

Utilization of Observational Data as a Proxy Cohort for Comparison Purposes with Open-Label Study Results: An Example from Alzheimer's Disease

C. Reed¹, M. Happich¹, J. Raskin², A. Tockhorn-Heidenreich¹, M. Belger¹

1. Eli Lilly and Company Limited, Windlesham, UK; 2. Eli Lilly and Company, Indianapolis, IN, USA

Corresponding Author: Catherine Reed, Eli Lilly and Company, Erl Wood Manor, Sunninghill Road, Windlesham, Surrey, GU20 6PH, UK, Phone: 0044 (0) 1276 483243; Fax: 0044 (0) 1276 483192; Email: reed_catherine@lilly.com

J Prev Alz Dis 2019;2(6):90-99
Published online February 11, 2019, <http://dx.doi.org/10.14283/jpad.2019.4>

Abstract

BACKGROUND: Randomized placebo-controlled trials in the development of disease-modifying treatments for Alzheimer's disease are typically of short duration (12–18 months), and health economic modeling requires extrapolation of treatment effects beyond the trial period.

OBJECTIVES: To investigate whether observational data can be used to extrapolate data from open-label trials, we compared outcomes (cognition, function, behavior) over 36 months for patients with mild Alzheimer's disease dementia in the GERAS observational study (proxy for placebo control) with those of the mild Alzheimer's disease population on active treatment (solanezumab) in two 18-month randomized placebo-controlled trials (EXPEDITION and EXPEDITION2) and the additional 18-month open-label extension study (EXPEDITION-EXT).

DESIGN & SETTING: Analysis of longitudinal data from patients with mild Alzheimer's disease dementia in the GERAS observational study (conducted in France, Germany and the United Kingdom) and the EXPEDITION program (conducted in Europe, North America, South America, Asia and Australia).

PARTICIPANTS: European and North American community-living patients, aged ≥ 55 years, with probable Alzheimer's disease dementia and their caregivers. Mild Alzheimer's disease dementia was defined as a Mini-Mental State Examination score of 20–26 in EXPEDITION and 21–26 in GERAS.

INTERVENTION: Active treatment in both randomized placebo-controlled trials and the open-label extension study was intravenous solanezumab 400 mg every 4 weeks. Patients in GERAS were receiving treatment as part of standard care.

MEASUREMENTS: Between-group differences for changes from baseline over 36 months in cognitive function, ability to perform activities of daily living, and behavioral and psychological symptoms of dementia were assessed using models stratified by propensity score.

RESULTS: At baseline, patients and caregivers participating in GERAS were significantly older than those in the EXPEDITION studies, and the GERAS patient cohort had fewer years of education and a shorter time since diagnosis of Alzheimer's disease. The baseline mean Mini-Mental State Examination score of the GERAS cohort was significantly higher (indicating better cognition) than that of patients receiving placebo or active treatment in the pooled EXPEDITION studies. Baseline functional ability scores were significantly lower for the GERAS cohort, indicating poorer functioning. Propensity score stratification achieved a good balance in the baseline variables

between GERAS and the two EXPEDITION arms. Over 18 months, least squares mean changes from baseline in outcome measures were similar in the GERAS cohort and the pooled placebo groups from the randomized controlled trials. Also, the 18-month results for the comparison between the GERAS cohort and the pooled active treatment groups from the randomized controlled trials were generally similar to those reported for the comparison with the control group in the randomized trial. Comparison of active treatment (EXPEDITION-EXT) and observational study (GERAS, as proxy control) results over 36 months of the open-label trial showed a significantly smaller decline in activities of daily living (instrumental and basic) in the active treatment group, reflecting better functioning, but no between-group differences at 36 months for cognitive function or behavioral and psychological symptoms of dementia.

CONCLUSIONS: Comparing results from clinical trials and observational studies (real-world data) may be a useful methodological approach for informing long-term outcomes in Alzheimer's disease drug development and could be used to inform health economic modeling. Further research using this methodological approach is needed.

Key words: Alzheimer's disease, cognition, function, real-world evidence, health economic modeling.

Introduction

Alzheimer's disease (AD) is a chronic, progressive disease characterized by cognitive and functional decline and behavioral changes prompted by neuropsychiatric symptoms along a disease continuum (1), although the clinical course is highly variable among affected individuals in the real-life setting (2). Drug development in AD is currently focused on disease-modifying treatments that target the underlying pathophysiology and prevent, delay or slow disease progression at an early stage in the disease course (3), although a successful disease-modifying therapy for AD is not yet available. Longer trial designs than 12-18 months may be needed to reliably detect changes in early AD and demonstrate treatment effects for drugs

that are expected to affect the underlying disease process and slow the rate of decline (4). As open-label extension studies of randomized controlled trials (RCTs) convert all patients remaining within the trial to active treatment, any long-term data collected in these studies cannot be compared directly with a control arm. Hence, demonstrating persistence of effect in the long term beyond the clinical trial period is another key challenge that needs to be addressed.

Health economic models are used by policy makers and payers to assess new treatments and make decisions about the allocation of healthcare resources (5). Health economic models generally rely on treatment data from RCTs, requiring extrapolation of treatment effects beyond the trial period, particularly when considering the long-term effectiveness of disease-modifying treatment in AD. Assumptions need to be made on whether the treatment effect continues to develop along the same trajectory as observed during the trial period or whether the trajectory changes after the end of the trial (6). Hence, extrapolation analyses that combine trial data and real-world evidence could help to improve the accuracy of economic modeling efforts (6).

Previous analysis of pooled data from the placebo arms of two RCTs (EXPEDITION and EXPEDITION2) has shown that patients with probable mild AD dementia had similar outcomes to an observational cohort with mild AD dementia from the same geographical region after adjustment for differences in baseline characteristics (7). At 18 months, the declines in cognitive, functional, and behavioral and psychological symptoms of dementia (BPSD) were similar in both groups (7). These findings suggest that data on disease progression in observational studies are complementary to those from RCTs and that an observational cohort could be used as a control group or proxy for placebo. This methodological approach may be useful for modeling the long-term effects of AD treatments.

