ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Themed Section: The Health Economics of Alzheimer's Disease and Related Dementias

IPECAD Modeling Workshop 2023 Cross-Comparison Challenge on Cost-Effectiveness Models in Alzheimer's Disease

Ron Handels, PhD, William L. Herring, PhD, Farzam Kamgar, MSc, Sandar Aye, MSc, Ashley Tate, PhD, Colin Green, PhD, Anders Gustavsson, PhD, Anders Wimo, MD, PhD, Bengt Winblad, MD, PhD, Anders Sköldunger, PhD, Lars Lau Raket, PhD, Chelsea Bedrejo Stellick, MSc, Eldon Spackman, PhD, Jakub Hlávka, PhD, Yifan Wei, PhD, Javier Mar, PhD, Myriam Soto-Gordoa, PhD, Inge de Kok, PhD, Chiara Brück, PhD, Robert Anderson, MA, Peter Pemberton-Ross, PhD, Michael Urbich, PhD, Linus Jönsson, MD, PhD

ABSTRACT

Objectives: Decision-analytic models assessing the value of emerging Alzheimer's disease (AD) treatments are challenged by limited evidence on short-term trial outcomes and uncertainty in extrapolating long-term patient-relevant outcomes. To improve understanding and foster transparency and credibility in modeling methods, we cross-compared AD decision models in a hypothetical context of disease-modifying treatment for mild cognitive impairment (MCI) due to AD.

Methods: A benchmark scenario (US setting) was used with target population MCI due to AD and a set of synthetically generated hypothetical trial efficacy estimates. Treatment costs were excluded. Model predictions (10-year horizon) were assessed and discussed during a 2-day workshop.

Results: Nine modeling groups provided model predictions. Implementation of treatment effectiveness varied across models based on trial efficacy outcome selection (clinical dementia rating – sum of boxes, clinical dementia rating – global, mini-mental state examination, functional activities questionnaire) and analysis method (observed severity transitions, change from baseline, progression hazard ratio, or calibration to these). Predicted mean time in MCI ranged from 2.6 to 5.2 years for control strategy and from 0.1 to 1.0 years for difference between intervention and control strategies. Predicted quality-adjusted life-year gains ranged from 0.0 to 0.6 and incremental costs (excluding treatment costs) from −US$66 897 to US$11 896.

Conclusions: Trial data can be implemented in different ways across health-economic models leading to large variation in model predictions. We recommend (1) addressing the choice of outcome measure and treatment effectiveness assumptions in sensitivity analysis, (2) a standardized reporting table for model predictions, and (3) exploring the use of registries for future AD treatments measuring long-term disease progression to reduce uncertainty of extrapolating short-term trial results by health-economic models.

Keywords: Alzheimer's disease, cross-validation, decision-analytic modeling, health-economic evaluation.

VALUE HEALTH. 2025; 28(4):497–510

Highlights

- The results of our cross-comparison study suggest that the following cascade of factors explained much of the variation in Alzheimer's disease (AD) health-economic model predictions: model design drove the choice of trial outcome, which determined the relative effect size, which together with natural history and waning/discontinuation assumptions drove the gained time in mild cognitive impairment, which drove quality-adjusted life-years gains and costs.
- For the evaluation of new AD treatments, we recommend (1) addressing trial outcome choice and treatment effectiveness assumptions in sensitivity analysis, (2) the use of a standardized reporting table for model predictions to support model cross-comparison, and (3) exploring the use of registries to measure long-term disease progression of future AD treatments to reduce the uncertainty of extrapolating short-term trial results by health-economic models.
- By identifying possible causes for variation between AD model predictions, our study supports AD model transparency and credibility and generates a set of recommendations for modelers and policy makers.

Introduction

Health-economic studies grounded in mathematical models play a pivotal role in supporting policy discussions concerning the accessibility, reimbursement, and pricing of interventions for Alzheimer's disease (AD). These model-based evaluations rely on results from clinical trials, necessitating the extrapolation of short-term clinical efficacy (eg, effects on biomarker targets and cognitive scales) to estimate the lifetime impact on patient-relevant outcomes (eg, function, care needs, and mortality). This in turn is often based on evidence from disease registries and cohort studies. Different health-economic models have shown to produce dissimilar results. Hence, transparency and credibility of these models are key to decision makers who rely on them[1] to facilitate timely access to new interventions for people living with AD. This article describes research aimed at better understanding why these differences among model predictions exist.

Comparative modeling has the potential to increase confidence in models if similar model predictions are observed.[2,3] Unexplained differences across AD model predictions could jeopardize

their credibility, making them unusable to support decision makers (eg, as reported in the field of abdominal aortic aneurysm screening[4,5]).

In 2020, the International Pharmaco-Economic Collaboration on Alzheimer's Disease (IPECAD) hosted a workshop to cross-compare AD models. Modeling groups were asked to implement a benchmark scenario consisting of a treatment with its effectiveness conceptually defined as a "30% reduction in disease progression."[6] Models were compared on the health-economic predictions they produced, which resulted in an increased understanding of model differences. However, because the benchmark treatment effect was defined on a conceptual level (ie, 30%), modelers did not have challenges on methodological (premodel analysis) choices on implementing real-world short-term clinical trial efficacy outcomes into their model, as well as extrapolation of effectiveness beyond trial follow-up evidence to long-term health-economic estimates. This challenge was discussed among modelers during the 2021 IPECAD follow-up workshop, highlighting the disconnect between efficacy estimates reported in AD clinical trials and effectiveness inputs required for their health-economic models. More specifically, challenges included (1) discrepancies in endpoints and scales in trials compared with models, (2) treatment effects reflecting single or multiple domains, (3) differences in statistical methods for analyzing trial efficacy data, and (4) differing analytical objectives for trials (identify a treatment effect in a clinical setting) compared with models (estimate the treatment benefit in a routine care setting) (see Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006 for details).[7]

Given our previous workshop recommendations for aligning models, the raised challenges around implementing a treatment effect, the recent Food and Drug Administration approvals of new AD treatments,[8,9] and the recently published significant results for a similar drug,[10] we believe there remains an urgent need to cross-compare AD models to increase our understanding of them and to support their credibility for evaluating new AD treatments.

This initiative aimed to cross-compare predictions from decision-analytic models for AD that start in mild cognitive impairment (MCI) and have implemented the same hypothetical trial efficacy estimates. The following research question was the starting point: What are the differences in key health-economic predictions across models that assess the cost-effectiveness of a hypothetical disease-modifying AD drug treatment, and what factors related to the use of clinical trial evidence explain those differences?

This article describes the methodology and results from the 2023 IPECAD workshop on the cross-comparison of AD models. The focus of this cross-comparison was on the challenges of implementing realistic short-term trial efficacy estimates in a decision model and extrapolating them to long-term health-economic estimates. Our goal was to describe treatment effect implementation methods and discuss how they translated into differences in health-economic model predictions, taking into account the specific model design, parameterization, and model assumptions.

Methods

The IPECAD steering group (ie, organizers) a priori defined a research question based on the previous workshop recommendations[6] (see introduction). Then, they designed a model cross-comparison study based on guidelines for multi-model comparisons[11] and foundations for model validation and comparison.[2] It consisted of the following 5 main steps.

First, the organizers identified and invited modeling groups (ie, participants).

Second, the organizers drafted a benchmark scenario, shared it with participating modeling groups, and provided them with the possibility to provide feedback to refine it.

Third, each modeling group implemented the benchmark scenario independently into their model as they deemed appropriate. Modeling groups were asked to maintain their model structures to avoid model convergence (ie, harmonization on [almost] all parameters omitting genuine discrepancies between models). Modeling groups were asked to report details on how they implemented the benchmark scenario into their model and to report their model predictions (ie, time in AD disease states, costs, and quality-adjusted life-years [QALYs]) to the organizers.

Fourth, organizers validated, cross-compared, and summarized all implementation details and model predictions and shared them with all modeling groups.

Fifth, we discussed cross-compared implementation details and model predictions during a workshop held on June 5 and 6, 2023, in Stockholm.

Model Identification, Selection, and Invitation (Step 1)

Models were identified through earlier participation in IPECAD workshops, systematic reviews, a PubMed search, an open call through ISPOR, snowballing, and ad hoc identification by the organizers. Duplicate models were omitted, leaving 46 identified models.

Modeling groups were invited (see Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006) if their models fulfilled the inclusion and exclusion criteria (reflected AD starting in MCI, AD disease progression, and health-economic outcomes in a US clinical setting) or could potentially fulfill the criteria after relatively small adjustments (n = 38). Among those invited, 12 did not respond, 2 considered their model outdated, 10 indicated insufficient resources to implement the benchmark scenario, and 4 did not have an operational model. This left 10 modeling groups for participation in the workshop (see Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006 for details).

Benchmark Scenario and Inputs (Step 2)

A hypothetical 18-month placebo-controlled trial for an AD treatment was developed in terms of eligibility criteria, baseline characteristics, and trial outcome estimates reflecting a population of persons with a diagnosis of MCI due to AD set in a US memory clinic setting. This was operationalized by creating summary tables and figures from synthetic individual patient-level data[12] on surrogate and secondary outcome estimates using observational data from the Alzheimer's Disease Neuroimaging Initiative[13] (see Appendix 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006).

