changes to patients/carers that could reasonably be influenced by the intervention (galantamine) and (3) that the goals could reasonably be achieved within the study period (4–8 months). GAS scores were summarized across patients and

domains using the following formula: $50 + \left[ \frac{\left( \sum w_i x_i \right)}{\left( 0.7 \sum w_i^2 + 0.3 (\sum x_i)^2 \right)^{1/2}} \right]$ where $w_i$ = weight assigned to the $i$th goal and $x_i$ = score of the $i$th goal. The formula results in a score of 50 when there is no net change from baseline (attainment = 0). Summary scores $> 50$ indicate more improvement than worsening, whereas scores $< 50$ indicate net decline. GAS summary scores range approximately from 17 to 82.

The CIBIC+ (Schneider and Olin, 1996), a global clinical measure of patients’ mental/cognitive abilities, behaviour and functional abilities, was a secondary outcome. After conducting separate interviews with the patient and carer, an experienced clinical-rater evaluated dementia severity at baseline (clinician’s interview-based impression of severity-plus caregiver input, CIBIS+) on a scale of 1 (no dementia) to 7 (very severe dementia), and change at post-baseline visits $(1 =$ very much improved, $4 =$ no change, $7 =$ very much worse). The CIBIC+ rater, who was also the facilitator for the patient/carer-GAS, was blind to all other post-baseline measurement and safety data (similarly, the clinician-GAS rater was blinded to the CIBIC+).

Outcomes were assessed at baseline, then at 8-week intervals.

Analysis

We report observed case outcomes by ADAS-cog response at 24 weeks to determine whether outcomes on the ADAS-cog predict outcomes on clinical measures both at the level of the individual patient and in terms of patient groupings. We report outcomes at the group level to provide an overview, and at the level of the individual patient to allow insight into the proposal that clinicians have more difficulty interpreting improvement than decline (Quinn et al., 2002). While imputation of missing data has a role in efficacy studies, its relevance in understanding measurement issues is less evident, hence we used only observed cases to assess true change. As our focus is on the instruments detecting change, we undertook these analyses without comparing treatment arms.

Patients were categorized into ADAS-cog response groups based on change from baseline after 24 weeks (Table 1). At the group level, we summarized patient characteristics, mean baseline outcomes and mean change ($m \Delta$) from baseline at 24 weeks (SD, 95% confidence intervals) for the ADAS-cog and the clinical measures. Change from baseline on the CIBIC+ was calculated as the difference between the rating given at follow-up (ratings of 1–7) and the equivalent to baseline rating of ‘no change’ (rating = 4). For group scores, Pearson correlation coefficients were calculated to compare change on the ADAS-cog with changes in the outcome measures after 24 weeks.

Individual patient changes on the ADAS-cog were cross-classified against individual patient responses on the clinical measures. Recalling that the clinical measures assay more domains than does the ADAS-cog, we approached this as a convergent construct validation exercise, computing correlations and comparing proportions, rather than calculating sensitivity and specificity. Similarly, as the various scores were grouped to indicate clinical response, we present absolute raw agreement (as a percentage) and not an inter-rater agreement measure (e.g. $k$). For the two GAS measures, we also separately compared ADAS-cog classifications with attainment of all goals (PGAS-total).

<table>
<thead>
<tr>
<th>Table 1 Criteria for classifying clinically meaningful change at the individual level on the ADAS-cog, CIBIC+ and GAS after 24 weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinically detectable improvement</strong></td>
</tr>
<tr>
<td>ADAS-cog</td>
</tr>
<tr>
<td>CIBIC+</td>
</tr>
<tr>
<td>GAS</td>
</tr>
</tbody>
</table>

ADAS-cog = Alzheimer’s disease assessment scale-cognitive subscale; CIBIC+ = clinicians’ interview-based impression of change-plus caregiver input; GAS = goal attainment scaling.