To provide further evidence for the critical assumptions in health economic modeling, we examined how observational data as real-world evidence could be used as a proxy for a placebo control in the open-label extrapolation of outcome data from RCTs beyond the traditional 18-month trial period. For this, we compared outcomes (cognition, function and behavior) over 36 months between patients with mild AD dementia in the GERAS observational study (proxy for placebo) and patients who were on active treatment (solanezumab) in EXPEDITION and EXPEDITION2 and in the open-label extension study following these two RCTs (EXPEDITION-EXT), for which no placebo group was available.

Methods

Study designs

EXPEDITION program

The EXPEDITION-EXT study (ClinicalTrials.gov identifier NCT01127633) was an open-label extension study offered to patients with AD who completed participation in either EXPEDITION or EXPEDITION2 (ClinicalTrials.gov identifiers NCT00905372 and NCT00904683, respectively), two identical phase 3, 18-month, randomized, double-blind, placebo-controlled studies investigating solanezumab treatment in patients with mild-to-moderate AD (8). The EXPEDITION program was designed to assess the effect of solanezumab, a humanized immunoglobulin G1 monoclonal antibody directed against the amyloid-beta ($A\beta$) peptide in the cerebral cortex and hippocampus, in the treatment of patients with mild-to-moderate AD (8); the accumulation of aggregated $A\beta$ peptide has been implicated in the early pathogenesis of AD (9). Briefly, these two EXPEDITION studies included community-dwelling patients who were ≥ 55 years old, met the criteria for probable AD (National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association [NINCDS-ADRDA] criteria) (10) and had a Mini-Mental State Examination (MMSE) score of 16–26 (11). Mild AD and moderate AD were defined as screening visit MMSE scores of 20–26 and 16–19, respectively (12, 13). Patients with vascular dementia or a current serious or unstable illness were excluded, and patients were required to have a caregiver who spent ≥ 10 hours/week with them and accompanied them to study visits (see Protocol in the online supplementary material for Doody et al. (8)).

Patients were recruited from Europe, North America, South America, Asia and Australia between May 2009 and December 2010 (8). The trial protocols were approved by ethical review boards at each study center, and written informed consent was provided by all participants. Patients were randomized to receive intravenous solanezumab 400 mg or placebo once every 4 weeks for 18 months. They were also allowed to continue with stable doses of standard-of-care AD treatments (e.g., acetylcholinesterase inhibitors and/or memantine) throughout the studies. Data from only the European and North American populations were used in the current analysis to minimize region-specific heterogeneity (7).

During the 18-month EXPEDITION-EXT, all participating patients received active treatment: i.e., patients who received solanezumab in the placebo-controlled period remained on solanezumab, and patients who received placebo were switched to solanezumab (12). However, patients and site personnel remained blinded to the patients' original treatment assignment

(solanezumab or placebo) in the placebo-controlled studies (12).

Patients included in the present analysis were those with mild AD dementia (MMSE score 20–26) at the time of enrolment (13).

GERAS

GERAS was an 18-month, prospective, observational study of the costs and resource use associated with AD for patients and caregivers in France, Germany and the UK, with the study extended to 36 months in France and Germany. The study design, methods and baseline characteristics in all three countries have been described previously (14). Briefly, the study included community-dwelling patients aged ≥ 55 years who had probable AD dementia (NINCDS-ADRDA criteria) (10) and an MMSE score of ≤ 26 (11) who presented within the normal course of care. Those with other potential causes of dementia were excluded from the study (see Wimo et al. (14)). Each patient was also required to have an informal caregiver who agreed to participate in the study and undertake responsibility for the patient for ≥ 6 months of the year.

Patients were enrolled between October 2010 and September 2011. All patients (or their legal representatives) and caregivers gave written informed consent before entering the study, which was approved by ethical review boards in each country according to country-specific regulations.

During the GERAS study, AD treatment could be prescribed according to standards of care, and treatment decisions were at the discretion of the physician and patient.

Patients included in the present analysis were those with mild AD dementia (MMSE score 21–26) at the time of enrolment, consistent with UK clinical guidelines (15).

Rationale for this analysis

In secondary outcome analyses of pooled data from the mild AD dementia populations in the two EXPEDITION studies, Siemers et al. (13) demonstrated a treatment effect of solanezumab versus placebo over 18 months, with significantly less cognitive decline in MMSE and the cognitive subscale of the Alzheimer Disease Assessment Scale (ADAS-Cog) and less functional decline in Alzheimer Disease Study Activities of Daily Living Inventory (ADCS-ADL)—instrumental activities of daily living (iADL). Additionally, Reed et al. (7) reported that, after adjustment for baseline characteristics, disease progression over 18 months in the European and North American placebo group from this pooled mild AD population was similar to that in the pooled mild AD dementia cohorts from France, Germany and the UK in the GERAS study.

As only active treatment was studied in the open-

label extension, we wanted to explore whether an observational study cohort treated with standard of care could be used as a control group for the active comparison. To demonstrate suitability as a control group, we first aimed to determine whether 18-month outcomes in the European and North American patients with mild AD dementia who received placebo in the EXPEDITION and EXPEDITION2 RCTs (pooled groups: EXP-placebo) were similar to those in GERAS participants with mild AD dementia. Then we verified whether the differences between the active treatment and placebo arms detected within the RCTs at 18 months by Siemers et al. (13) were also apparent when GERAS participants were used as a proxy for the pooled placebo group. The overall aim of the current study was to compare the 36-month clinical outcomes between the matched mild AD dementia groups in EXPEDITION-EXT who had received active treatment in EXPEDITION and EXPEDITION2 (EXP-active) and those in mild AD dementia GERAS participants (control).

Data collected

Data collected in all studies included patient and caregiver demographics as well as patient clinical characteristics at baseline, such as comorbidities, medication use, cognitive function, ability to perform basic activities of daily living (bADL) and iADL, BPSD and health-related quality of life (HRQoL). Time spent by the caregiver on assisting the patient with bADL, iADL and supervision was also recorded.

Cognitive function was assessed using the MMSE (11) and the 14-item subscale of the ADAS-Cog (ADAS-Cog14) (16); these scales were completed by the investigators. The MMSE total score ranges from 0 to 30, with lower scores indicating poorer cognition. The ADAS-Cog14 has a score range of 0–90, with higher scores indicating poorer cognition.