The benchmark scenario to be implemented in the model was based on previous workshop recommendations[6] as follows. The target population was set as a standardized patient aged 70 years, male or female (modeled separately), with a diagnosis of amyloid-confirmed AD-type MCI on their first visit to a memory clinic in the United States. To focus on treatment implementation only, costs and effects related to diagnostics fell outside the scenario. The intervention strategy was a hypothetical treatment in addition to standard of care (ie, as reflected in the hypothetical 18-month trial results). The control strategy was standard of care. The treatment effect was considered disease modifying, although the implications of disease modification in a modeling context were not specified. Treatment discontinuation was set at 10% of patients

Table 1. Implementation of population and setting and natural history progression in MCI.

Group	Markov type					Microsimulation type			
Participant	**Biogen**	**CPEC**	**IPECAD**	**Bedrejo and colleagues**	**SveDem**	**BASQDEM**	**FEM**	**RTI-HS**	**MISCAN-Dementia**
Model information									
Abbreviation	(company name)	Care Policy Evaluation Centre	International Pharmaco-Economic Collaboration on Alzheimer's Disease	(author name)	Swedish Dementia Registry	Basque dementia model	Future Elderly Model*	(company name)	Microsimulation Screening Analysis (Dementia)
Developer(s)	P.P.-R.† M.U.† (Biogen), W.L.H.† (RTI-HS)	R.A.† (London School of Economics), R.H.† (Maastricht University)	R.H.† (Maastricht University), C.G.† (Biogen), A.G.† (Quantify Research)	C.B.S.† (University of Alberta), E.S. (University of Calgary)	A.W., R.H.† (Karolinska Institute)	J.M.† (Osakidetza Basque Health Service), M.S.-G.† (Mondragon Unibertsitatea)	B.T. (USC), Y.W.† (USC), J.H.† (USC, Masaryk University)	W.L.H.,† F.K.† (RTI Health Solutions)	C.B.,† I.d.K.† (Erasmus Medical Center)
Reference	Herring et al[14]	Anderson et al[15]	Green et al[16]		Wimo et al[17]	Mar et al[18]	Goldman et al[19]	Herring et al[20]; additional details in Handels et al[6,21]	Bruck et al[22]
Open source	No	No	Yes v1.1.2 www.ipecad.org	No	Yes v2.2.0 https://github.com/ronhandels/IPECAD	No	Yes: https://hrsdata.isr.umich.edu/data-products/future-elderly-model-fem-files‡	No	No
MCI natural history									
Data source(s)	Synthetic trial control arm	Vos et al[23]	Vos et al[23]	Potashman et al[40] [2021]	Synthetic trial control arm	Synthetic trial control arm; van Oudenhoven et al[44] [2019]	Cognitive status of US population in the Health and Retirement Study (HRS); Wei et al[46] (2022)	Synthetic trial control arm; Sapkota et al[42] [2021]	Synthetic trial control arm; Vermunt et al[45] [2019]
Source's estimate (and implementation)	Annualized CDR-global transition probability from CDR-global 0.5 (reflecting MCI with Aβ+) to mild AD with CDR-global 1; backward transitions (CDR-global 0.5 to CDR-global 0) to no AD assumed to remain in MCI	Annualized transition probability for IWG-2 prodromal AD criteria (reflecting MCI with Aβ+) to AD dementia	Annualized transition probability based on weighted average of NIA-AA high AD likelihood group and NIA-AA conflicting IAP group (reflecting MCI with Aβ+) to AD dementia	Annual CDR-SB transition probability from CDR-SB 0.5-4.0 (reflecting MCI with Aβ+) to mild AD with CDR-SB 4.5-9; backward transitions to asymptomatic AD assumed to remain in MCI	Annualized CDR-global transition probability from CDR-global 0.5 (reflecting MCI with Aβ+) to mild AD with CDR-global 1; backward transitions (CDR-global 0.5 to CDR-global 0) to no AD assumed to remain in MCI	Mixed model (originally developed based on van Oudenhoven et al[44] [2019] CDR-SB data) fit to synthetic trial CDR-SB data; differentiation between moderate and severe added to model. Model simulates times until mild, moderate, and severe using cut points CDR-SB 4.5, 9.5, and 16, respectively.	ADNI: linear regression model for CDR-SB using splines of lagged CDR-SB, age, gender, and race as covariates. Synthetic trial: baseline CDR-SB and gender (separate) used as input for linear model and cutoff applied to classify as mild, moderate, and severe dementia	Sapkota: annual decline in MMSE and DAD (reflecting MCI with Aβ+) using MMSE and FAQ changes from baseline from synthetic trial (FAQ mapped to DAD) Synthetic trial: baseline MMSE and DAD (mapped from FAQ)	Vermunt: duration from MCI with Aβ+ to AD dementia, calibrated to proportion transitions to CDR-global 1 (from synthetic trial) by restricting fast progressing individuals; backward transitions (CDR 0.5 to CDR 0) to no AD not considered
Source's age and sex (and adjustments)	Age: 73 years, gender: 42% female	Age: 71.4 years, gender: 48% female)	Age: 71.4 and 66.1 years, gender: 47% and 43% female, respectively	Age and gender for MCI patients not reported separately	Age: 73 years, gender: 42% female	Age: 73 years, gender: 42% female	Age: 70 years (starting age in model) Gender: gender-specific progression	Age: 73 years, gender: 42% female	Age: 70 years, gender: 42% (reflected as observed in data source), with gender-specific progression in the model calibrated to gender-average rate observed from trial

continued on next page

Table 1. Continued

Group	Markov type					Microsimulation type			
Participant	**Biogen**	**CPEC**	**IPECAD**	**Bedrejo and colleagues**	**SveDem**	**BASQDEM**	**FEM**	**RTI-HS**	**MISCAN-Dementia**
Source's syndrome(s) (and adjustments)	Prevalent MCI defined by episodic memory impairment and CDR-global (0.5) and MMSE (≥24) at baseline; conversion to dementia defined by CDR-global score	Prevalent MCI based on Petersen et al[49] criteria; conversion to AD-type dementia defined by clinical criteria	Prevalent MCI based on Petersen et al (2004) criteria; conversion to AD-type dementia defined by clinical criteria	Incident MCI defined by CDR-SB (0.5-4.0) at baseline; conversion to dementia defined by CDR-SOB	Prevalent MCI defined by episodic memory impairment and CDR-global (0.5) and MMSE (≥24) at baseline; conversion to dementia defined by CDR-global score	Prevalent MCI defined by episodic memory impairment and CDR-global (0.5) and MMSE (≥24) at baseline; conversion to dementia defined by CDR-global score	Prevalent MCI defined by CDR-global (0.5) at baseline; conversion to dementia defined by CDR-global (≥1)	Prevalent MCI defined by CDR-global (0.5) and MMSE (≥24) at baseline; AD dementia diagnosis at MMSE ≥1.5 SDs below normal and DAD ≤93 applied to simulated data (informed by McKhann et al[37] [2011])	Prevalent MCI defined by episodic memory impairment and CDR-global (0.5) and MMSE (≥24) at baseline; conversion to dementia defined by CDR-global score
Source's pathology (and adjustments)	Amyloid positive via abnormal CSF amyloid beta or amyloid PET; no criteria on tau	IWG-2 criteria: "Abnormal CSF amyloid-beta1-42 and CSF tau" or "abnormal amyloid PET"	NIA-AA high criteria: Abnormal amyloid (CSF or PET) and abnormal neuronal injury markers (CSF, PET, or MRI); NIA-AA conflicting IAP criteria: abnormal amyloid (CSF or PET) and normal neuronal injury markers (CSF, PET, or MRI)	Amyloid positive via abnormal CSF amyloid beta, amyloid PET, or (retrospectively) autopsy result; no criteria on tau or neuronal injury	Amyloid positive via abnormal CSF amyloid beta or amyloid PET; no criteria on tau	Amyloid positive via abnormal CSF amyloid beta or amyloid PET; no criteria on tau	No amyloid parameter or assumption has been made	Synthetic trial: Amyloid positive via abnormal CSF amyloid beta or amyloid PET; no criteria on tau	Synthetic trial: Amyloid positive via abnormal CSF amyloid beta or amyloid PET; no criteria on tau
Source's clinical setting (and adjustments)	Clinical trial	Combination of local, academic, and tertiary memory clinic, and research care settings	Combination of local, academic, and tertiary memory clinic, and research care settings	Research care settings	Clinical trial	Clinical trial	ADNI: combination of local, academic, and tertiary memory clinic and research care settings	Clinical trial	Clinical trial
Model severity level cutoffs	CDR-SB, MCI: 0.5-4, mild: 4.5-9, moderate: 9.5-15.5, severe: 16-18 [O'Bryant et al,[39] 2010]	MMSE, mild: 30-20, moderate: 19-10, severe: 9-0	MMSE, mild: 30-20, moderate: 19-10, severe: 9-0 FAQ, mild: 0-8, moderate: 9-23, severe: 24-30) NPI-Q, mild: each item ≤1, moderate: each item ≤2 with at least one 2, severe: at least one item with 3	CDR-SB, MCI: 0.5-4, mild: 4.5-9, moderate: 9.5-15.5, severe: 16-18 [O'Bryant et al,[39] 2010]	MMSE, mild: 30-20, moderate: 19-10, severe: 9-0	CDR-SB, MCI: ≤4, mild: 4.5-9, moderate: 9.5-15.5, severe: 16-18 [O'Bryant et al,[39] 2010]	Crosswalk [O'Bryant et al,[39] 2008: table 2] to generate global CDR from simulated CDR-SB	MMSE, mild: >20, moderate: 10-20, severe: <10	CDR-global, MCI: 0.5, mild: 1, moderate: 2, severe: 3

