Supplemental material 2 corresponding to manuscript: "Cost-effectiveness Models for Alzheimer's disease and related dementias: IPECAD Modelling Workshop Cross Comparison Challenge". See www.ipecad.org.

International Pharmaco-Economic Collaboration on Alzheimer's Disease (IPECAD) Modeling Workshop

September 19-20, 2019

Venue: Vår Gård Saltsjöbaden, Stockholm, Sweden.

Meeting Minutes

Introduction

About 20 researchers from academia, consultancy and the life sciences/pharmaceutical industry with a common interest in Alzheimer's disease (AD) modeling convened for a two-day workshop in Stockholm. The aim of this workshop was to bring together those who are research active in the field of AD decision modeling, in terms of developing modeling methods in the context of cost-effectiveness analyses and health technology assessment, or disease progression models that may be used in this way.

The stated long-term goal was to create a dedicated forum for exchange of ideas on health economic modeling in AD, to discuss and align on key methodological issues and to foster the development of robust and transparent models to guide decision making around novel treatment scenarios/pathways for AD. The ambition was for this workshop to be the first of an ongoing series of meetings and activities to further the development of modeling methods, and the identification of data to populate decision-models, in the area of AD. Whilst the initial focus was on AD, we hoped to be able to extend the scope of our future activity into and across related dementias, and across the full diagnostic and treatment pathways associated with dementia. An additional aim was to encourage and facilitate, where possible, a greater number of open access modeling studies/frameworks in AD.

Agenda

Fourteen of the participants reported on work from their respective research groups based in different parts of the world. This included a wide range of models with varying research questions, contexts, outcomes, methods and underlying data. Some included a wide range of risk factors and biomarkers whereas others focused on key symptoms including different measures of cognition, behavior and function. Several models were based on the United States National Alzheimer's Coordinating Center (NACC) data while others included data from the Aging, Demographics and Memory Study (ADAMS), the GERAS study, the Alzheimer's Disease Neuroimaging Initiative (ADNI) study and the Health and Retirement Study (HRS). Outcomes included resource utilization, costs, health utility, mortality, time to institutionalization and dependence. Both cohort and individual

patient simulation models were represented. Other presentations included overviews of challenges and proposed methods in AD modeling as well as presentations on related issues to inform AD models (see detailed agenda below).

The mount hood challenge

The Mount Hood Diabetes Challenge Network is dedicated to promoting an exchange of ideas and information between those developing and using health economic diabetes simulation models.

The Network has run a diabetes computer simulation modeling conference bi-annually since 1999. A major focal point of their conferences are comparisons of health economic diabetes models both in terms of their structure and performance. Michael Willis (Swedish Insitutet for Health Economics, IHE) from the Mount Hood Challenge Network gave a presentation on their experiences in organizing their series of workshops, where modelers have convened to compete, compare and validate models in diabetes. He advised that a modeling challenge scenario was typically agreed in advance and participants presented their individual model simulation results at the meeting. Such workshops may aid knowledge sharing, improve the quality of models, help their validation and engage end users. Dr. Willis encouraged similar activity in the AD context but stressed that substantial planning is required for a successful workshop. In particular, he recommended that any challenge scenario results from contributing models/groups be compiled in advance of the planned meeting to enable efficient comparison across models. Resulting outcomes of the meeting may include a meta-database of models, reporting standards and modeling guidelines.

Summary of issues from presentations and discussion sessions

AD modeling started about 25 years ago with symptom-based models (primarily cognition) and have advanced to simulating levels of care (or institutionalization), progression across multiple symptom domains (e.g., cognition, behavior, function), and most recently the disease pathology and biomarkers of underlying disease progression. This steady increase in complexity has been coupled with more advanced modeling techniques and increasing data needs.

Past reviews of AD models have offered a multitude of recommendations for the improvement of future models, and these viewpoints were summarized at the beginning of the workshop (see agenda for presentations from Will Herring and Ron Handels).

We considered the focus of drug development in **earlier stages** of AD and heard how this is paralleled by an increased focus on epidemiological data on natural disease progression in these early pre-clinical and pre-dementia stages. In the absence of clear symptoms, the outcomes in early disease are different and focused on biomarkers, risk factors and sensitive composite measures. Because these outcomes are intermediate endpoints (potential surrogate outcomes) one challenge is on establishing the clinical relevance of such measures, with a requirement to link (translate) these outcomes to more immediate clinically relevant and policy-relevant outcomes in later stages of disease. Still, the evidence linking early- and late-stage markers is largely missing. Although epidemiological data are available across all stages of AD, separate studies tend to focus on different stages of disease. Moreover, their designs often differ where early stages are mostly studied in population-based elderly cohorts and highly selected biomarker cohorts whereas later stages are generally studied in clinical patient cohorts identified in memory clinics. Models tend to utilize data from different cohorts but instead of bespoke linkage for individual models an alternative idea may be to link data sources together in a separate effort and make the resulting data available for model developers.

We recognized that AD is a highly **heterogenous disease** with a wide range of individual outcomes, often reported using measures that may be considered as having limitations (e.g., cognitive measures). Therefore, use of average outcomes, and mean scores, may not be representative of a common individual patient's profile or outcome, thus limiting the generalizability and use of models for individual level model predictions. Models are commonly developed to address policy questions, and data to inform models (e.g., effectiveness) is typically at a sample (policy) level. Models may be able to offer an estimate of the average (costs, outcomes) and differences between scenarios, but it is recognized that they are less accurate and less useful in predicting individual trajectories of individual patients. This issue may partly be attenuated by more careful patient selection, not least with help from better biomarkers. The meeting heard presentations across a range of topics related to this area (see agenda). For example, presentations were heard on patterns cognitive decline and on some of the latest findings within the development and use of biomarkers. Biomarkers may both play an important role in identification of patients at risk of progression and as a relevant outcome in trials as well as models.