Patient functional ability was assessed using the 23-item ADCS-ADL (17). This caregiver-rated scale assesses bADL, such as dressing or bathing, and iADL, which are more complex everyday tasks, such as cooking, driving or managing finances, providing separate subscale scores for bADL (ADCS-bADL; range 0–22) and iADL (ADCS-iADL; range 0–56), with lower scores indicating poorer functioning.

Patient BPSD, and caregiver distress caused by these symptoms, were evaluated by caregivers using the 12-item Neuropsychiatric Inventory (NPI-12) (18). The NPI-12 total score ranges from 0 to 144, with higher scores representing worse behavior problems.

Time spent by the caregiver on assisting the patient with bADL, iADL and supervision was recorded using the Resource Utilization in Dementia instrument (adapted from version 3.0) (19).

Information on HRQoL was collected for both the patient and the caregiver using the EuroQoL-5

Table 1. Demographic and baseline characteristics of patients and caregivers in the two arms of the EXPEDITION studies (pooled mild AD population) and the mild AD cohort of the GERAS study

Characteristic*	EXP-placebo (N=466)	EXP-active (N=455)	GERAS (N=566)	p-value EXP-placebo vs GERAS	p-value EXP-active vs GERAS
<i>Patients</i>					
Age, years	73.2 ± 8.1	74.2 ± 8.2	77.3 ± 6.9	<0.001	<0.001
Male	227 (48.7)	235 (51.6)	295 (52.1)	0.553	0.989
Body mass index, kg/m ²	25.8 ± 4.4	26.1 ± 4.5	25.4 ± 3.8	0.218	0.025
Living alone	52 (11.8)	53 (12.4)	103 (18.6)	0.002	<0.001
Time since AD diagnosis, years	2.1 ± 2.0	2.0 ± 1.9	1.7 ± 2.0	0.001	0.009
Years of education	13.3 ± 3.7	13.0 ± 3.8	11.1 ± 3.3	<0.001	<0.001
AD medication	412 (88.4)	393 (86.4)	478 (84.5)	0.185	0.690
AD medication category				<0.001	<0.001
AChEI only	247 (53.0)	234 (51.4)	411 (72.6)		
Memantine only	29 (6.2)	16 (3.5)	41 (7.2)		
AChEI + memantine	136 (29.2)	143 (31.4)	26 (4.6)		
No AD treatment	54 (11.6)	62 (13.6)	88 (15.5)		
<i>Caregivers</i>					
Age, years	65.0 ± 12.5	65.0 ± 12.8	68.1 ± 11.6	<0.001	<0.001
Male	178 (38.4)	146 (32.5)	178 (31.5)	0.119	0.176
Caregiver relationship, spouse	318 (68.5)	297 (66.1)	399 (70.6)	0.781	0.067
Caregiver time spent on patient, hours in past month					
Basic ADL	9.1 ± 30.0	11.7 ± 43.9	16.4 ± 44.7	<0.001	<0.001
Instrumental ADL	58.9 ± 76.3	68.2 ± 96.2	61.0 ± 83.1	0.800	0.431
Supervision	25.7 ± 96.6	38.5 ± 118.6	48.4 ± 123.6	<0.001	0.017
Total	90.3 ± 134.9	112.4 ± 169.4	121.6 ± 174.2	0.021	0.326
NPI Caregiver distress score	5.7 ± 6.3	4.9 ± 6.0	6.4 ± 6.2	0.072	<0.001

Abbreviations: AChEI, acetylcholinesterase inhibitor; AD, Alzheimer's disease; ADL, activities of daily living; EXP, EXPEDITION study; NPI, Neuropsychiatric Inventory; * Data are presented as mean ± standard deviation or n (%) for patients/caregivers with non-missing data.

Dimensions (EQ-5D) (20). Both UK population-based index scores and visual analog scale (VAS) scores were recorded.

Given the differing nature of the GERAS and EXPEDITION studies, some assessments were made more frequently in the EXPEDITION trials. Data were generally collected at baseline and every 6 months up to 36 months, although ADAS-Cog14 and ADCS-ADL data were collected only at baseline and at 18 and 36 months in the GERAS study. The analyses reported here use those time points common to both studies.

Statistical analysis

Demographic and baseline characteristics were summarized using descriptive statistics (mean and standard deviation [SD] or frequency) based on non-missing observations. Differences at baseline between the EXP-placebo and GERAS cohorts, and between the EXP-active and GERAS cohorts, were examined using the Tukey test for continuous variables (e.g., age) and the

Cochran–Mantel–Haenszel test for categorical variables (e.g., gender). However, the Mann–Whitney test was used for caregiver time comparisons at baseline between EXPEDITION and GERAS cohorts because of the skewed distribution of data.

Propensity score stratification was used to adjust for differences in baseline characteristics between the GERAS and the two EXPEDITION cohorts. Propensity scores were derived using all available baseline covariates, including interaction terms, stratified (10 strata were used) and compared across cohorts. Absolute standardized differences (21) (pooled across strata) were calculated for baseline covariates before and after propensity score adjustment; a difference of <0.1 was considered an acceptable balance between the covariates. Propensity scores were stratified with and without non-overlapping propensity scores; the primary analysis excluded patients with non-overlapping propensity scores.