continued on next page

Table 1. Continued

Group	Markov type					Microsimulation type			
Participant	**Biogen**	**CPEC**	**IPECAD**	**Bedrejo and colleagues**	**SveDem**	**BASQDEM**	**FEM**	**RTI-HS**	**MISCAN-Dementia**
Model settings for the United States									
Costs	Direct medical costs: Leibson et al[35] [2015] Direct nonmedical costs: Robinson et al[41] [2020]; Gustavsson et al[30] [2011]	Gustavsson et al[30] [2011] and assumed to be over and above the age-specific nondementia cost of health services. MCI was assumed to not impose excess cost.	Dementia: Gustavsson et al[30] [2011]; MCI: assumed 50% of mild dementia [Darba et al,[27] 2015; Jonsson et al,[31] 2006]	Lin et al[36] [2021] ICER report (derived from Leibson et al[35] [2015])	Tahami et al[43] [2022] (derived from Robinson et al[41] [2020] and Genworth et al[29] [2022]) weighted for institutionalization proportions (derived from Tahami Monfared et al[24] referring to Neumann et al[38] [1999]).	Robinson et al[41] [2020]; Gustavsson et al[30] [2011]	Medicare Current Beneficiary Survey (2007-2010)	Robinson et al[41] [2020]; Gustavsson et al[30] [2011]	Robinson et al[41] [2020]; Gustavsson et al[30] [2011]
Utilities	Landeiro et al[33] [2020]	Neumann et al[38] [1999]; Karlawish et al[32] [2008]	Neumann et al[38] [1999]	Neumann et al[38] [1999]	Neumann et al[38] [1999]	Neumann et al[38] [1999]	Leaf[34] [2015]	Landeiro et al[33] [2020]	Neumann et al[38] [1999]
Mortality	RRs for MCI and dementia [Wilson et al,[47] 2009] applied to US general population mortality 2019	Excess mortality in moderate and severe dementia Brookmeyer et al[26] [2007] applied to US general population mortality 2019	RRs for dementia only [Andersen et al,[25] 2010] applied to US general population mortality 2019	RRs for MCI and dementia [Andersen et al,[25] 2010] applied to US general population mortality 2020	RRs for dementia only[17] relative to very mild dementia and Andersen et al[25] [2010] very mild dementia to normal) applied to US general population mortality 2019	Time to death from mild-moderate-severe dementia [Dodge et al,[28] 2003]; time to death from other causes adjusted to US setting from US life table	Mortality risk matches those observed in HRS adjusted for risk factors. Data from the 2018 wave of HRS are used	RRs for MCI and dementia[17] (see also Wilson et al [47] 2009) applied to US general population mortality 2019	Time to death from dementia from Rotterdam study [Wolters et al,[48] 2019]; time to death from other causes from US general population life tables 2019

Aβ+ indicates amyloid beta positive; AD, Alzheimer's Disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CDR, clinical dementia rating; CDR-SB, clinical dementia rating – sum of boxes; CSF, cerebrospinal fluid; DAD, disability assessment for dementia; FAQ; functional activities questionnaire; FEM, future elderly model; HRS, Health and Retirement Study; IAP, isolated amyloid pathology; MCI, mild cognitive impairment; MMSE, mini-mental state examination; MRI, magnetic resonance imaging; NIA-AA, National Institute on Aging and the Alzheimer's Association; PET, positron emission tomography; RR, relative risk; SD, standard deviation.
*FEM draws on a cognitive and functional impairment transition model that has been validated for a 5-year period. Long-term estimates introduce considerable uncertainty and may underestimate or overestimate the value of treatment in the scenario studied.
†Indicates present during the workshop.
‡Details can be found here: https://www.rand.org/content/dam/rand/pubs/rgs_dissertations/RGSDA1400/RGSDA1439-1/RAND_RGSDA1439-1.pdf; please contact authors for details on methods when replicating the results of this study.

over the 18-month trial follow-up period, which included discontinuation due to any adverse events (including amyloid-related imaging abnormalities). Treatment was to be applied up to and including mild dementia. Treatment waning was set at 5% per year. US-specific background mortality and consumer price index data were specified for the year 2019 (reflecting pre-COVID-19 levels). The time horizon was set at 25 years and the discount rate at 3.5% per year for both costs and health outcomes. Suggestions to reflect a US setting in terms of mortality, utility, and cost estimates were also provided (see Appendix 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006).

Implement Benchmark Scenario and Model Predictions (Step 3)

Both the synthetic individual-patient level data set and summary tables and figures were provided to all modeling groups. If applicable, modeling groups were expected to apply mapping to translate the benchmark scenario outcomes to the outcomes used in their model.

Each modeling group was asked to implement the benchmark scenario. Model predictions consisted of proportions in AD disease states over time, cumulative time in AD disease states, and health-economic outcomes (QALYs and costs) by control/intervention strategy and sex. These model predictions were requested to be shared in a predefined format (see Tables 1[14–48], 2, and 3).

Validation and Analysis (Step 4)

The participating modeling groups submitted their benchmark scenario implementation details and model predictions, and the organizers synthesized them into summary tables and graphs. An ad hoc quality check was performed to compare the person-years alive with the person-years in disease states. A qualitative review of the implementation description was also performed. If issues were identified, either from these consistency checks or during the workshop discussion, modeling groups had the opportunity to revisit their approach and submit updated results. This resulted in the following changes after the workshop: (1) adjusting MCI progression to better reflect the changes from baseline over time observed in the synthetic trial data (RTI-HS model), (2) use of an updated open-source version for reasons unrelated to the results (SveDem model), (3) redesign of the benchmark implementation due to incorrect interpretation (Bedrejo model), and (4) omitting the results due to lack of face validity (ADDITION model[50]). A member of the future elderly model (FEM) group participated in the workshop discussions without having submitted their model predictions due to time limitations. We relied on any validation and verification of the original models, and we similarly relied on

the individual modeling groups to validate their implementations of the benchmark scenario. Modeling groups did not have access to one another's model predictions before submitting their results, except for the FEM, although the use of other model predictions for preparing their own model predictions was viewed as unlikely.

Workshop and Results Synthesis (Step 5)

All results (implementation details, model predictions, and summarized cross-comparison results) were shared with all modeling groups before the workshop. Next, participants presented and discussed these results (except for FEM results) during the workshop in a hybrid setting, with online participants' discussion occurring separately among themselves and summarized to the face-to-face participants. Given that there was no established methodology for evaluating concordance in health-economic simulation modeling, between-model differences were interpreted qualitatively.

Results

In this section, we summarize the treatment implementation methods and the variability between model predictions. In addition, we describe possible causes for variation based on discussions during the workshop and on supporting post hoc correlation analyses provided in Appendix 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006.

Treatment Implementation Methods

The following 9 participating models (5 Markov and 4 microsimulation type) were part of this cross-comparison: BASQDEM,[18] Biogen,[14] CPEC,[15] FEM,[19] RTI-HS,[20] IPECAD open-source model,[16] MISCAN,[22] Bedrejo and colleagues, and SveDem open-source model[17] (Table 1).[14-48]

The sources used to represent the target population's natural disease progression varied in setting (clinical vs general population) and biomarker status (abnormal amyloid only, both abnormal amyloid and neuronal injury, implied amyloid abnormality). Some models modified their MCI natural history to fully (Biogen, SveDem) or in part (BASQDEM, RTI-HS, MISCAN) reflect the synthetic trial placebo arm. The specific starting age of 70 years and separately modeled sex were mainly applied to mortality whereas the average age and sex of the source for natural history were assumed representative for the modeled population (except for MISCAN, BASQDEM, FEM). See Table 1[14-48] for details.

The implementation of the treatment effect varied in choice of outcome domain (cognition, function, and/or composite), choice of scale (clinical dementia rating – sum of boxes [CDR-SB], CDR-global, mini-mental state examination [MMSE], and/or functional activities questionnaire [FAQ]), and choice of unit (health state transition or change from baseline). These choices were typically made to match the domain and scale used in the model and the model type (Markov or microsimulation). In addition, implementations differed in the level of premodel analysis conducted using the hypothetical trial data, such as estimating the ratio in mean change from baseline between the intervention and control arms, fitting a survival model to the individual patient-level data to obtain a hazard ratio, applying the relative effect on a continuous outcome as a relative risk (RR) reduction on a transition probability, or calibration. None of the included models used the hypothetical trial biomarker outcomes to implement a treatment effect. Discontinuation and waning varied in terms of being implemented at all (6 of 9 models) and in their timing and duration. See Table 2 for details.