$^a$The ADAS-cog measures cognitive errors; lower scores indicate better outcomes.
and CGAS-total) and with attainment of only cognition goals (PGAS-cognition and CGAS-cognition). Criteria for determining clinically meaningful responses were specified a priori (Table 1). Note that a minority of patients/carers (8%) identified treatment goals in which the patient had experienced only minor decline, and where the goal for treatment was 'no further worsening' (e.g., to maintain abilities as described at baseline with no discernable decline by the end of 4 months [+1] and 8 months [+2]). Such goals demonstrate that 'no change' over several months with a progressive neurodegenerative disease is commonly regarded as a positive outcome (Burns et al., 2008). However, the current analysis attempts to compare clinical judgments of observable change (positive and negative) with what is accepted as a meaningful accumulation of points (gained/lost) on the ADAS-cog. Consequently, maintenance of the baseline status, even where sought as a goal of treatment, was not categorized as 'improved'.

Proportional differences between response groups were tested using $\chi^2$. To evaluate whether and/or how differences in the cut-point for 'meaningful' change on the ADAS-cog would affect the results, we repeated the analysis using change thresholds of 2- and 3-points. All statistical tests were interpreted at the 5% significance level and carried out using SAS$^R$ version 9.1.3.

Ethics

The VISTA protocol was approved by the Therapeutics Product Directorate of Health Canada and by each institution's Research Ethics Board. The primary study centre, located in Halifax, received ethics approval from the Capital District Health Authority. All patients and carers provided written, informed consent.

Results

Of 130 randomized patients, 101 (78%) were assessed at week 24, of whom 50 had received galantamine treatment continuously and 51 had been started on galantamine only 8 weeks earlier (Figure 1). Two patients did not complete the ADAS-cog at week 24 and were excluded. Patients who were included in this secondary analysis ($n = 99$) performed slightly better on baseline outcome measures than those who were not included ($n = 31$), but only significantly so for the CIBIS+ (3.5 SD = 0.8 vs. 3.8 SD = 0.9, $p = 0.03$). The 24-week subsample was also younger (76.7 SD = 7.2 vs. 79.1 SD = 9.1, $p = 0.03$) and included more men (42 vs. 19%, $p = 0.02$).

ADAS-cog response at 24 weeks

After 24 weeks, 37/99 patients improved by $\geq$ 4 points on the ADAS-cog ($m \Delta = -6.8$, SD = 2.7, range = $-4$ to $-15$), 46 patients showed no change ($m \Delta = -0.4$, SD = 2.0) and 16 patients declined ($m \Delta = 6.3$, SD = 3.3 range = 4-14). Patients with ADAS-cog improvement at 24 weeks had better cognition at baseline than patients with 24-week stability or worsening, but otherwise patients in these three groups did not differ in their baseline characteristics (Table 2).

One patient who provided data for the ADAS-cog at 24 weeks did not provide data for two of the three clinical measures, and is only included in comparisons with the clinician-GAS.

Comparison with judgement-based outcomes after 24 weeks

At the group level, people who improved by $\geq$ 4 points on the ADAS-cog at 24 weeks also showed mean improvement compared to baseline on each of the judgment-based measures (Figure 2). Mean CGAS improvement was evident in patients classified as stable by the ADAS-cog, but otherwise these patients showed no clinical change. Patients with a $\geq$ 4-point ADAS-cog decline showed a general trend towards worsening. Although between group differences on the clinical measures were significant ($p < .01$–$p < .001$), correlations between overall outcomes on the ADAS-cog and the clinical measures were in the low-moderate range: $r = -0.29$ for the PGAS, $r = -0.31$ for the CGAS and $r = 0.31$ for the CIBIC+.

ADAS-cog versus the patient/carer-GAS. The most common response on the patient/carer-GAS (CGAS-total) after 24 weeks was 'no change' (55/98, 56%); 32 patients 'improved' on 2/3 of their patient/carer goals and 11 patients 'worsened' (Table 3). Almost half of the patients classified as 'improved' by $\geq$ 4 points on the ADAS-Cog were similarly classified on the patient/carer-GAS (17/36 improved on both). Agreement between measures was greatest when the response was 'no change' (27/46, 59%) and lowest when the response was 'worsened' (6/16, 38%). Overall, 50/98 patients were classified the same by both measures, however, complete mismatching of responses was uncommon: no one who improved on the ADAS-cog responded as...
Table 2 Baseline characteristics of patients with mild-moderate Alzheimer's disease by ADAS-cog response after 24 weeks

<table>
<thead>
<tr>
<th>ADAS-cog response</th>
<th>Improved by ≥ 4 points</th>
<th>No change (±3 points)</th>
<th>Worsened by ≥ 4 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>37</td>
<td>46</td>
<td>16</td>
</tr>
</tbody>
</table>

