

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol
AUTHORS	Solomon, Alina; Kivipelto, Miia; Molinuevo, José; Tom, Brian; Ritchie, Craig

VERSION 1 – REVIEW

REVIEWER	Betty Tijms Alzheimer Center and Department of Neurology, VUmc, Amsterdam Neuroscience, Amsterdam, The Netherlands.
REVIEW RETURNED	03-Jan-2018

GENERAL COMMENTS	<p>This study protocol describes the protocol for a longitudinal study on predementia Alzheimer's disease, which is part of the European Prevention of Alzheimer's Dementia. This study will be very important for the field, and overall it is well described, but some issues require further clarification:</p> <ul style="list-style-type: none">-The key concept of this project is the 'probability continuum spectrum' for AD dementia, but it remains unclear how this continuum is defined. Throughout the manuscript it remains vague whether this continuum is assumed and that the precise definition is one of the objectives of the present study (in this case this should be part of the objective descriptions of the study at p6) or whether the project already has a working definition of this continuum (in which case this working definition should be clearly stated with a corresponding rationale), and that this will be refined in this project.-One of the main objectives of the present project is to define a probability continuum for AD dementia. The primary outcome measure (to determine this probability continuum?) will be composite scores of neuropsychological test. Clinically relevant measures, i.e., diagnosis of dementia, MMSE and CDR scores are listed as 'additional assessments' in tables 4 and 5. Progression to dementia however is the most relevant primary outcome measure, and CDR and MMSE also are clinically relevant measures, and so I find the strategy to list these outcomes as 'additional' measures amongst factors such as 'handedness' somewhat difficult to follow.- A data analysis plan including more detailed planned statistical analysis is missing. This part is important, as one of the study objectives is disease modelling, but no details of potential disease models that will be tested (and validated) are given. When the aim is to develop novel disease models, the authors should make this more clear in the text, as well as what the strategy to develop disease model adequate disease models will be. Such a data analysis plan should also more clearly describe the analysis strategy to define a probability continuum.
-------------------------	--

-Although this protocol is specific for the LCS substudy, the primary objective of this study is the creation of a 'trial ready' cohort for the PoC trial. However, the Bayesian adaptive aspect of this PoC is not explained, and so it is unclear how such alignment will be ensured, as well as how this could complement or rather be a potential limitation for the secondary objectives of the LCS.

-p19 section 'Sample Size'. In this paragraph it is unclear how many subjects are already eligible for the LCS study, and so it is unclear whether it is feasible to establish an ongoing cohort of n=6000 within time period of 4 years of this study. Similarly, subjects will be (temporarily) lost when they are selected and participating in the PoC trial, and so it is conceivable that this group will be the largest to be re-sampling (possibly this is part of the rationale for over- under sampling, but not this is not clear from the manuscript text). It would be helpful to provide estimations of the number of subjects expected to be included to maintain n=6000 [and whether these are selected specifically to replace lost subjects; or whether subjects will not be selected and continue to be recruited according to inclusion criteria].

minor:

-start and end dates of the project are missing

-p7 "(Table 1) to ensure fast recruitment of a probability-spectrum population covering the entire continuum of probability for AD dementia development' It is not evident that the use of parent cohorts will ensure fast recruitment, and/or that this will cover the entire continuum.

-p7 paragraph starting with "To ease the search process, " a 'flexible search algorithm' is stated as a tool to select subjects for the LCS study, but, at this point it is unclear what inclusion criteria for subjects and so this part is difficult to understand. The authors may consider to move this part of the manuscript after subject inclusion/exclusion criteria are discussed, as well as providing a clearer explanation of the 'flexible algorithm' (is this like a flow chart?).

-p8 : the PrePAD velocity section is unclear, it seems that in this case only amyloid positive subjects will be eligible for LCS? It is also unclear why this approach would 'optimise the balance in the LCS towards as large a proportion as possible to be amyloid positive by existing thresholds' (this sentence seems to imply that not a balance, but a bias is the purpose. If an as 'large a proportion as possible to be amyloid positive' is important, then it would make more sense to use amyloid positivity as an inclusion criterium for LCS?).

-p10 "Population size will be maintained over time by continuous refilling from the PCs or via PrePAD Velocity" please clarify whether it is meant that all PCs will be continuously screened for eligible participants in an un-selective manner, or, whether subjects will be selected to match 'lost/to be replaced' participants in their subject characteristics.

