Alzheimer's & Dementia 13 (2017) 1098-1106

Featured Article

Clinical meaningfulness of Alzheimer's Disease Assessment Scale–Cognitive subscale change in relation to goal attainment in patients on cholinesterase inhibitors

Kenneth Rockwood[a,b,*], Susan E. Howlett[b,c], Deborah Hoffman[d], Rachel Schindler[d], Arnold Mitnitski[a,b]

[a]DGI Clinical Inc., Halifax, Nova Scotia, Canada
[b]Division of Geriatric Medicine, Dalhousie University, Halifax, Nova Scotia, Canada
[c]Department of Pharmacology, Dalhousie University, Halifax, Nova Scotia, Canada
[d]Pfizer Global Outcomes Research, New London, Connecticut, USA

Abstract

Introduction: The clinical meaningfulness of Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog) subscale change is disputed. We compared 2- to 4-point ADAS-Cog changes with changes in Goal Attainment Scaling (GAS) and everyday function across initial ADAS-Cog scores and treatment responses.

Methods: This exploratory analysis evaluated mild-moderate Alzheimer's disease patients treated with donepezil (12 months) or galantamine (8 months). Clinical meaningfulness was defined as concomitant ADAS-Cog and GAS changes of ±3 points and/or functional improvement.

Results: Patients with ≥3-point ADAS-Cog improvement significantly improved on GAS but not on standard tests of everyday function. ADAS-Cog "no change" (≤±3 points) was seen with mean GAS improvement. Initial ADAS-Cog improvement made endpoint improvement (ADAS-Cog 3 points and GAS 1 point) more likely (odds ratio = 6.9; 95% confidence interval = 2.5–19.5). In contrast, initial deterioration made endpoint improvement unlikely (0.33; 0.14–0.64).

Discussion: ADAS-Cog improvement and no change were each associated with GAS improvement. Initial ADAS-Cog worsening was unlikely to result in later improvement.

Clinical trial registration number: ISRCTN26167328.

Keywords: Alzheimer; Clinical meaningfulness; ADAS-Cog; Goal Attainment Scaling; Function; Clinical trial; Dementia; Cholinesterase inhibitor; Donepezil; Galantamine

1. Introduction

Treating Alzheimer's disease is proving to be a difficult challenge. Many trials have failed in phase III, raising questions about both the compounds and how they are being tested [1]. How we evaluate cognition in clinical trials is crucial, even in early stage dementia where hopes for a biomarker-dominant assessment have not yet materialized [1]. The Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog) has been used both in clinical trials and to characterize the natural history of Alzheimer's disease [2–4]. It remains widely used, despite concerns about its psychometric properties [5–7], especially in the early stages of dementia [8,9].

The ADAS-Cog informs the clinical meaningfulness of treatments [10]. By consensus, a net difference between treatment arms of 3 or 4 points commonly is considered clinically meaningful [10,11] leading to claims of ineffectiveness when it is not met [10]. Attention has been drawn to the shaky

*Corresponding author. Tel.: +1-902-473-8687; Fax: +1-902-473-1050.
E-mail address: krockwood@dgiclinical.com

http://dx.doi.org/10.1016/j.jalz.2017.02.005
1552-5260/

ground on which such consensus stands, based as it is on inferences about the extent of deterioration in the untreated natural history [12] and without regard to varying baseline scores, from which varying degrees of change might be expected, given acceleration of impairment across the disease course [1,13]. In contrast, there is evidence that meeting personalized goals is both significantly responsive to change and viewed by patients and carers as clinically meaningful [13–15].

Our objective was to compare the direction and degree of change in ADAS-Cog scores with clinically meaningful changes in Goal Attainment Scaling (GAS) and in everyday function in people using cholinesterase inhibitors for mild-moderate Alzheimer's disease. Specifically, we aimed to (1) describe changes in ADAS-Cog scores in relation to changes in GAS and in daily function; (2) evaluate the impact of varying the ADAS-Cog change cutpoints by $\pm 2, 3$, or 4 points; (3) evaluate the effect of baseline cognitive impairment on the response to treatment; (4) evaluate which goal domains were most related to ADAS-Cog change; and (5) determine whether initial treatment effects (in the first 8–12 weeks) were associated with later outcomes (at 32–36 weeks).