Baseline characteristics
- age, mean (SD) years
- female, n (%)
- education, mean (SD) years
- severity — mild? n (%)

Baseline outcomes
- MMSE, mean (SD)
- ADAS-cog?, mean (SD)
- CIBIS+, mean (SD)

Goals set per patient at baseline
- PGAS-total, mean (SD)
- CGAS-total, mean (SD)
- PGAS-cognition', mean (SD)
- CGAS-cognition', mean (SD)

ADAS-cog = Alzheimer's disease assessment scale-cognitive subscale; CIBIS+ = clinicians' interview-based impression of severity-plus caregiver input; CGAS = clinician's goal attainment scaling; MMSE = mini-mental state examination; PGAS = patient/carer's goal attainment scaling; SD = standard deviation.

aMMSE score ≥ 20.

Higher scores indicate greater impairment (ADAS-Cog) or dementia severity (CIBIS-Plus).

b87/96 patients/carers set PGAS-cognition goals; respectively 33/37, 41/46 and 13/16.
c61/96 patients/carers set CGAS-cognition goals, respectively 25/37, 28/46 and 8/16.

Figure 2 Mean change from baseline and 95% confidence intervals on ADAS-cog and judgment-based measures by 24-week ADAS-cog response. Statistically significant differences between ADAS-Cog response groups are noted in the legend. ***p < .0001, **p < .001, *p < .01. ADAS-cog = Alzheimer's disease assessment scale-cognitive subscale; CIBIC+ = clinicians' interview-based impression of change-plus caregiver input; CGAS = clinician's goal attainment scaling; PGAS = patient/carer goal attainment scaling.

Copyright © 2009 John Wiley & Sons, Ltd.

Table 3 Distribution of 24-week responses on the ADAS-cog by responses on clinical measures, observed cases

<table>
<thead>
<tr>
<th>ADAS-cog, n (%)a</th>
<th>4-point change</th>
<th>3-point change</th>
<th>2-point change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Improved</td>
<td>No change</td>
<td>Worsened</td>
</tr>
<tr>
<td>PGAS-total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>36</td>
<td>46</td>
<td>16</td>
</tr>
<tr>
<td>Improved</td>
<td>17 (47.2)</td>
<td>14 (30.4)</td>
<td>1 (6.2)</td>
</tr>
<tr>
<td>No change</td>
<td>19 (52.8)</td>
<td>27 (58.7)</td>
<td>9 (56.3)</td>
</tr>
<tr>
<td>Worsened</td>
<td>0 (0)</td>
<td>5 (10.9)</td>
<td>6 (37.6)</td>
</tr>
<tr>
<td>PGAS-cognition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>33</td>
<td>41</td>
<td>13</td>
</tr>
<tr>
<td>Improved</td>
<td>18 (54.5)</td>
<td>13 (31.7)</td>
<td>2 (15.4)</td>
</tr>
<tr>
<td>No change</td>
<td>15 (45.5)</td>
<td>23 (56.1)</td>
<td>6 (40.3)</td>
</tr>
<tr>
<td>Worsened</td>
<td>0 (0)</td>
<td>5 (12.2)</td>
<td>5 (38.6)</td>
</tr>
<tr>
<td>CGAS-total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>37</td>
<td>46</td>
<td>18</td>
</tr>
<tr>
<td>Improved</td>
<td>18 (48.6)</td>
<td>14 (30.4)</td>
<td>2 (12.5)</td>
</tr>
<tr>
<td>No change</td>
<td>17 (46.0)</td>
<td>26 (56.5)</td>
<td>8 (50.0)</td>
</tr>
<tr>
<td>Worsened</td>
<td>2 (5.4)</td>
<td>6 (13.0)</td>
<td>6 (37.6)</td>
</tr>
<tr>
<td>CGAS-cognition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>25</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>Improved</td>
<td>13 (52.0)</td>
<td>11 (39.3)</td>
<td>1 (12.5)</td>
</tr>
<tr>
<td>No change</td>
<td>11 (44.0)</td>
<td>11 (39.3)</td>
<td>2 (25.0)</td>
</tr>
<tr>
<td>Worsened</td>
<td>1 (4.0)</td>
<td>6 (21.4)</td>
<td>5 (26.2)</td>
</tr>
<tr>
<td>CIBIC+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>36</td>
<td>46</td>
<td>16</td>
</tr>
<tr>
<td>Improved</td>
<td>24 (66.7)</td>
<td>15 (32.6)</td>
<td>1 (2.5)</td>
</tr>
<tr>
<td>No change</td>
<td>6 (16.7)</td>
<td>12 (25.1)</td>
<td>6 (37.5)</td>
</tr>
<tr>
<td>Worsened</td>
<td>6 (16.7)</td>
<td>19 (41.3)</td>
<td>8 (50.0)</td>
</tr>
</tbody>
</table>