Between-Model Variability

Disease progression in the control strategy and the difference with the intervention strategy varied across models. Mean time spent in MCI over a 10-year time horizon in the control strategy ranged from 2.6 to 5.2 years (see Fig. 1). The difference in MCI between intervention and control strategy ranged approximately 0.1 to 1.0 years (−0.8 to 0.3 for mild, −0.5 to 0.0 for moderate, −0.6 to 0.0 for severe; Fig. 2). The proportion in MCI over time in the control strategy varied between models, eg, range 0.16 to 0.60 at year 3 and 0.04 to 0.42 at year 5 (see Appendix Fig. 6.1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006). Two models (BASQDEM and RTI-HS) showed a relatively slower progression than other models in year 1 and faster progression for year 2 onward.

QALYs and costs in the control strategy ranged from 4.2 to 6.3 and from $68 558 to $451 421, respectively. The difference in QALYs and costs (excluding treatment costs) between intervention and control ranged from 0 to 0.6 and −$66 897 to $11 896, respectively (see Fig. 3). One model (RTI-HS) showed a relatively large cost saving and only one model (BASQDEM) showed small increased costs.

Sex differences (women minus men) in incremental 10-year time in MCI, cumulative QALYs, and cumulative costs ranged from 0.01 to 0.07, −0.01 to 0.05, and −$13 147 to $349 respectively, with the largest differences by 2 models (RTI-HS and BASQDEM).

Possible Causes for Variation

We performed additional post hoc analyses (see Appendix 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006) to assess the correlation among time gained in disease state (Fig. 2), QALY gains and differences in costs (Fig. 3), and implementation details (Table 2).

The variation in predicted gained time in MCI seemed relatively strongly related to the treatment effect estimate (see Appendix Fig. 6.2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006). The trial outcomes chosen by the modeling groups from the benchmark hypothetical trial varied meaningfully. This resulted in the use of different treatment effect estimates (ie, the relative effect of the hypothetical benchmark intervention expressed as the ratio in average change from baseline between control and intervention arm, being CDR-SB = 0.70, CDR-global = 0.93, MMSE = 0.65, and FAQ = 0.73, and expressed as the hazard ratio from trial individual patient-level data Cox survival analysis being CDR-SB = 0.80). Nevertheless, 5 models that selected the same treatment effect outcome (CDR-SB) deviated relatively widely in their gained time in MCI (FEM = 0.3 years, IPECAD = 0.4, Biogen = 0.6, Bedrejo = 0.6, and BASQDEM = 0.9). Among other factors described below, the difference between 2 of these models (IPECAD and Biogen) could be explained by one of them (Biogen) implementing the trial ratio in change from baseline of 0.70 directly as an RR for transitioning to dementia, whereas the other (IPECAD) fitted a survival model that produced a hazard ratio of 0.80 implemented as an RR for transitioning to dementia. The difference between 2 other models (IPECAD and Bedrejo), which both applied the same RR of 0.80, could be explained by one of them (Bedrejo) applying a relatively small treatment effect waning. The similarity between 2 other models (Biogen and Bedrejo) could be explained by different assumptions canceling one another out: one of them (Biogen) applied a higher treatment effect (RR 0.70) but with waning whereas the other (Bedrejo) applied a lower treatment effect (RR 0.80) but with limited waning. The difference between these models and

Table 2. Implementation of treatment effect.

Group	Markov type					Microsimulation type			
Participant	**Biogen**	**CPEC**	**IPECAD**	**Bedrejo and colleagues**	**SveDem**	**BASQDEM**	**FEM**	**RTI-HS**	**MISCAN-Dementia**
Implementation									
Choice of outcome(s)	CDR-SB, change from baseline	MMSE, change from baseline	CDR-SB, time to dementia	CDR-SB, time to dementia	CDR-global, change from baseline	CDR-SB, change from baseline	CDR-SB, change from baseline	MMSE and FAQ, changes from baseline	CDR-global, change from baseline
Rationale	Health states in the model defined using CDR-SB ranges	Health states in the model defined using MMSE ranges	Measure captures multiple domains and is sensitive in early AD.	Health states in the model defined using CDR-SB ranges	Transition to dementia is clinically relevant; aligns with use of synthetic control arm for MCI natural history	Time to dementia in the model relies on equation for CDR-SB progression over time.	Cognitive states in the model are defined using CDR-SB, which also contributes to the staging of dementia.	Aligned with the cognitive (MMSE) and functional (DAD) domains in the model (with mapping from FAQ to DAD)	Appropriate for the model's time to dementia approach; aligns with use of synthetic control arm for MCI natural history
Estimate(s) used/ metric used	Relative difference in CDR-SB change from baseline at 18 months	Relative difference in MMSE change from baseline at 18 months	HR for time to CDR-SB ≥4.5 events	[Same as IPECAD]	Relative difference in proportions transitioning to CDR-global 1 at 18 months	CDR-SB change from baseline at each time point in both arms	Relative difference in CDR-SB change from baseline at 18 months	Relative reduction in MMSE and FAQ change from baseline at 18 months	CDR-global change from baseline at 18 months
Relative effect observed in trial of the chosen outcome	HR = 0.70	RR = 0.65	HR = 0.80	HR = 0.80	RR = 0.93	The regression mixed model contains a beta to include the trial arm	HR = 0.70	Relative decline = 0.65 (MMSE) and 0.73 (FAQ)	RR = 0.93
Synthetic trial individual-level data analysis, if applicable	None	None	Cox survival model using time to CDR-SB ≥4.5 events	[Same as IPECAD]	None*	Mixed regression model fitted to CDR-SB individual-level synthetic trial data	None	None	None
Implementation in model	Applied as a HR to transition probabilities from MCI to mild dementia and from mild dementia to moderate dementia	Applied as an RR to transition probabilities from MCI to mild dementia and from mild dementia to moderate dementia	HR applied directly to transition probability from MCI to mild dementia; calibration required to model treatment effect in mild dementia	Applied as a HR to transition probabilities from MCI to mild dementia and from mild dementia to moderate dementia	Applied as an RR to transition probabilities from MCI to mild dementia and from mild dementia to moderate dementia	Parameter in mixed regression model to reflect difference between control and intervention arm CDR-SB	Applied as an HR to transition probabilities from MCI to mild dementia and from mild dementia to moderate dementia	Relative reduction in MMSE and FAQ change from baseline applied to MMSE and DAD annual rates of decline, assuming a linear mapping between FAQ and DAD	Multiplicative factor for time in MCI calibrated to the CDR-global transitions at 18 months for intervention arm; calibrated factor also applied to time in mild dementia
Extrapolation beyond 18-month synthetic trial data									
Extrapolation under treatment†	Effectiveness maintained in MCI and mild dementia	Effectiveness maintained in MCI and mild dementia	Effectiveness maintained in MCI and mild dementia	Effectiveness maintained in MCI and mild dementia	Effectiveness maintained in MCI and mild dementia	Effectiveness maintained in MCI	Effectiveness maintained in MCI and mild dementia	Effectiveness maintained in MCI and mild dementia	Effectiveness maintained in MCI and mild dementia
Type of extrapolation after treatment discontinuation†	Disease-modifying: revert to natural history progression (gains maintained)	Disease-modifying: revert to natural history progression (gains maintained)	Disease-modifying: revert to natural history progression (gains maintained)	Disease-modifying: revert to natural history progression (gains maintained)	Disease-modifying: revert to natural history progression (gains maintained)	Partly disease-modifying: revert to natural history progression (gains maintained); partly symptomatic: revert to natural history state (gains lost)	Disease-modifying: revert to natural history progression (gains maintained)	Disease-modifying: revert to natural history progression (gains maintained)	Disease-modifying: revert to natural history progression; gains maintained

continued on next page

Table 2. Continued

Group	Markov type					Microsimulation type			
Participant	**Biogen**	**CPEC**	**IPECAD**	**Bedrejo and colleagues**	**SveDem**	**BASQDEM**	**FEM**	**RTI-HS**	**MISCAN-Dementia**
Waning description	Waning of 5% of effect lost per year starting in year 2	No waning implemented	Waning of 5% of effect lost per year in MCI starting in year 2; time-dependent waning not applied to calibrated multi-domain effect in mild dementia, opting instead for a constant waning factor set to the 10-year average of the MCI waning	Waning was assumed to add up to 15% in the transition from mild to moderate, implemented as 0.80-(1-0.15)	Waning of 5% of effect lost per year starting in year 2	No waning implemented	No waning implemented	Waning of 5% of effect lost per year starting in year 2	5% of patients lose effect per year; implemented by proportionally scaling the incremental time in MCI and mild dementia dependent on the timing of waning
Discontinuation									
Description	10% over 18 months from synthetic trial intervention group; converted to 6.8% per year in all years	Not implemented	10% in year 1; further discontinuation at 4% per year in all subsequent years	10% in year 1 only	10% in year 1 only	Not implemented	Not implemented	10% over 18 months from synthetic trial intervention group; converted to 6.8% per year in all years	10% over 18 months from synthetic trial intervention group; no discontinuation assumed after 18 months
Stopping rule	At moderate dementia	At moderate dementia	At moderate dementia	At moderate dementia	At moderate dementia	After a maximum of 10 years on treatment	At moderate dementia	At moderate dementia	At moderate dementia
Adverse events									
Were adverse events implemented and, if so, how?	Implemented ARIA events in year 1 only and fall and headache event rates per year from synthetic trial intervention and control data; costs and disutilities assigned per event	AEs and ARIA not explicitly considered	Accounted for ARIA events in year 1 only from incremental synthetic trial event data; implemented via a single disutility and cost	Disutility of 0.14 in year 1 was applied for ARIA-E to the proportion of the baseline population, estimate was obtained from the symptomatic ARIA-E difference between synthetic data intervention and control; assumed 3 MRI scans for each individual in year 1	AEs accounted for via a cost of $1000 in year 1 only	AEs and ARIA not considered	AEs and ARIA not considered	AEs and ARIA not explicitly considered; assumed to be reflected in treatment discontinuation rate	AEs and ARIA not considered
Model outcomes									
Proportion in dementia year 1 (conditional on survival) in control strategy	15%	27%	22%	23%	14%	3%	14%	10%	14%
Ratio of proportion in dementia between control and intervention	0.72	0.65	0.83	0.80	0.94	0.43	0.71	0.18	0.89