The **external validity** of available models is believed to be largely unknown, at least in part due to lack of data available for validation. The results of model validation exercises were presented, but comparable data showed high heterogeneity in subject populations and outcomes and it was found difficult to find matching cohorts to the data used within the models.

On **transparency**, all models ideally are published with sufficient information for them to be understood, scrutinized and validated by other researchers and decision-makers. Simpler approaches could sometimes be preferable, but the appropriate level of complexity is dependent on context. We discussed the role for both simple and complex models.

The **discrepancy between onset of disease versus diagnosis** is interesting and particularly challenging to disease modeling. Because of its progressive nature, the onset of AD is defined by a complex combination of pathological and clinical constructs, and is therefore difficult to place in time. Moreover, the timing of diagnosis likely depends on the timing and setting of the assessment. Thus far, dementia diagnosis is probably considered a relevant outcome to most stakeholders, while we questioned whether it is necessary for the simulation of disease progression. Biomarkers and more granular measurements of early symptoms may show to be better predictors of disease progression than the presence of a diagnosis.

The **long-term impact** of disease modifying treatment will most probably rely on assumptions on the potential attenuation of effects beyond the duration of the trial as well as continuing effects after treatment discontinuation. We discussed that such assumptions should ideally be considered in model scenario analyses but may also specifically be studied in post-authorization studies. A related issue is the role of stopping rules and the definition/assessment of treatment response which, together with the definition/identification of eligible patients, are expected to be highly influential on the cost-effectiveness of treatments.

Selection of outcomes varies across all types of studies in AD, including modeling studies. Past appraisals have stressed the value of considering the hallmark symptoms: cognition, behavior and function. Delay in institutionalization has often been selected as both a clinically relevant outcome, but also an outcome relevant to payers due to its associated costs. However, institutionalization is dependent on context, which is why the concept of dependence has been suggested as an outcome that may be more stable across countries but nevertheless could be connected to country-specific care patterns/indications. The selection of outcomes also will likely depend on the possibility of linking trial endpoints to outcomes of relevance to payers and other stakeholders, quite often involving intermediate or surrogate endpoints.

The impact of **model uncertainty** should ideally be assessed in sensitivity analyses but may further be assessed by analyses of expected value of perfect information (EVPI). Such analyses may inform where additional efforts should be made to fill data gaps and lower uncertainty.

Group discussions

Participants divided into three groups to answer two sets of questions. The questions and group recommendations are summarized below.

Which are the most important areas of development for the field of health economic modeling in AD? What is the role of IPECAD in this?

1. Address the lack of comparability across function/Activities of Daily Life (ADL) via the mapping of ADL measures. Conduct a review of available outcome measures and compile data on their validity, reliability, usability, comprehensiveness, strengths, weaknesses in different stages of AD. Compare the inclusion of specific domains with measures and provide guidance on preferred measures. This is currently being done for 5-10 measures within the MODEM project.
2. Identify priority data gaps on surrogate outcomes, their linkage to biomarkers and payer outcomes and needs for modeling.
3. Provide guidance on model validation including minimum reporting standards including recommendations on outcomes that should be reported on to increase comparability across models. Should be aligned with ISPOR guidance.
4. Stitch databases together to form a composite AD continuum dataset.
5. Offer a repository for data (including info on data accessibility), a model catalogue, and mapping of outcome measures on www.ipecad.org.
6. Develop a gold standard approach for modeling mortality.

Should we organize a model cross-validation exercise (similar to the Mount Hood Challenge)?

There was broad support for an initiative in AD similar to that seen in Diabetes (Mt Hood).

1. Avoid too high ambition in the first round. Focus on comparison across currently available models.
2. Define a simple scenario with a list of inputs and preferred outcomes to be modeled. Ask for participants to apply the scenario as far as possible with currently available models (to avoid additional work on redesign).
3. Ask modeling groups to present at next workshop and compare findings. Ask participants to submit results in advance of meeting for compilation.
4. Invite expert on machine learning techniques and their potential in AD modeling.
5. In forthcoming rounds, share a data cut from NACC with modeling groups and ask them to simulate the next follow-up to be compared with a new data cut once available.

Conclusions

The organizers suggested three potential work streams on which the IPECAD modelling workshop participants may focus:

Tools and resources:

- Scale catalogue, crosswalks, recommendations on scale selection.
- Data repository, expand on Real world Outcomes across the Alzheimer's Disease spectrum for better care: Multi-modal data Access Platform (ROADMAP) data cube, access to data.
- Model repository with details on available models.

Best practices, potential editorial:

- Translating intermediate to final outcomes.
- Key model endpoints: e.g., diagnosis, dependency.
- Stakeholder engagement.
- Reporting standards, apply ISPOR standard to AD modeling.

Validation:

- Define outcome of interest: e.g., conversion to dementia, time by disease stage, mortality.
- Define scenarios, patient population, treatment.
- Define a tadpole challenge on data prediction.

Next steps

the workshop concluded with a discussion of next steps and an intention to convene a second workshop in the second half of 2020, in terms of an initial model-comparison challenge.