A mixed-effects model for repeated measures (MMRM) approach was used to analyze differences between the EXPEDITION and GERAS studies for changes in patient

Table 2. Baseline clinical characteristics of patients with mild AD in the two arms of the EXPEDITION studies and the mild cohort of the GERAS study

Characteristic*	EXP-placebo (N=466)	EXP-active (N=455)	GERAS (N=566)	p-value EXP-placebo vs GERAS	p-value EXP-active vs GERAS
MMSE†	23.0 ± 2.0	22.8 ± 1.9	23.3 ± 1.6	0.005	<0.001
ADAS-Cog14‡	29.1 ± 9.2	29.8 ± 8.9	30.4 ± 7.8	0.013	0.250
ADCS-ADL total score§	64.7 ± 10.7	63.8 ± 11.4	58.4 ± 14.2	<0.001	<0.001
ADCS-bADL	20.9 ± 2.2	20.9 ± 2.0	19.8 ± 3.1	<0.001	<0.001
ADCS-iADL	43.8 ± 9.3	42.9 ± 10.2	38.5 ± 11.8	<0.001	<0.001
NPI-12	9.8 ± 11.4	8.5 ± 11.0	10.2 ± 10.7	0.588	0.015
EQ-5D index score{	0.79 ± 0.19	0.80 ± 0.19	0.71 ± 0.24	<0.001	<0.001

Abbreviations: AD, Alzheimer's disease; ADAS-Cog14, Alzheimer's Disease Assessment Scale-cognitive subscale; ADCS-ADL, Alzheimer's Disease Cooperative Study Activities of Daily Living; bADL, basic activities of daily living; iADL, instrumental activities of daily living; EQ-5D, EuroQol-5 Dimension; EXP, EXPEDITION study; MMSE, Mini-Mental State Examination; NPI, Neuropsychiatric Inventory; *Data are presented as mean ± standard deviation; †MMSE score range = 0–30; higher score indicates better functioning. For the EXPEDITION studies, the MMSE score was taken from the prescreening visit (visit 1); ‡ADAS-Cog14 score range = 0–90; higher score indicates greater severity; §ADCS-ADL total score range = 0–78; bADL = 0–22; iADL = 0–56; higher scores indicate better functioning; || NPI-12 score range = 0–144; higher score indicates greater severity; {EQ-5D index score (UK population-based) range = 0–1; higher score indicates better HRQoL.

clinical measures (MMSE, ADAS-Cog14, ADCS-iADL, ADCS-bADL, NPI-12) over 18 months (EXP-placebo vs GERAS; EXP-active vs GERAS) and 36 months (EXP-active vs GERAS). MMRM models included patients with at least one post-baseline visit. The models included the baseline value of the outcome measure of interest, study group (GERAS or EXP), visit, visit-by-treatment group interaction and propensity score stratum. Results from the models are reported as least squares (LS) mean changes from baseline. For the 36-month analysis, the EXP-active group only includes subjects who were randomized to active treatment in the EXPEDITION and EXPEDITION2 studies; those who were randomized to placebo and then switched to active treatment for the EXPEDITION-EXT study are not included in the EXP-active versus GERAS 36-month analysis. The analysis of data from the GERAS study includes patients from all three participating countries up to 18 months and from France and Germany only between 18 and 36 months.

As sensitivity analyses, MMRM models were also run where the propensity score stratum was replaced with the propensity score and including patients with non-overlapping propensity scores.

All data were analyzed using SAS software, version 9.2 (SAS Institute, Cary, NC, USA).

Results

Of the 1495 patients enrolled in the GERAS study (France, Germany and the UK), 566 had mild AD dementia at baseline. In EXPEDITION and EXPEDITION2, a total of 466 patients from North America and Europe with mild AD dementia at baseline were randomized to placebo and 455 were randomized to active treatment.

Baseline characteristics of patients with mild AD dementia

The baseline demographics of the patients and caregivers included in our analyses are summarized in Table 1. Patients and caregivers participating in GERAS were significantly older than those in the EXPEDITION studies, and the GERAS patient cohort had fewer years of education and a shorter time since diagnosis of AD. Also, some differences were observed at baseline in the patient clinical measures of interest (Table 2). Although all patients in the analysis had mild AD dementia, the baseline mean MMSE score of the GERAS cohort was significantly higher (indicating better cognition) than that of the EXP-placebo or EXP-active treatment groups. However, the higher ADAS-Cog14 scores indicated the GERAS cohort had statistically poorer cognition at baseline than the EXP-placebo group. Functional ability scores (ADCS-bADL and ADCS-iADL) were significantly lower for the GERAS cohort, indicating poorer functioning; however, baseline scores overall indicated little impairment in bADL in any group. Also, the baseline NPI-12 scores indicated few problems due to BPSD, although they were significantly worse in the GERAS cohort than in the EXP-active treatment group.

Propensity score stratification

The absolute standardized differences (with non-overlapping propensity scores removed) between GERAS and the placebo and active treatment arms of the EXPEDITION RCTs (EXP-placebo and EXP-active, respectively) for the baseline characteristics and clinical measures before and after propensity score adjustment are shown in Figure 1A (EXP-placebo vs GERAS) and Figure 1B (EXP-active vs GERAS). The results show that propensity score stratification achieved a good balance in the baseline variables between GERAS and the two arms

Table 3. Difference between cohorts of least squares mean change from baseline (95% confidence intervals) over 18 and 36 months in patient clinical measures*

Outcome	Time point	EXP-placebo vs GERAS		EXP-active vs GERAS	
		Estimate (95% CI)	p-value	Estimate (95% CI)	p-value
MMSE†	18 months	-0.69 (-1.48–0.09)	0.084	0.95 (0.17–1.74)	0.018
	36 months	N/A		0.32 (-0.92–1.57)	0.610
ADAS-Cog14‡	18 months	0.91 (-1.12–2.94)	0.377	-2.00 (-4.09–0.08)	0.060
	36 months	N/A		0.95 (-2.37–4.28)	0.573
ADCS-bADL§	18 months	0.22 (-0.39–0.83)	0.484	0.39 (-0.30–1.08)	0.270
	36 months	N/A		1.22 (0.18–2.27)	0.022
ADCS-iADL§	18 months	0.85 (-1.01–2.72)	0.369	2.54 (0.46–4.62)	0.017
	36 months	N/A		4.18 (1.33–7.03)	0.004
NPI-12	18 months	-0.61 (-2.49–1.27)	0.525	-1.37 (-3.45–0.71)	0.196
	36 months	N/A		0.42 (-2.19–3.02)	0.754

Abbreviations: ADAS-Cog14, Alzheimer's Disease Assessment Scale-cognitive subscale; ADCS, Alzheimer's Disease Cooperative Study; bADL, basic activities of daily living; CI, confidence interval; EXP, EXPEDITION study; iADL, instrumental activities of daily living; MMSE, Mini-Mental State Examination; N/A, not assessed; NPI, Neuropsychiatric Inventory; *Mixed-effects models for repeated measures included baseline value of outcome measure, study group (GERAS or EXP), visit, visit-by-treatment group interaction and propensity score stratum; patients with non-overlapping propensity scores were excluded. p-Values in bold represent a significant difference between EXPEDITION and GERAS; †MMSE score range 0–30; higher score indicates better functioning; ‡ADAS-Cog14 score range 0–90; higher score indicates greater severity; §ADCS-bADL score range 0–22 and ADCS-iADL score range 0–56; higher scores indicate better functioning; || NPI-12 score range 0–144; higher score indicates greater severity.

of EXPEDITION; the adjusted standardized differences were <0.1 for many of the baseline variables.