ADAS-cog = Alzheimer's disease assessment scale-cognitive subscale; CIBIC-Plus = clinician's interview-based impression of change-plus caregiver input; CGAS = clinician's goal attainment scaling; PGAS = patient/caregiver goal attainment scaling. Data are bolded where there is agreement between measures.

aColumn per cent;

*p < 0.05; **p < 0.01; ***p < 0.001 ns = not significant.
worse on the patient/carer-GAS, but the opposite was true for one patient. When considering only cognitive goals (PGAS-cognition), agreement on improvement between the measures was greater (18/33, 55%) (Table 3).

ADAS-cog versus the clinician-GAS. Despite blinded ratings, classifications of response on the clinician-GAS (CGAS-total) mirrored the patient/carer-GAS both in how it adjudicated the 24-week response—51/99 patients classified as ‘no change’, 34 as ‘improved’ and 14 as ‘worsened’—and in comparison with the 4-point change response on the ADAS-cog (Table 3). Individual ADAS-cog and clinician-GAS responses were in agreement for 50/99 patients. Total mismatching occurred in only 4%; i.e. 2/37 patients ‘improved’ on the ADAS-cog but ‘worsened’ on the clinician-GAS and 2/16 patients with ADAS-cog decline ‘improved’ on the clinician-GAS. When the ADAS-cog detected change, agreement between measures was higher when considering cognition goals only (CGAS-cognition): 52% improved on both (vs. 49% with ADAS-cog/CGAS-total improvement) and 63% (vs. 38%) worsened on both. Patients classified as ‘no change’ on the ADAS-cog were more likely to have changed on their CGAS-cognition goals than to have the same response (61% classified as either ‘improved’ or ‘worsened’).

ADAS-cog versus the CIBIC+. Unlike the two GAS measures, the CIBIC+ detected more change than maintenance at 24 weeks: 41/98 patients were rated as ‘improved’, 38 had ‘worsened’ and 24 met the criterion for ‘no change’ (Table 3). Most patients who improved by four points on the ADAS-cog also improved on the CIBIC+ (24/36, 67%) and half of the patients with ADAS-cog worsening had a similar CIBIC+ response, but overall response agreement between measures occurred for only 44/98 patients. As with the CGAS-cognition goals, the CIBIC+ often detected change (34/46, 74%) when the ADAS-cog detected none. Complete misclassifications were more common with the CIBIC+, occurring in 8% of patients: 6/37 patients with ADAS-cog improvement had CIBIC+ worsening (including one patient with a CIBIC+ rating of six or ‘much worse’) and 2/16 patients with ADAS-cog worsening had CIBIC+ improvement.

Evaluating the change threshold. The analyses were repeated using a 2- and 3-point change on the ADAS-cog to define response. At the group level, the patterns evident in Figure 2 persisted but with increasingly smaller statistical significance between the ADAS-cog response groups. In general, however, when comparing individual responses, as the range of ADAS-Cog became smaller, agreement on ‘no change’ increased, but agreement on improvement and decline lessened (Table 3).

Discussion

At the group level, patients in the VISTA study who had ≥ 4-point improvement on the ADAS-cog at 24 weeks tended to show improvement on judgment-based clinical measures, patients who declined by ≥ 4 points on the ADAS-cog tended to have declined clinically and patients who showed no change on the ADAS-cog tended to show no discernable clinical change. At the individual level however, we noted more variability, so that only about half the time did the ADAS-Cog and the clinical measures agree. Complete disagreement (improvement on the ADAS-Cog vs. worsening on the clinical measures, or vice versa) was uncommon. Most disagreements, instead, were about whether there had been any detectable change. These disagreements generally arose because the clinical measures detected more change than did the ADAS-Cog. In short, neither clinically detectable improvement on the ADAS-cog, nor clinically detectable worsening, necessarily translates into clinically important change for an individual patient. We found no cut-point on the ADAS-cog that optimally classified patients in respect of their clinical response.

