Supplemental material 4 corresponding to manuscript: "Cost-effectiveness Models for Alzheimer's disease and related dementias: IPECAD Modelling Workshop Cross Comparison Challenge". See www.ipecad.org.

WORKING PROCEDURES AND POST-WORKSHOP REQUESTED INFORMATION

WORKING PROCEDURES

A selection of the organizers (AG, CG, RH) checked the outcomes send in by the contributors in an iterative approach on transparency, suspected errors or misinterpretations and plausibility of the results. Issues raising from these checks were communicated to the specific modelling group with a request to clarify and resubmit if necessary. See supplemental 2 for details.

The contributions were compiled into a database and summarized in graphs and tables. Due to a relatively large variation in the ICER the organizers suggested to focus on mean person-years alive and in disease stage. Model outcomes were grouped based on their scenario starting in MCI (A) or starting in mild dementia (B).

The database and the summary graphs and tables on the selected outcomes were shared with all contributors. The contributors were asked to share a pre-recorded presentation of about 15 minutes explaining their model's background for all contributors to review, on a voluntary basis. During the workshop each contributor had the opportunity to present the model (8 minutes) and a summary on the selected outcomes was presented by the organizers. All contributors were invited to discuss 3 questions: 1) How did you manage applying the benchmark scenario? 2) How do you interpret the variation in the selected outcomes? 3) What are the strongest drivers of this variation? Please focus on design choices and parameterization (in terms of population, setting, mortality, disease progression, treatment operationalization and model assumptions).

Pre-defined checklist and results:

1. The model description:
    a. will be checked for transparency and clarity;
    b. Will be cross-checked against the criteria of the common scenario.
2. The model results will be checked to identify suspected errors or misinterpretations of the instruction. This will be done by:
    a. checking the deviation to the range of expected plausible results (i.e. unexpected discrepancy), unit and format;
    b. Checking whether the person-years alive are plausible given the model description on mortality;
    c. Checking whether the costs are plausible given the model description on cost inputs.
    d. Checking whether the state trace is plausible.
3. Issues raising from both above checks will be communicated to the specific modelling group with a request to clarify and resubmit if necessary. Any additional clarification to the instruction will be communicated to all participants.
4. The models will be grouped based on similarities in their model specification. This will take into account starting population, domains reflected, modelling type, and possible other factors that will be identified along the way. The purpose of this comparison is to accommodate efficient discussion at the modelling workshop by enabling comparing groups of models.
5. The distribution of outcomes will be described. Individual outcomes may be plotted against differentiating model characteristics. All analyses will be descriptive but mean and variation measures across models may be calculated for reference.
6. A results document will be shared with all participants.

7. A presentation will be made at the cross-comparison workshop – based around the summary document – with key discussion points suggested.

These model checks revealed several issues, which were solved by interaction with the contributors. These included unclarity on e.g. the definition of the starting population or implementation of the treatment and impossible mean time in states due to inconsistency within the estimates or with the time horizon. All were solved or marked as a deviation from the benchmark scenario due to impossibility to implement it in the model.

POST-WORKSHOP REQUESTED INFORMATION

(1) reflecting on the model characteristics and model outcomes shared in the workshop how do you think models compare?

(2) How would you compare your model to others, either to all models/broadly or to specific models that are clear 'comparators'?

(3) What are your take-aways or learnings from the workshop after having seen your model outcomes in comparison to the other model outcomes?