Changes in outcome measures over 18 and 36 months

LS mean changes from baseline for all outcomes of interest were similar in the EXP-placebo arm and the GERAS arm over 18 months (Table 3). In the EXP-active and GERAS arm comparison, LS mean changes from baseline for MMSE and iADL were significantly different at 18 months (Table 3), indicating a slower decline in the EXP-active group than in the GERAS cohort (Figure 2).

Table 3 and Figure 2 show the LS mean change from baseline for the different outcome measures over 36 months in the EXP-active and GERAS arms. For MMSE, the decline from baseline (worsening cognition reflecting disease progression) was similar between the EXP-active and GERAS (control) groups, with no significant difference between the two groups at 36 months (Table 3; Figure 2A). The significant divergence between groups seen at 18 months did not continue, and the lines showed a visually parallel trajectory from 24 months (Figure 2A). For ADAS-Cog14 (Figure 2B), there was again some visual divergence between the two groups at 18 months, with a smaller increase from baseline in the EXP-active group (reflecting less cognitive decline); however, the between-group difference in LS mean change from baseline was not significant at 18 or 36 months (Table 3).

For patient functional ability (ADCS-bADL and ADCS-iADL) (Table 3; Figures 2C and 2D), the divergence between the EXP-active and GERAS groups seen at 18 months (significant for iADL) had increased at 36 months, showing a larger decrease from baseline in the GERAS

group (reflecting poorer functioning) that was significant between groups at 36 months for both ADCS-iADL and ADCS-bADL.

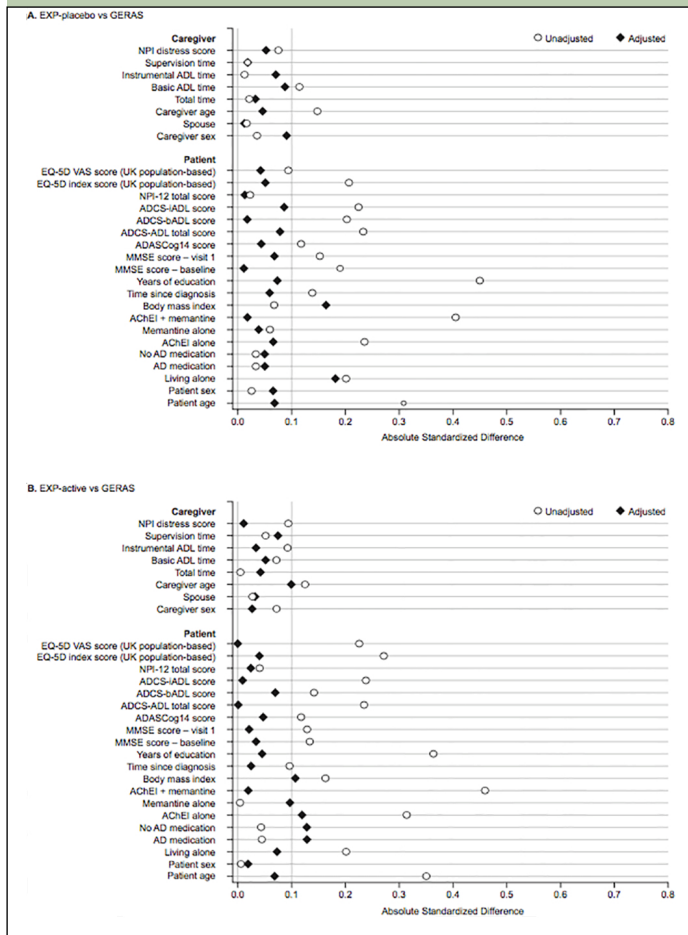
For NPI-12, the LS mean change from baseline was not significantly different between the EXP-active and GERAS groups at 18 or 36 months (Figure 2E, Table 3), showing similar worsening of BPSD among patients in both studies.

Sensitivity analyses, in which MMRM models were run replacing the propensity score stratum with the propensity score and including patients with non-overlapping propensity scores, gave similar results (see Supplementary Tables).

Discussion

Our analysis demonstrated that observational data as real-world evidence can be used to extrapolate findings from clinical trials when collecting longer term open label data using an example from AD. The changes in cognition, function and BPSD over 36 months in patients with mild AD dementia receiving active treatment in the EXPEDITION and EXPEDITION2 RCTs, and their open-label extension study (EXPEDITION-EXT), and in an observational study cohort (serving as a proxy control group) showed a significant difference in functional decline, but not cognitive decline, between the studies at 36 months, after controlling for baseline differences.