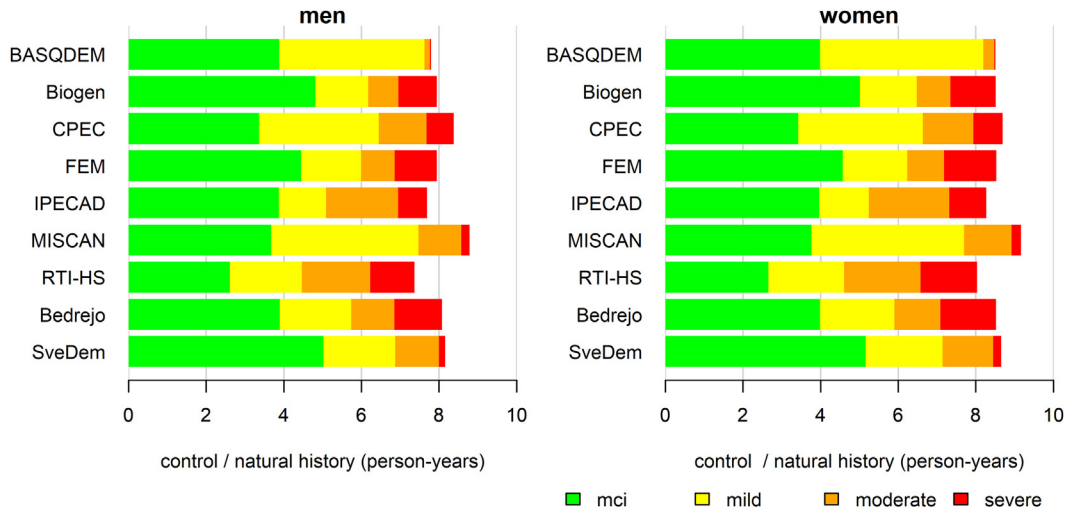
AD indicates Alzheimer's disease; AE, adverse event; ARIA, amyloid-related imaging abnormality; CDR-SB, clinical dementia rating – sum of boxes; DAD, disability assessment for dementia; FAQ, functional activities questionnaire; HR, hazard ratio; MCI, mild cognitive impairment; MMSE, mini-mental state examination; MRI, magnetic resonance imaging; RR, relative risk.

*An attempt was made, but the result was considered not valid and therefore not used.

†Extrapolation categories:

- Absolute control counterfactual / gains lost / symptomatic: revert to control strategy counterfactual (ie, natural progression) in terms of absolute symptoms at same time.
- Relative counterfactual / gains maintained / disease-modifying: revert to control strategy counterfactual (ie, natural progression) in terms of rate of decline at same state.
- Effectiveness maintained / cure: effect of the treatment remains applicable.

Figure 1. Model prediction of time spent in states over a 10-year time horizon for the control strategy, for men (left) and women (right).



another model (BASQDEM) is possibly due to faster natural progression creating a larger room for improvement as described below (and vice versa for FEM due to slower progression creating a smaller room for improvement).

Predicted gained time in MCI seemed related to natural disease history (ie, time spent in MCI in the control strategy) (see Appendix Fig. 6.3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006), although somewhat less than the relation with "choice of outcome(s)" as described earlier. Spending less time in MCI implies a higher risk of dementia onset, which generates a greater room for improvement (ie, a larger proportion to be prevented). Nevertheless, 4 models (BASQDEM, IPECAD, Bedrejo, and MISCAN) showed a relatively large variation, which could be explained by the treatment effect (increasing respectively with their deviation); the same accounts for the difference between 2 models (SveDem and Biogen). One model (BASQDEM) showed almost no time in moderate or severe states leaving less room to reduce time spent in these states, which are associated with low quality of life and high cost.

QALY gains could be explained by the gained time in MCI given that they seemed relatively strongly related (see Appendix Fig. 6.4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006). Nevertheless, 4 models (BASQDEM, CPEC, Biogen, and Bedrejo) showed a relatively large range in QALY gain at similar gained time in MCI. The difference between 2 models (Biogen and Bedrejo) could be explained by one of them (Bedrejo) applying lower waning in mild dementia and therefore gaining more QALYs from the treatment in that disease stage. The differences between the 2 other models (BASQDEM and CPEC) were difficult to explain and possibly related to differences in dementia natural history between them.

Differences in costs (excluding treatment costs) could partly be explained by the gained time in MCI given that they seemed somewhat related (see Appendix Fig. 6.5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.09.006).

Figure 2. Model prediction of difference (in terms of intervention minus control strategy) in time spent in states over a 10-year time horizon for men (left) and women (right).

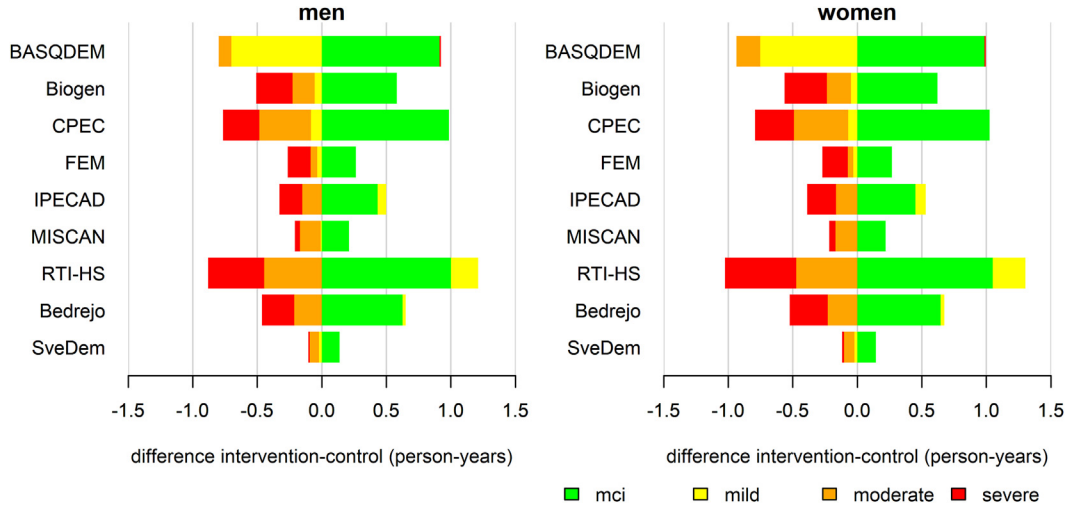
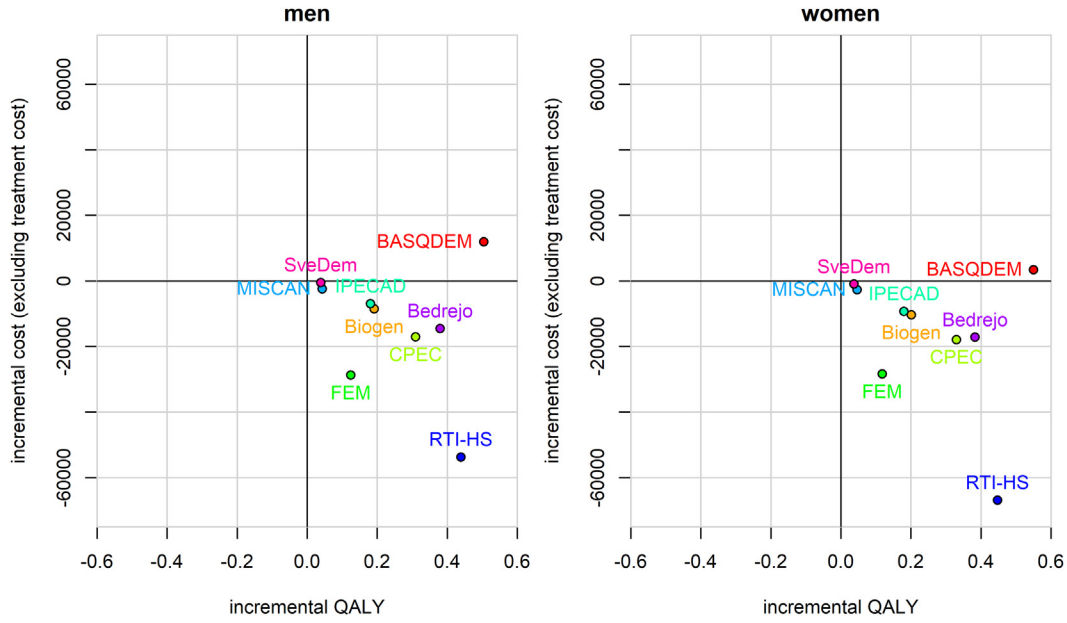


Figure 3. Model prediction of incremental QALYs and incremental costs (without treatment costs) for men (left) and woman (right) over a restricted 10-year time horizon.