Our data should be interpreted with caution. The sample size was small. As argued elsewhere, the interpretability of any ‘raw’ n-point change is suspect on several grounds (Rockwood and MacKnight, 2001; Rockwood, 2004). In particular, it will depend to some extent on average variability—either at baseline or in the change scores. For this reason, calculation of standardized effect sizes is preferred (Rockwood, 2004). Even so, spirited regulatory and payer discussions and much current literature focus on the unadjusted number, so these analyses are directly relevant to those discussions (Birks, 2006; Takeda et al., 2006; Gammapil et al., 2007; Helmer et al., 2007; Vellas et al., 2007; Raina et al., 2008).

We used an observed case analysis, not the last observation carried forward, which is conventional for efficacy studies but which introduces a bias in measurement studies. This also allowed us to use 6-month data, for greater relevance to the many 6-month trial (Birks, 2006) and to our earlier inquiry (Rockwood et al., 2007a). Nevertheless, we conducted
additional post hoc analyses of the 4 months, endpoint data (after which all patients were assigned to galantamine), stratified by treatment group, which verified a treatment response in favour of initial assignment to galantamine.

Our data showed that the answer to understanding clinical meaningfulness is unlikely to be had by fiddling with the cut-point. Rather, adjudicating clinical meaningfulness requires several considerations, including timing, dose-response, biological plausibility, reproducibility, convergence, measures and, as here, understanding clinical judgments and patient/carer preferences (Rockwood and MacKnight, 2001). For that reason, more research into what constitutes individualized assessments of treatment response is necessary (Qaseem et al., 2008). The VISTA trial, having aimed at understanding what constitutes individual benefit, means to contribute to that information. For example, we have shown that for patients in whom repetitive questioning is problematic, those in whom repetition lessened tended to have clinically meaningful benefit overall (Rockwood et al., 2007b).

We recognize that misclassifications between the ADAS-cog and the clinical measures might simply reflect that the scales were designed to capture different things. The ADAS-cog measures impairment in a single domain—cognition—whereas the CIBIC+ and GAS tend to be more global, so that change in non-cognitive areas might not be detected by the ADAS-cog. Similarly, given that CIBIC+ raters tend to set summary scores such that improvement in one area might be offset by worsening in another (Joffres et al., 2006; Joffres et al., 2003), disagreement between the ADAS-cog and the CIBIC+ might be expected. Similarly, the GAS summary formula controls for differences in attainment scaling across individual goals and domains. Even so, when we compared the ADAS-cog to just the cognition goals, the distribution continued to mirror the overall GAS trend. In short, there remains a gap between what the ADAS-cog measures and what is important and detectable in patients’ lives.

The measures also showed differences in the proportions of people estimated as having clinically meaningful decline. The CIBIC+ was the most conservative, rating 34% of patients as worse, whereas the other measures found fewer declines—i.e. the ADAS-cog detected decline in 16%, the clinician-GAS in 14% and the patient/carer GAS in 11%. This finding does not disagree with the contention that CIBIC+ raters better detect decline than improvement (Quinn et al., 2002). On the other hand, the CIBIC+ also rated more patients as improved (42%) than did the other clinical measures (33–34%) or the ADAS-cog (37%).

This secondary analysis differs from the last (Rockwood et al., 2007a) in using a more recent dataset from a controlled trial, with blinding between those who administered the clinical measures. The earlier study identified fewer patients as having a 4-point improvement at 6 months (20%) than here (37%). This likely reflects the initial improvement seen between 4 and 6 months in the VISTA trial from patients newly receiving galantamine. Even so, in this analysis of the 6-month ADAS-cog effect, we are appropriately indifferent to treatment status, as our focus is on clinical meaningfulness of the effects, however achieved.