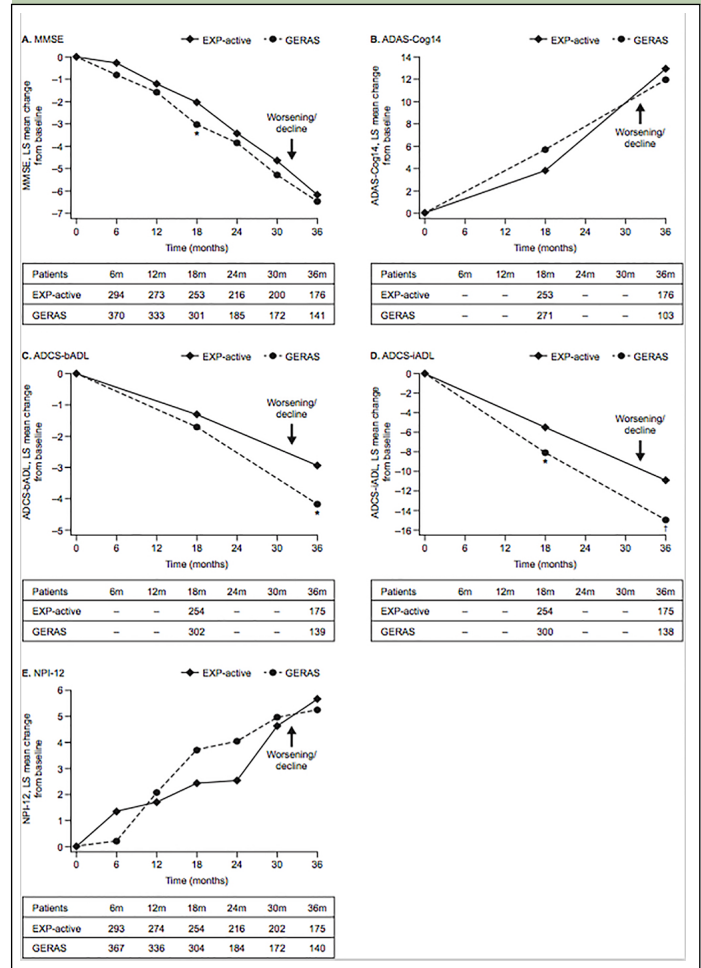
Figure 1. Absolute standardized differences for baseline covariates before and after propensity score adjustment with non-overlapping propensity scores removed for (A) EXP-placebo vs GERAS, (B) EXP-active vs GERAS



Abbreviations: AChEI, acetylcholinesterase inhibitor; AD, Alzheimer's disease; ADAS-Cog14, Alzheimer's Disease Assessment Scale-cognitive subscale; ADCS, Alzheimer's Disease Cooperative Study; ADL, activities of daily living; bADL, basic activities of daily living; EQ-5D, EuroQol-5 Dimension; EXP, EXPEDITION study; iADL, instrumental activities of daily living; MMSE, Mini-Mental State Examination; NPI, Neuropsychiatric Inventory; VAS, visual analog scale. Note: In the EXPEDITION studies, MMSE was measured at prescreening (Visit 1; to assign patients into mild or moderate AD dementia groups) and at baseline (randomization to treatment).

Although the outcome comparisons showed significance on ADCS-iADL and ADCS-bADL at 36 months, our findings were mostly non-significant. This can be attributed to a lack of efficacy of the active treatment rather than an issue with the methods used: the active treatment investigated in the EXPEDITION and EXPEDITION2 trials did not meet their co-primary objectives nor did the subsequent EXPEDITION3 which had the primary endpoint of decreasing cognitive decline at 18 months (assessed using ADAS-Cog14 in 2129 patients with mild AD (9)). Instead our findings highlight the benefit of collecting real-world data to support the clinical development of new treatments.

Figure 2. Least squares (LS) mean change from baseline in scores for (A) MMSE, (B) ADAS-Cog14, (C) ADCS-bADL, (D) ADCS-iADL and (E) NPI-12 over 36 months for the mild AD cohorts in the EXPEDITION (EXP-active) and GERAS studies (excluding non-overlapping propensity scores)



Cognitive function was assessed using the Mini-Mental State Examination (MMSE; score range 0–30) and the cognitive subscale of the Alzheimer's Disease Assessment Scale (ADAS-Cog14; score range 0–90). Functional ability was assessed using the Alzheimer's Disease Study Activities of Daily Living Inventory (ADCS-ADL), with separate scores for basic activities of daily living (ADCS-bADL; score range 0–22) and instrumental activities of daily living (ADCS-iADL; score range 0–56). Patient behavioral and psychological symptoms were assessed using the Neuropsychiatric Inventory (NPI-12; score range 0–144). * $p < 0.05$ and † $p < 0.01$ indicate a significant difference between the EXP-active and GERAS cohorts in the LS mean change from baseline at this timepoint; the difference between cohorts at all other timepoints was non-significant. Note: patients in the GERAS arm were from France, Germany and the UK up to 18 months and from France and Germany only at 24–36 months.

Evaluating the long-term impact of a treatment, especially in a slow progressive disease within an elderly population, such as is common in AD, can be difficult as study dropout rates (due to death, institutionalization, comorbidities, etc.) tend to be high. Being able to use real-world observational data as a proxy for a placebo control group provides a useful method for evaluating the long-term impact of a treatment from studies, such as EXPEDITION-EXT, where no control/placebo group is available or where long-term use of placebo would be considered unethical. This type of analysis demonstrates

how real-world evidence can be used to support long-term outcomes in AD, where one of the major challenges is to measure outcomes some time after an intervention (such as starting treatment) early in the disease process (6).

Functional decline (with regards to both ADCS-iADL and ADCS-bADL scores) over 36 months was significantly less in the group that received active treatment (solanezumab) than in the GERAS (control) group. Over 3 years, the decline in ADCS-iADL was approximately 15 points in the GERAS group and 11 points in the EXPEDITION group; the decline can be considered clinically meaningful in both groups, as can the between-group difference of 4 points at 36 months (22). Although these findings are consistent with those of Siemers et al. (13), who found a significant treatment effect of solanezumab versus placebo on ADCS-iADL at 80 weeks in patients with mild AD dementia in their secondary analysis of EXPEDITION and EXPEDITION2, this was not reproduced in the EXPEDITION3 trial (9).

Basic functional abilities declined over 36 months in both groups, with the between-group difference in change scores for ADCS-bADL reaching statistical significance at 36 months. In contrast to ADCS-iADL scores, however, there was no significant difference in ADCS-bADL scores between the EXP-active and GERAS arms at 18 months, consistent with the findings of Siemers et al. (13), who reported no difference between the solanezumab and placebo groups on the change in ADCS-bADL score at 80 weeks in the EXPEDITION studies. A possible explanation for our differing findings with regards to ADCS-iADL and -bADL scores is that our study cohorts showed little impairment in bADL at baseline, and the ADCS-bADL score decreased by only 3–4 points over 3 years in both groups in the current study. This is consistent with previous reports that iADL become impaired before bADL (23) and that iADL measures are more sensitive to change than bADL measures in patients with mild AD (17). Research has also indicated that cognition has a stronger correlation with iADL than with bADL in patients with mild AD dementia (24). Nevertheless, this is the first demonstration of a slowing of decline in bADL scores over 36 months with an active treatment for AD compared with a placebo proxy control group in patients with mild AD.