QALY indicates quality-adjusted life-year.

Nevertheless, 3 models with similar gained time in MCI (BASQDEM, CPEC, RTI-HS) showed a relatively large variation in costs. This could be explained by one of them (BASQDEM) "trading" time in mild dementia for time in MCI and another (RTI-HS) "trading" time in moderate and severe for time in MCI and mild. With higher costs in more severe stages, the benefits of "trading" those stages are higher. In one model (BASQDEM), the costs related to gained life-years spent in MCI were likely not offset by savings related to prevention of spending life-years in mild and moderate dementia. This balanced effect in costs does not apply to QALYs where all additional life-years result in QALY increase rather than additional care costs in life-years gained. Other causes could relate to cost input estimates used or impact on mortality making it difficult to pinpoint the precise cause for variation in costs.

We explored various other factors to explain model differences. Microsimulation-type versus Markov-type differences were not uniformly related to gained time in MCI (see Fig. 2) and varied relatively largely in QALY gains and costs (excluding treatment costs) (see Fig. 3). The differences seem to be better explained by the other factors described earlier. Sex differences showed slightly higher benefits in women in gained time in MCI (see Fig. 2) and health-economic predictions (see Fig. 3), most likely related to a higher life expectancy providing a larger room for treatment benefit.

None of the models used the benchmark biomarker results, none selected the composite outcome (AD composite score), none implemented a direct effect on mortality (ie, unrelated to disease state), and all assumed the treatment to be (partly) of disease-modifying nature in terms of reverting to natural history progression rate with treatment gains maintained once treatment is ceased. These factors remain undetermined as possible factors that could cause variation across model predictions.

Table 3. Suggested standard template to report predictions of disease progression from future AD simulation models: proportion of persons in each state over time in the control and intervention strategy over a 25-year time horizon separately for men and women; undiscounted and no half-cycle correction (ie, proportion at end of year; eg, at year 2, the proportion at day 365 × 2 is to be provided).

Year	MCI	Mild dementia	Moderate dementia	Severe dementia	Death
Control					
0	1	0	0	0	0
1					
2					
3					
…					
25					
Intervention					
0	1	0	0	0	0
1					
2					
3					
…					
25					

AD indicates Alzheimer's disease; MCI, mild cognitive impairment.

Discussion

This study cross-compared the predictions from 9 AD decision-analytic models that started in MCI and independently implemented efficacy estimates from the same benchmark hypothetical trial for an AD treatment. Variation was found in model

Table 4. Suggested standard template to report health-economic predictions from future AD simulation models: cumulative outcomes over 10-year time horizon separately for men and women (mean per person).

Outcome	Control	Intervention
Time (years) on intervention (undiscounted)		
Time in full-time/institutionalized care (undiscounted)		
Direct costs medical (discounted and undiscounted)		
Direct costs nonmedical (discounted and undiscounted)		
Patient QALY (discounted and undiscounted)		
Informal carer QALY (discounted and undiscounted)		

AD indicates Alzheimer's disease; QALY, quality-adjusted life-year.

predictions of natural disease history, time gained in MCI, QALYs, and costs.

Our results suggest the following cascade of factors that explained differences in model predictions: model design drove the choice of the outcome selected from trial evidence (eg, CDR-SB, MMSE), which was associated with the relative effect size from the trial evidence (7%-35%), which together with natural history (faster progression creating a larger room for improvement) and assumptions on waning and discontinuation (eg, effectiveness maintained under and after treatment, reversion to natural history progression after discontinuation) drove the gained time in less severe disease states (particularly MCI), which drove the QALY gain (by both more time in less severe states and increased life expectancy) and costs (although less straightforward as cost-savings due to more time in less severe states are offset by additional costs in life-years gained). Time spent on treatment and model type (Markov or microsimulation) did not seem to explain model differences. We note treatment costs were not included, although these are expected to be influenced by time on treatment.

Although models varied in their choice of treatment effect outcome, the modeling groups' rationales for linking their selected outcomes with the approaches used to model natural disease progression suggested face validity. However, we think applying the ratio between mean change from baseline on continuous outcomes such as CDR-SB and MMSE as an RR of conversion from MCI to dementia or from mild to moderate dementia (as applied by Biogen and CPEC models in this analysis) is uncertain due to limited empirical evidence. However, this choice might have been driven by limited resources not allowing exploration of alternative options that are likely more time consuming such as analyzing the synthetic individual-level trial data or addressing this choice in sensitivity analysis.

The observed variation in natural history in the current IPECAD workshop cross-comparison study showed similarities to the variation in the previous IPECAD workshop cross-comparison in 2020.[6] The previous workshop specified the benchmark as a 30% reduction in progression from MCI to dementia rather than a set of plausible trial outcomes with different relative effects in our current study. The range of time (years) spent in MCI (natural history) in our current study (2.4 to 5.2) was lower and wider than the range reported by the previous cross-comparison (3.4 to 5.6). We think this is due to an emphasis on the more specific amyloid-confirmed MCI starting population, which reflects a higher risk of dementia onset (ie, shorter time in MCI). The difference in time spent in MCI between intervention and control in our current study (0.1-1.0) was lower than the previous cross-comparison (approximately 0.3-1.7 with 1.1 being the second highest). We consider this is due to the relative treatment effect ranging from 7% to 35% compared with the uniform application of a 30% reduction in the previous cross-comparison.

A previous study focusing on structural uncertainty and cross-validation of decision models in AD[51] applied a hypothetical 30% reduction in disease progression from MCI to dementia using 3 different model types (1 Markov and 2 microsimulation) in combination with 4 data sources for natural progression, holding other inputs and assumptions constant. Incremental QALYs ranged from approximately 0.35 to 0.50 but when applying a 15% treatment effect instead of 30% it ranged from 0.15 to 0.25. We judge that these results compare favorably with the range in our current cross-comparison (0.05 to 0.55), anticipating this corresponds to a 7% to 35% treatment effect.

Recommendations for Reimbursement Agencies and Modelers

Based on our findings, we have 5 key recommendations for researchers and experts. First, for the evaluation of treatments for AD, we recommend that modelers address and discuss choices and assumptions in sensitivity analyses, especially those listed in Tables 1[14-48] and 2 along with other assumptions highlighted by earlier cross-comparison and reviews, such as mortality[6] and alternative data sources.[52,53] We recommend that reimbursement agencies request the results of such analyses or test them themselves, for example, using one of the available AD open-source models.[16,54,55] We note that our study did not produce the currently lacking empirical evidence on what choices and assumptions are correct, but merely the uncertainty around their choice.

Second, we recommend for modelers to report, and for reimbursement agencies to request, a standardized table reporting the proportion in different states at each year (timepoint), as designed in Tables 3 and 4. This adds to generic[56] and AD-specific[57] reporting recommendations on (dis)aggregated health-economic outcomes. For this table, we acknowledge the possible requirement of categorization and/or mapping[58] into disease stages being a potential source for variation in itself, but we consider it the optimal balance between the level of detail and outcomes most AD models can provide.

Third, we recommend that reimbursement agencies support the generation of long-term real-world evidence registries (incorporating both clinical and health-economic outcomes). Such evidence can be used for extrapolating treatment effects beyond short-term trial follow-up[59,60] to reduce uncertainty related to assumptions on extrapolation in health-economic models.

Fourth, we invite modelers to submit their results to our IPECAD registry for open continuous cross-comparison (https://osf.io/jv85a/) in the format of Tables 3 and 4 as well as earlier Tables 1[14-48] and 2 to further support the transparency and credibility of AD models. We recommend this both for our benchmark scenario and for recent AD drug trials such as aducanumab,[8] lecanemab,[9] and donanemab[10] to support ongoing model cross-validation in AD.

Fifth, because gained time in MCI was an important driver of health-economic model predictions, we recommend modelers to detail the MCI stage of a model in at least as much detail as the dementia stage, noting that earlier reviews recommended reflecting dementia as continuous disease progression on multiple domains.[61,62] This could be done, for example, by explicitly modeling or using subpopulation-specific input estimates for cognitive status (eg, early/late or [non]amnestic cognitive impairment) and biomarker status ([ab]normal amyloid, [ab]normal neuronal injury), as provided by Vos et al.[23] Because "trading" time between less and more severe stages (as described earlier) seemed to drive health-economic model predictions, it remains important to reflect dementia stages in some detail.

Strengths/Limitations

Of the modeling groups invited to participate, 12 groups did not participate due to nonresponse and 10 groups could not participate due to insufficient resources, which limits the generalizability of our results. However, among the 12 nonresponders, most models were relatively old or reflected categorized disease progression states, a well-represented model type in our cross-comparison. Nevertheless, of the 8 models with limited resources to participate, one explicitly modeled biomarkers[24] and one was used for the evaluation of aducanumab and lecanemab,[63] which would likely have enriched the data of our cross-comparison.