There is still more to be done to understand the adjudication of individual benefit to treatment in AD. For example, compared with the ADAS-cog and the CIBIC+, the two GAS measures rated fewer people as having declined. It is not yet clear whether this means that GAS is worse at capturing decline, or whether GAS is measuring something to which the other measures are insensitive, but which is nevertheless clinically important; repetitive questioning, as noted, might be one such example (Rockwood et al., 2007b). The further disentangling of individual profiles of successful treatment is motivating additional inquiries by our group.

Conclusions

The ADAS-cog appears to be a reasonable measure for regulatory purposes. In general, the results of this study

Key Points

- Although the ADAS-Cog is widely used as a primary outcome measure in evaluating dementia treatment, there is little evidence to justify consensus-based cut-points in relation to clinical meaningfulness.
- At the group level, changes in three judgement-based clinical measures tended to mirror ADAS-cog changes.
- At the individual level, there was substantial variability within ADAS-cog profiles: patients who improved on the ADAS-cog were unlikely to decline clinically while patients who declined on the ADAS-cog often showed little clinical decline.
- In general, there is no cut-point on the ADAS-cog that optimally classifies patients in respect of their clinical response.
acknowledgements

vista was an investigator-initiated study, co-funded by janssen-ortho canada (80%) and the canadian institutes for health research (20%, grant no. dci-49981). these secondary analyses were sponsored by the alzheimer society of canada (grant no. asc 05-54) and the dalhousie medical research foundation, which supports kr as the kathryn allen weldon professor of alzheimer’s disease research. janssen-ortho canada provided all medications and matching placebos, conducted on-site monitoring and gathered and electronically coded the case report forms. gas and cibic+ data were collected by the coordinating site in halifax, nova scotia. all data are held by the principal investigator (kr) who initiated and supervised all analyses. janssen-ortho canada received a draft of this paper 45 days before submission to verify protocol details. the decision to publish rested entirely with the principal investigator (kr). the vista study investigators, dr c. macknight (halifax, ns), dr f. veloso (regina, sk), dr j. laing (victoria, bc), dr a. roch (sudbury, on), dr p. jarrett (saint john, nb), dr m. gorman (antigonish, ns), dr e. de coteau (saskatoon, sk), dr m. winger (windsor, on), dr j. ali (sydney, ns), dr r. arts (barrie, on), dr d. tal (toronto, on), dr m. renuka-prasad (saskatoon, sk), dr b. campbell (winnipeg, mb) and dr j. sommers (halifax, ns). the authors thank heather jourey (msc) for assistance with the analyses and peter nicholl for assistance with the figures.

references

birks j. 2006. cholinesterase inhibitors for alzheimer’s disease. cochrane database syst reviews, january 25, issue (1), cd005593.

burack d, molnar fj, st john p, man-song-hing m. 1999. key methodological features of randomized controlled trials of alzheimer’s disease therapy: minimal clinically important difference, sample size and trial duration. dement geriatr cog disord 10: 534-540. doi: 10.1159/000017201


burns a, o’brien j, bap dementia consensus group, auriacombe s, ballard c, broich k, bullock r, feldman h, ford g, knapp m, mcgaddon a, illiffe s, jacova c, jones r, lennon s, mcKay l, orgogozo jm, purandare n, richardson m, ritchie c, thomas a, warner j, wilcock g, wilkinson d. british association for psychopharmacology. 2006. clinical practice with anti-dementia drugs: a consensus statement from british association for psychopharmacology. j psychopharmacology 20: 733-755.


frank lb. 2007. commentary on “health economics and the value of therapy in alzheimer’s disease.” report from the alzheimer’s association research roundtable on patient-reported outcomes and dementia research. alzheimers demen 3: 162-165.

gammanpila up, burns a, heller rf, purandare n. 2007. what are the benefits of cognitive enhancers for alzheimer’s disease: use of population impact measures. BMC geriatr 7: 25. doi: 10.1186/1471-2318-7-25.


hirsch c. 2006. galantamine was effective in mild to moderate Alzheimer disease. ACP J Club 145: 50.


Rockwood k, Fay s, jarrett p, asp e. 2007b. Effect of galantamine on verbal repetition in AD: a secondary analysis of the VISTA trial. neurology 68: 1116-1121.

Rockwood k, Fay s, song x, MacKnight c, gorman m. 2006. attainment of treatment goals by people with Alzheimer’s disease receiving galanta-