Our finding of a significant between-group difference in functional decline but not cognitive decline over 36 months was unexpected given that studies of disease progression in patients with mild AD have suggested that cognitive decline generally precedes and predicts functional decline (25). Possible reasons for this include that patients in the GERAS and EXPEDITION cohorts in this study had an MMSE score of ~23 at baseline (towards the middle of the mild MMSE score spectrum [21–26]) and would probably have progressed to moderate cognitive decline (MMSE score <20) within 3 years; solanezumab, as a disease-modifying drug, was

shown to have no effect on cognition in patients with moderate AD dementia in the EXPEDITION studies (8). Also, although patients in the GERAS study received treatment as part of standard care, this treatment may have been more rigorously applied than usual because they were participating in the study. Furthermore, in the early stages of AD, a patient's inability to perform ADLs may be more apparent than cognitive loss to caregivers (who rate patient's functional capability on the ADCS-ADL) (22).

Although the pattern of changes in NPI varied over time for the two groups (as seen in Figure 2E), the overall change in BPSD was small (LS mean increase of about 5–6 points in NPI-12 total score from a low baseline score), with no significant difference between the EXP-active group and the GERAS cohort at 18 or 36 months. Most patients with AD dementia exhibit BPSD, which can be present in the early prodementia stages and increase over time, although the course tends to be variable and episodic (26).

Given the restricted duration of RCTs in AD, interest in using real-world data generated during routine clinical practice to complement the data provided by RCTs is growing (27). Such real-world evidence can inform the application of evidence from RCTs to healthcare decision making (28). The approach used in the current analysis is just one of a number of different approaches which could be taken, including matching algorithms, entropy balancing, and applying different inclusion and exclusion criteria. We chose two different approaches of propensity score adjustment to establish the concept we were trying to demonstrate. Methods for using real-world evidence in health economic models need to be developed further. The Real world Outcomes across the Alzheimer's Disease spectrum for better care: Multimodal data Access Platform (ROADMAP) project is a public-private partnership in Europe formed under the Innovative Medicine Initiative, which aims to explore the usability of all data resources in the decision-making process and develop efficient uses of real-world evidence for the benefit of patients with AD and their caregivers (see <https://roadmap-alzheimer.org/>).

There is a need for suitable real-world evidence in the valuation of emerging new AD treatments currently in development by key decision makers including for HTA. Utilizing observational data alongside clinical trials will allow the exploration of assumptions in economic models around extrapolation of treatment effects beyond the typical RCT period as well as other relevant information such as health related quality of life and real-world cost drivers.

Our analysis has several strengths. The rationale for our study was that patients with mild AD dementia in the EXP-placebo arm and the GERAS cohort would exhibit similar disease progression over 18 months (as reported by Reed et al. (7)) and that differences over 18 months would be apparent between the GERAS cohort

and the EXP-active arm and similar to those seen for the comparison between solanezumab and placebo reported by Siemers et al. (13) in their pooled analysis of the 18-month EXPEDITION and EXPEDITION2 studies. Given that our findings supported both these assumptions, we consider that using 36-month mild AD GERAS observational data as a proxy placebo group for comparison with 36-month outcomes in the mild AD EXP-active group is a valid strategy. In addition, combining this strategy with the use of propensity score stratification and adjustment to ensure the patient mix was similar between the two data sources (given the differences in key baseline characteristics between the GERAS and EXPEDITION studies, despite efforts to minimize region-specific heterogeneity by restricting analysis of EXPEDITION to patients from Europe and North America) is a novel approach that can be used to support assumptions in economic modeling. Further research is needed to look at alternative approaches for matching patients. While we used stratification on the propensity score, other approaches could be considered, including propensity score as a covariate in the models and matching on the propensity score (29). The same analyses could also be performed for validation purposes using alternative proxy placebo arms from similar observational studies e.g. ADNI although it can be challenging to find other observational studies that include the same outcome measures as in the RCT. For example, the ADNI study uses a different measure for function, so unless validated mapping algorithms exist, further uncertainty will be introduced. It is also unlikely to have such an extensive list of baseline characteristics as was available in GERAS and EXPEDITION, which can lead to issues of unmeasured confounding. Other strengths include the use of large studies of patients with mild AD dementia at baseline and the assessment of multiple, widely used, standardized outcome measures.

Some limitations must be acknowledged. As mentioned, the active treatment investigated in the EXPEDITION trials did not meet the primary endpoint of decreasing cognitive decline in a randomized, placebo-controlled phase 3 trial in patients with mild AD dementia (9). Hence, the current paper should be considered preliminary work on a methodological approach for using real-world evidence to model long-term effects of treatments. Further research is needed to replicate this methodological approach using a treatment with proven efficacy. Second, although propensity scores adjusted for baseline differences, the study discontinuation criteria may have differed between the two studies (i.e., patients in EXPEDITION-EXT may have been more impaired at study discontinuation than those in GERAS). Moreover, our analysis was limited to patients who completed 36 months of follow-up, and those who discontinued earlier (e.g., due to institutionalization or death) may have had a more severe decline in outcome measures. In addition, despite

controlling for baseline characteristics, unmeasured confounders may have impacted on our results. This may include unrecognized differences in how the trials were performed. Another limitation is that the sample size may be reduced if there is a high number of non-overlapping propensity scores (as used in the current primary analyses), as this excludes outlying individuals for whom no match is available in the control arm. Also, ADCS-ADL was measured less frequently than the other measures (at 18 and 36 months only), which may have impacted on the results as functional decline in AD is reported to be variable and non-linear (30). Finally, as we did not measure the amyloid status of patients participating in the GERAS study, it is likely that some of these patients did not have AD.

Conclusion

Our analysis demonstrates that comparing the results from RCTs and real-world evidence can be a useful methodological approach for informing long-term efficacy/effectiveness outcomes in the development of treatments for AD and could be used to inform health economic modeling. Further research using this methodological approach is needed.

Acknowledgments: The authors would like to acknowledge Deirdre Elmhirst and Gill Gummer (Rx Communications, Mold, UK) for medical writing assistance with the preparation of this manuscript, funded by Eli Lilly and Company.