The analyses described here omit treatment costs and hence do not inform cost-effectiveness estimates for these therapies. In addition, our benchmark scenario did not cover any potential diagnostic procedures, limiting generalizability to the assessment of real-world pharmacological interventions. We note the method to implement mortality did not differ between models but has previously been shown to affect health-economic model predictions.[6,21,64,65]

Although we sought to maintain independence among the participating modeling groups, this was limited for SveDem given that it received technical model coding support and absent for Bedrejo given that it received recommendations on implementing the benchmark scenario from an IPECAD steering member (R.H., also part of the IPECAD modeling group). Some modeling groups participated in the original development of multiple models (R.H., W.L.H.) without being involved in the implementation of the benchmark scenario. In addition, some results were adjusted (RTI-HS) or submitted (FEM) after participation in the workshop.

Conclusions

Our benchmark scenario for a hypothetical set of trial results for an AD treatment was implemented independently in different ways across 9 distinct health-economic models and resulted in a relatively large variation in model predictions. Our results suggest the following cascade of effect: model design drove the choice of outcome, which was associated with relative effect size, which together with natural history (faster progression creating a larger room for improvement) and assumptions on waning and discontinuation drove the gained time in less severe disease states, which drove the QALY gain and costs. We recommend (1) addressing outcome choice and treatment effectiveness assumptions as part of sensitivity analysis, (2) using a standardized table for reporting AD model predictions, and (3) exploring the use of registries for future AD treatments measuring long-term disease progression to reduce the uncertainty of extrapolating trial results in health-economic models.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2024.09.006.

Article and Author Information

Accepted for Publication: September 1, 2024

Published Online: November 21, 2024

doi: https://doi.org/10.1016/j.jval.2024.09.006

Author Affiliations: Alzheimer Centre Limburg, Faculty of Health Medicine and Life Sciences, School for Mental Health and Neuroscience, Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, The Netherlands (Handels); Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Solna, Sweden (Handels, Herring, Aye, Tate, Green, Gustavsson, Wimo, Winblad, Sköldunger, Jönsson); Health Economics, RTI Health Solutions, Research Triangle Park, NC, USA (Herring, Kamgar); Biogen Idec Ltd, Maidenhead, England, UK (Green); Quantify Research, Stockholm, Sweden (Gustavsson); Theme Inflammation and Aging, Karolinska University Hospital, Huddinge, Sweden (Winblad); Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Lund, Sweden (Raket); Community Health Sciences & O'Brien Institute of Public Health, University of Calgary, Calgary, Alberta, Canada (Stellick, Spackman); Health Economics, Policy and Innovation Institute, Masaryk University, Brno, Czech Republic (Hlávka); USC Price School of Public Policy and Schaeffer Center for Health Policy and Economics, Los Angeles, CA, USA (Hlávka, Wei); Basque Health Service (Osakidetza), Debagoiena Integrated Healthcare Organisation, Research Unit, Arrasate-Mondragón, Spain (Mar); Biogipuzkoa Health Research Institute, Donostia-San Sebastián, Spain (Mar); Biosistemak Institute for Health Service Research, Barakaldo, Spain (Mar); Faculty of Engineering, Electronics and Computing Department, Mondragon Unibertsitatea, Mondragon, Gipuzkoa, Spain (Soto-Gordoa); Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands (de Kok, Brück); Care Policy and Evaluation Centre, London School of Economics, London, England, UK (Anderson); Biogen International GmbH, Baar, Switzerland (Pemberton-Ross, Urbich).

Correspondence: Ron Handels, PhD, Department of Psychiatry and Neuropsychology, Maastricht University, Universiteitssingel 40, 6200 MD, Maastricht, The Netherlands. Email: ron.handels@maastrichtuniversity.nl

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: Workshop travel support from Karolinska Institutet Department of Neurobiology, Care Sciences and Society to persons R.H. and W.L.H.

Role of the Funders/Sponsors: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: Synthetic clinical trial data generated for the benchmark scenario are available upon reasonable request. We acknowledge Dr Mike Paulden and Dr Andy Greenshaw for their contribution to the Bedrejo and colleagues model. We acknowledge Raphael Wittenberg, Martin Knapp, and Ron Handels for their contribution to the CPEC model. W.L.H. is a full-time employee of RTI Health Solutions, an independent nonprofit research organization. His compensation is unconnected to the studies on which he works. RTI Health Solutions did not receive funding for the current study. Outside of this study, RTI Health Solutions receives funding pursuant to research contracts with pharmaceutical and biotechnology companies. W.L.H. is

affiliated to research at the Karolinska Institutet and receives no compensation as part of this affiliation. J.H. and Y.W. have received grant support from the National Institutes of Health (under award R01AG062277). The funding agencies were not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