2. Methods

2.1. Patients and setting

This is an exploratory analysis of two multicenter Canadian cholinesterase inhibitor clinical trials [14,15]. Each used the ADAS-Cog [16], tested everyday function, and measured clinical meaningfulness through tracking attainment of personalized treatment goals in patients with mild-moderate disease. The Atlantic Canada Alzheimer's Disease Investigation of Expectations (ACADIE) study was a 52-week open-label, prospective study of donepezil (no placebo group) [14]. The Video-Imaging Synthesis of Treating Alzheimer's disease (VISTA) study was a 32-week multicentered, double-blinded, placebo-controlled trial of galantamine [15].

2.2. Measures

Both trials staged baseline cognitive function with the Mini–Mental State Examination (MMSE) score [16]. Both assessed cognition using the 70-point ADAS-Cog, with 11 tasks measuring memory, language, praxis, attention, and other cognitive abilities [17]. Higher scores indicate worse cognition. Baseline severity was classified as follows: least impaired (scores = 8–19), two intermediate categories (20–24 and 25–30), and most impaired (31–54).

VISTA and ACADIE used different measures of function. ACADIE used three scales: (1) the 30-point Functional Activities Questionnaire evaluated six instrumental activities of daily living (IADLs) and four high-order functions (skilled games/hobbies, current events, reading, and remembering appointments) [18]; (2) the 31-point Lawton-Brody IADL scale evaluated telephone use, shopping, food preparation, housekeeping, laundry, transportation, medications, and finances [19]; and (3) the 30-point Lawton-Brody Physical Self-Maintenance Scale evaluated toileting, feeding, dressing, grooming, walking, and bathing [19]. For each scale, a higher score indicates worse function. In VISTA, the 100-point Disability Assessment for Dementia (DAD) [20] (a higher score is better) rated eating, meal preparation, telephoning, hygiene, dressing, medication, finances, correspondence, leisure, and housework. Each item is scored for planning, initiation, and performance. For comparability, we rescaled them. In VISTA, we retained the DAD scoring, expressed as a percentage of the total number of items sampled, similar to an earlier integrated analysis across trials [21]. In ACADIE, we standardized the score as $(1 - [\text{FAQ} + \text{IADL} + \text{PSMS}])/91 \times 100$ (where FAQ is the Functional Activities Questionnaire and PSMS is the Physical Self-Maintenance Scale), also yielding a 0 to 100 range, with a higher score being better.

Patient/carer GAS, the coprimary outcome in each study, was used here with function to define clinical meaningfulness [22]. Personalized goals were set at baseline by the caregiver/patient using a five-step process in which problem areas were selected, goals were set, varying degrees of plausible worsening or improvement were specified, problems were weighted, and then each was scored at follow-up. Personalization means that subjects had their own individual outcome based on their expected level of performance, allowing both different goals and different number of goals between subjects. GAS is a change score and in both trials was scored as 0 at baseline, so that when the score remains 0 there is no change. Scores >1 indicate improvement and scores <1 indicate worsening. For both function and GAS, we report change from baseline at the endpoint (8 months in VISTA and 9 months in ACADIE) [14,15]. Goals can be analyzed as a group or by goal domain, classified as cognition, executive function, behavior, daily function, and physical manifestations [12,23].

2.3. Analysis

ADAS-Cog change scores were calculated by subtracting the baseline scores from the endpoint scores. Each subject was classified as *improved* (ADAS-Cog change scores ≤ -3), *no change* (ADAS-Cog change scores between -3 and 3), or *worse* (ADAS-Cog change scores ≥ 3). Analyses were repeated using 2- and 4-point intervals (e.g., as improved = ADAS-Cog change scores ≤ -2, no change = ADAS-Cog change scores between -2 and 2, and worse = ADAS-Cog change scores ≥ 2).

All measures at all time points were tested for outliers using Chauvenet's criterion [24] and for normality using the Shapiro-Wilk test. Correlations are Pearson's correlation coefficient (r). Two-way analysis of variance was used to test mean changes in function and GAS between ADAS-COG cutpoint groups and again by ADAS-Cog change groups; the comparisons were baseline versus endpoint. One-way