Funding and disclosure statements: This study was funded by Eli Lilly and Company.

Conflict of interest: All authors are employees of Eli Lilly and Company.

Ethical standards: The study was reviewed by ethical review boards in each country according to country-specific regulations.

Open Access: This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

References

1. Aisen PS, Cummings J, Jack CR Jr, et al. On the path to 2025: understanding the Alzheimer's disease continuum. *Alzheimers Res Ther* 2017;9:60. doi: 10.1186/s13195-017-0283-5.
2. Gillette-Guyonnet S, Andrieu S, Nourhashemi F, et al. Long-term progression of Alzheimer's disease in patients under antidementia drugs. *Alzheimers Dement* 2011;7:579-592.
3. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The neuropsychiatric inventory: comprehensive assessment of psychopathology in dementia. *Neurology* 1994;44:2306-2314.
4. Schneider LS, Sano M. Current Alzheimer's disease clinical trials: methods and placebo outcomes. *Alzheimers Dement* 2009;5:388-397.
5. Hernandez L, Ozen A, DosSantos R, Getsios D. Systematic review of model-based economic evaluations of treatment for Alzheimer's disease. *Pharmacoeconomics* 2016;34:681-707.
6. Gustavsson A, Green C, Jones RW, et al. Current issues and future research priorities for health economic modelling across the full continuum of Alzheimer's disease. *Alzheimers Dement* 2017;13:312-321.
7. Reed C, Belger M, Dell'Agnello G, et al. Representativeness of European clinical trial populations in mild Alzheimer's disease dementia – a comparison of 18-month outcomes with real-world data from the GERAS observational study. *Alzheimers Res Ther* 2018;10:36. doi: 10.1186/s13195-018-0360-4.
8. Doody RS, Thomas RG, Farlow M, et al. Phase 3 trials of solanezumab for

- mild-to-moderate Alzheimer's disease. *N Engl J Med* 2014;370:311-321.
9. Honig LS, Vellas B, Woodward M, et al. Trial of solanezumab for mild dementia due to Alzheimer's Disease. *N Engl J Med* 2018;378:321-330.
 10. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34:939-944.
 11. Folstein M, Folstein S, McHugh P. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *J Psychiat Res* 1975;12:189-198.
 12. Liu-Seifert H, Siemers E, Holdridge KC, et al. Delayed-start analysis: mild Alzheimer's disease patients in solanezumab trials, 3.5 years. *Alzheimers Dement (N Y)* 2015;1:111-121.
 13. Siemers ER, Sundell KL, Carlson C, et al. Phase 3 solanezumab trials: secondary outcomes in mild Alzheimer's disease patients. *Alzheimers Dement* 2016;12:110-120.
 14. Wimo A, Reed CC, Dodel R, et al. The GERAS study: a prospective observational study of costs and resource use in Alzheimer's disease of community dwellers in three European countries – study design and baseline findings. *J Alzheimers Dis* 2013;36:385-399.
 15. National Institute for Health and Care Excellence. NICE Technology Appraisal Guidance TA217. Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease. <http://www.nice.org.uk/guidance/TA217>. 20 June 2018. Accessed October 10, 2018.
 16. Mohs RC, Knopman D, Petersen RC, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. *The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord* 1997;11:S13-S21.
 17. Galasko D, Bennett D, Sano M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. *Alzheimer Dis Assoc Disord* 1997;11(Suppl 2):S33-S39.
 18. Cummings J, Lee G, Mortsdorf T, Ritter A, Zhong K. Alzheimer's disease drug development pipeline: 2017. *Alzheimers Dement (N Y)* 2017;3:367-384.
 19. Wimo A, Wetterholm A-L, Mastey V, Winblad B. Evaluation of the healthcare resource utilization and caregiver time in anti-dementia drug trials. In: Wimo A, Jönsson B, Karlsson G, Winblad B (eds) *The health economics of dementia*. 1998. John Wiley and Sons, London, pp 465-499.
 20. Kind P. The EuroQol instrument: An index of health-related quality of life. In: Spilker B (ed) *Quality of life and pharmacoeconomics in clinical trials*, 2nd edn. 1996. Lippincott-Raven Publishers, Philadelphia, pp 191-201.
 21. Faries DE, Leon AC, Haro JM, Obenchain RL. Propensity score stratification and regression. In: Chapter 2: Analysis of observational health care data using SAS (section 2.5). 2010. SAS Institute Inc., Cary (NC), pp 27-29.
 22. Dysken MW, Sano M, Asthana S, et al. Effect of vitamin E and memantine on functional decline in Alzheimer disease. The TEAM-AD VA Cooperative Randomized Trial. *JAMA* 2014;311:33-44.
 23. Marshall GA, Amariglio RE, Sperling RA, Rentz DM. Activities of daily living: where do they fit in the diagnosis of Alzheimer's disease? *Neurodegener Dis Manag* 2012;2:483-491.
 24. Liu-Seifert H, Siemers E, Selzler K, et al. Correlation between cognition and function across the spectrum of Alzheimer's disease. *J Prev Alzheimers Dis* 2016;3:138-144.
 25. Liu-Seifert H, Siemers E, Price K, et al. Cognitive impairment precedes and predicts functional impairment in mild Alzheimer's disease. *J Alzheimers Dis* 2015;47:205-214.
 26. Lanctôt KL, Amatniek J, Ancoli-Israel S, et al. Neuropsychiatric signs and symptoms of Alzheimer's disease: new treatment paradigms. *Alzheimers Dement (N Y)* 2017;3:440-449.
 27. Atri A, Rountree SD, Lopez OL, Doody RS. Validity, significance, strengths, limitations, and evidentiary value of real-world clinical data for combination therapy in Alzheimer's disease: comparison of efficacy and effectiveness studies. *Neurodegener Dis* 2012;10:170-174.
 28. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf* 2017;26:1033-1039.
 29. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46:399-424.
 30. Mohs RC, Schmeidler J, Aryan M. Longitudinal studies of cognitive, functional and behavioural change in patients with Alzheimer's disease. *Stat Med* 2000;19:1401-1409.