REFERENCES

1. Oliver K, Innvar S, Lorenc T, Woodman J, Thomas J. A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Serv Res*. 2014;14(1):2.
2. Eddy DM, Hollingworth W, Caro JJ, et al. Force I-SMGRPT. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–7. *Value Health*. 2012;15(6):843–850.
3. Vemer P, Corro Ramos I, van Voorn GA, Al MJ, Feenstra TL. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics*. 2016;34(4):349–361.
4. Campbell H, Briggs A, Buxton M, Kim L, Thompson S. The credibility of health economic models for health policy decision-making: the case of population screening for abdominal aortic aneurysm. *J Health Serv Res Policy*. 2007;12(1):11–17.
5. Sogaard R, Lindholt J. Evidence for the credibility of health economic models for health policy decision-making: a systematic literature review of screening for abdominal aortic aneurysms. *J Health Serv Res Policy*. 2012;17(1):44–52.
6. Handels RLH, Green C, Gustavsson A, et al. Cost-effectiveness models for Alzheimer's disease and related dementias: IPECAD modeling workshop cross-comparison challenge. *Alzheimers Dement*. 2023;19(5):1800–1820.
7. Jonsson L, Handels R, Green C. Bridging clinical trials and health economic models in Alzheimer's disease. In: J Prev Alzheimers dis *15th Conference Clinical Trials Alzheimer's Disease. November 29 - December 2, 2022. San Francisco, USA: Posters (Clinical Trial Alzheimer's Disease)*. 2022. Vol. 9(suppl 1):S51-S248.
8. Budd Haeberlein S, Aisen PS, Barkhof F, et al. Two randomized phase 3 studies of aducanumab in early Alzheimer's disease. *J Prev Alzheimers Dis*. 2022;9(2):197–210.
9. van Dyck CH, Sabbagh M, Cohen S. Lecanemab in early Alzheimer's disease. *N Engl J Med*. 2023;388(17):1631–1632.
10. Sims JR, Zimmer JA, Evans CD, et al. Donanemab in early symptomatic Alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA*. 2023;330(6):512–527.
11. den Boon S, Jit M, Brisson M, et al. Guidelines for multi-model comparisons of the impact of infectious disease interventions. *BMC Med*. 2019;17(1):163.
12. Handels R, Jonsson L, Raket LL. Generate synthetic data in R for a hypothetical Alzheimer's disease trial. medRxiv. https://doi.org/10.1101/2024.02.05.24302140
13. Mueller SG, Weiner MW, Thal LJ, et al. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement*. 2005;1(1):55–66.
14. Herring WL, Gould IG, Fillit H, et al. Predicted lifetime health outcomes for aducanumab in patients with early Alzheimer's disease. *Neurol Ther*. 2021;10(2):919–940.
15. Anderson R, Knapp M, Wittenberg R, Handels R, Schott J. Economic modelling of disease-modifying therapies in Alzheimer's disease. London School of Economics and Political Science. https://www.lse.ac.uk/cpec/assets/documents/EconomicmodellingAD.pdf; Published 2018. Accessed August 27, 2024.
16. Green C, Handels R, Gustavsson A, et al. Assessing cost-effectiveness of early intervention in Alzheimer's disease: an open-source modeling framework. *Alzheimers Dement*. 2019;15(10):1309–1321.
17. Wimo A, Handels R, Winblad B, et al. Quantifying and describing the natural history and costs of Alzheimer's disease and effects of hypothetical interventions. *J Alzheimers Dis*. 2020;75(3):891–902.
18. Mar J, Gorostiza A, Ibarrondo O, et al. Economic evaluation of supplementing the diet with Souvenaid in patients with prodromal Alzheimer's disease. *Alzheimers Res Ther*. 2020;12(1):166.
19. Goldman DP, Ermini Leaf D, Sullivan J, Tysinger B. The future elderly model: technical documentation. https://healthpolicy.box.com/v/FEM-Appendix; Published 2018. Accessed August 27, 2024.
20. Herring W, Keenan A, Mauskopf J, Michael T, Wiegand F. The potential economic value of disease-modifying treatments in Alzheimer's disease: patient-level simulation of predementia symptom trajectories. *Value Health*. 2017;20(5):A12.
21. Handels R, Grimm S, Blokland A, et al. The value of maintaining cognition in patients with mild cognitive impairment: the innovation headroom and potential cost-effectiveness of Roflumilast. *Alzheimers Dement*. 2023;19(8):3458–3471.
22. Bruck CC, Wolters FJ, Ikram MA, de Kok I. Projections of costs and quality adjusted life years lost due to dementia from 2020 to 2050: a population-based microsimulation study. *Alzheimers Dement*. 2023;19(10):4532–4541.
23. Vos SJ, Verhey F, Frolich L, et al. Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage. *Brain*. 2015;138(5):1327–1338.
24. Tahami Monfared AA, Ye W, Sardesai A, et al. Estimated societal value of lecanemab in patients with early Alzheimer's disease using simulation modeling. *Neurol Ther*. 2023;12(3):795–814.
25. Andersen K, Lolk A, Martinussen T, Kragh-Sørensen P. Very mild to severe dementia and mortality: a 14-year follow-up - the Odense study. *Dement Geriatr Cogn Disord*. 2010;29(1):61–67.
26. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement*. 2007;3(3):186–191.
27. Darba J, Kaskens L, Lacey L. Relationship between global severity of patients with Alzheimer's disease and costs of care in Spain; results from the co-dependence study in Spain. *Eur J Health Econ*. 2015;16(8):895–905.
28. Dodge HH, Shen C, Pandav R, DeKosky ST, Ganguli M. Functional transitions and active life expectancy associated with Alzheimer disease. *Arch Neurol*. 2003;60(2):253–259.
29. Genworth. Cost of care survey 2022. https://www.genworth.com/aging-and-you/finances/cost-of-care.html. Accessed April 14, 2022.
30. Gustavsson A, Brinck P, Bergvall N, et al. Predictors of costs of care in Alzheimer's disease: a multinational sample of 1222 patients. *Alzheimers Dement*. 2011;7(3):318–327.
31. Jonsson L, Eriksdotter Jönhagen M, Kilander L, et al. Determinants of costs of care for patients with Alzheimer's disease. *Int J Geriatr Psychiatry*. 2006;21(5):449–459.
32. Karlawish JH, Zbrozek A, Kinosian B, Gregory A, Ferguson A, Glick HA. Preference-based quality of life in patients with Alzheimer's disease. *Alzheimers Dement*. 2008;4(3):193–202.
33. Landeiro F, Mughal S, Walsh K, et al. Health-related quality of life in people with predementia Alzheimer's disease, mild cognitive impairment or dementia measured with preference-based instruments: a systematic literature review. *Alzheimers Res Ther*. 2020;12(1):154.
34. Leaf DE. Health utilities index data derived from the 2000 health and retirement study. https://hrsdata.isr.umich.edu/sites/default/files/documentation/data-descriptions/hrs2000_hui3_doc.pdf; 2015. Accessed August 27, 2024.
35. Leibson CL, Long KH, Ransom JE, et al. Direct medical costs and source of cost differences across the spectrum of cognitive decline: a population-based study. *Alzheimers Dement*. 2015;11(8):917–932.
36. Lin GA. Aducanumab for Alzheimer's disease: effectiveness and value; final evidence report and meeting summary. *Institute for Clinical and Economic Review*; Published 2021. https://icer.org/assessment/alzheimers-disease-2021/. Accessed August 27, 2024.
37. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):263–269.
38. Neumann PJ, Hermann RC, Kuntz KM, et al. Cost-effectiveness of donepezil in the treatment of mild or moderate Alzheimer's disease. *Neurology*. 1999;52(6):1138–1145.
39. O'Bryant SE, Waring SC, Cullum CM, et al. Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's Research Consortium study. *Arch Neurol*. 2008;65(8):1091–1095.
40. Potashman M, Buessing M, Levitchi Benea M, et al. Estimating progression rates across the spectrum of Alzheimer's disease for amyloid-positive individuals using national Alzheimer's coordinating center data. *Neurol Ther*. 2021;10(2):941–953.
41. Robinson RL, Rentz DM, Andrews JS, et al. Costs of early stage Alzheimer's disease in the United States: cross-sectional analysis of a prospective cohort study (GERAS-US)1. *J Alzheimers Dis*. 2020;75(2):437–450.
42. Sapkota S, Ramirez J, Yhap V, Masellis M, Black SE. Alzheimer's Disease Neuroimaging Initiative. Brain atrophy trajectories predict differential functional performance in Alzheimer's disease: moderations with apolipoprotein E and sex. *Alzheimers Dement (Amst)*. 2021;13(1):e12244.
43. Tahami Monfared AA, Tafazzoli A, Chavan A, Ye W, Zhang Q. The potential economic value of lecanemab in patients with early Alzheimer's disease using simulation modeling. *Neurol Ther*. 2022;11(3):1285–1307.
44. van Oudenhoven FM, Swinkels SHN, Hartmann T, Soininen H, van Hees AMJ, Rizopoulos D. Using joint models to disentangle intervention effect types and baseline confounding: an application within an intervention study in prodromal Alzheimer's disease with Fortasyn Connect. *BMC Med Res Methodol*. 2019;19(1):163.
45. Vermunt L, Sikkes SAM, van den Hout A, et al. Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype. *Alzheimers Dement*. 2019;15(7):888–898.
46. Wei Y, Heun-Johnson H, Tysinger B. Using dynamic microsimulation to project cognitive function in the elderly population. *PLoS One*. 2022;17(9):e0274417.
47. Wilson RS, Aggarwal NT, Barnes LL, Bienias JL, Mendes de Leon CF, Evans DA. Biracial population study of mortality in mild cognitive impairment and Alzheimer disease. *Arch Neurol*. 2009;66(6):767–772.
48. Wolters FJ, Tinga LM, Dhana K, et al. Life expectancy with and without dementia: a population-based study of dementia burden and preventive potential. *Am J Epidemiol*. 2019;188(2):372–381.
49. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med*. 2004;256(3):183–194.
50. Tate AE, Bouteloup V, van Maurik IS, et al. Predicting sojourn times across dementia disease stages, institutionalization, and mortality. *Alzheimers Dement*. 2024;20(2):809–818.
51. Handels R, Jonsson L, Green C, et al. Structural uncertainty and cross-validation of decision models in the ad setting: answering questions from policy

makers on the reliability of health-economic models. *Alzheimers Dement*. 2019;15(7S_Part 31). P1637-P1637.
52. Hernandez L, Ozen A, DosSantos R, Getsios D. Systematic review of model-based economic evaluations of treatments for Alzheimer's disease. *Pharmacoeconomics*. 2016;34(7):681–707.
53. Nguyen KH, Comans TA, Green C. Where are we at with model-based economic evaluations of interventions for dementia? a systematic review and quality assessment. *Int Psychogeriatr*. 2018;30(11):1593–1605.
54. Handels R, Herring WL, Grimm S, et al. New IPECAD open-source model framework for the health technology assessment of early Alzheimer's disease treatment: development and use cases [Published online April 7, 2024]. Value Health. https://doi.org/10.1101/2024.04.05.24305373
55. Roybal Center for Health Policy Simulation. Future elderly model (FEM) files. Health and retirement study. https://hrsdata.isr.umich.edu/data-products/future-elderly-model-fem-files. Accessed February 16, 2024.
56. Husereau D, Drummond M, Augustovski F, et al. Consolidated health economic evaluation reporting standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *BMC Med*. 2022;20(23):1–8.
57. Landeiro F, Morton J, Gustavsson A, et al. Health economic modeling for Alzheimer's disease: expert perspectives. *Alzheimers Dement (N Y)*. 2022;8(1):e12360.
58. Hlavka JP, Kinoshita AT, Fang S, Hunt A. Clinical outcome measure crosswalks in Alzheimer's disease: a systematic review. *J Alzheimers Dis*. 2021;83(2):591–608.
59. Jonsson L, Wimo A, Handels R, et al. The affordability of lecanemab, an amyloid-targeting therapy for Alzheimer's disease: an EADC-EC viewpoint. *Lancet Reg Health Eur*. 2023;29:100657.
60. Wahlberg K, Winblad B, Cole A, et al. People get ready! A new generation of Alzheimer's therapies may require new ways to deliver and pay for healthcare. *J Intern Med*. 2024;295(3):281–291.
61. Green C, Shearer J, Ritchie CW, Zajicek JP. Model-based economic evaluation in Alzheimer's disease: a review of the methods available to model Alzheimer's disease progression. *Value Health*. 2011;14(5):621–630.
62. Gustavsson A, Pemberton-Ross P, Gomez Montero M, Hashim M, Thompson R. Challenges in demonstrating the value of disease-modifying therapies for Alzheimer's disease. *Expert Rev Pharmacoecon Outcomes Res*. 2020;20(6):563–570.
63. Wright AC, Lin GA, Whittington MD, et al. The effectiveness and value of lecanemab for early Alzheimer disease: a summary from the Institute for Clinical and Economic Review's California Technology Assessment Forum. *J Manag Care Spec Pharm*. 2023;29(9):1078–1083.
64. Skoldunger A, Johnell K, Winblad B, Wimo A. Mortality and treatment costs have a great impact on the cost-effectiveness of disease modifying treatment in Alzheimer's disease-a simulation study. *Curr Alzheimer Res*. 2013;10(2):207–216.
65. Aye S, Jonsson L, Gustavsson A, Tate A, Ptacek SG, Eriksdotter M. Effect of mortality in cost-effectiveness modeling of disease-modifying treatment for Alzheimer's disease. *Alzheimers Dement (Amst)*. 2023;15(1):e12422.