-p10 section "EPAD LCS participant selection process" the first part of this section discusses generically the need for disease modelling, this should be moved and more explicitly explained as part of the analysis plan, and this paragraph should focus on clearly explaining to the selection process. Here a working definition of the probability spectrum is necessary, as it reads now it as if estimated disease probabilities from to-be-developed

	<p>models are already used it as a selection criterium at the same time.</p> <p>-p16 Explanation of primary, secondary and exploratory outcome measures should be moved to the general section that gives an overview of all outcomes, as this explanation is not specific for cognitive outcomes.</p> <p>-p17 Sudden change of future to past tense in Neuroimaging paragraph (this paragraph is difficult to follow, mostly due to the lack of data analysis plan. This section seems to imply that subjects will be compared on amyloid / APOE status, it is unclear what how this relates to the primary/secondary objectives of the present project).</p> <p>-p19 last sentence: 'after 500, 1000, 2000' participants?</p> <p>-p21 section 'informed consent' (and or p22 privacy of personal data; and or dissemination plan p23): the authors stated that data will be shared with 'researchers', please clarify whether 'researchers' indicates the global scientific community, or will this be restricted to EPAD consortium members; in the first case does the informed consent specifically mentions such wide data sharing?</p> <p>p25. "because different contributions from various components in each dimension may results in a similar overall probability, flexible algorithms are more suitable than simple cut-offs" this statement is inaccurate, as the use of cut-offs also allows to combine various components as part of a continuum (which in a more simple way is already done in research criteria when defining multiple disease stages (0, 1, 2) in preclinical/prodromal AD).</p> <p>-p25 "... and for providing a sufficient number of potential trial participants" it is not evident why a probability continuum to estimate a subject's dementia risk necessarily results in 'sufficient number of potential trial participants'.</p> <p>Table 1: Is this table necessary? Possibly, it would be more informative to add/merge this information to figure 1, as there is detail lacking in this table (e.g., the numbers of cohorts, references to the studies referred to), as well as the mixed presentation of types of cohorts with the cohort eligibility criteria.</p> <p>Table 3: Please provide in the manuscript text a rationale for under- and over sampling (to ensure high amyloid positive proportion ...?), as well as the corresponding proportions. it is confusing to list factors that will be considered to be important for the development of the probability continuum in the same table as the selection process for participants.</p>
--	--

REVIEWER	Deborah Gustafson State University of New York - Downstate Medical Center
REVIEW RETURNED	12-Jan-2018

GENERAL COMMENTS	<p>This paper describes the Protocol development and implementation for EPAD LCS, a cohort designed to characterize adults age 50 years and older in a rigorous manner, consistent with that for RCT enrollment. The goal is to identify candidates for secondary prevention trials of AD dementia, defined as AD neuropathology and clinical, symptomatic AD.</p> <p>Some comments are as follows:</p>
-------------------------	---

	<p>Page 4 – Strengths and Limitations – There are no limitations listed.</p> <p>Secondary prevention is used too broadly in this manuscript. It is used to describe both preclinical biomarker only-based as well as clinical symptom-based intervention windows. This term should be defined and used consistently. Then, seemingly in contrast, at the end of the paper, Page 23, line 48, re: Dissemination, it is stated that EPAD LCS will help researchers, ‘... improve their understanding of the early, pre-dementia phase of AD, and facilitate collaborations.’ Note: ‘early, pre-dementia phase of AD’. Again, consistency regarding the exposure and outcomes windows needs to be enhanced.</p> <p>Page 7, line 51 - Flexible search algorithm is not defined here or later in the manuscript.</p> <p>Page 8, line 22 – without knowing the details of the flexible search algorithm, and noting Table 2, Inclusion and exclusion criteria, mention of amyloid positivity is a surprise. In addition, on page 10, line 41, there is mention of limiting selection bias by not overspecifying inclusion criteria. Some idea of the algorithm is finally provided in Table 3, but remains a black box.</p> <p>Table 2 – Eligibility criteria: it is unclear as to how the second bullet point is an eligibility criterion, rather it describes an aspect of protocol. In addition, the fifth bullet point is not only eligibility but reasons for the eligibility criterion. Reasons should be in the text, not in the table.</p> <p>Fulfill - check spelling</p> <p>Exclusion criteria - are there definitions for these criteria, e.g., uncontrolled diabetes or hypertension, ‘severe’ vision impairment, etc.</p> <p>Page 10, line 15 and elsewhere, there is reference to ‘continuous refilling of the cohort’ to maintain N=6000. What does this mean?</p> <p>Page 10, last sentence of the first paragraph, beginning line 24 is very difficult to understand. Please clarify.</p> <p>Page 12, Table 3 – it is interesting to note that only BMI is listed as a risk phenotype and not blood pressure or lipid levels or other risk factors for that matter, including control of vascular risk. Instead several of these are listed as Outcomes in Table 4. In addition, Secondary neuroimaging outcomes listed in Table 4 are only volumetric. There is no mention of amyloid except for CSF levels. This also raises the question of whether all enrolled in the EPAD LCS will be imaged and experience LP.</p> <p>Page 16, line 14 – modifiable is a word and does not need to be in “ “</p> <p>Page 16, lines 40-42 – be careful of grammar, consistent use of singular, plurals e.g., ‘... outcome measures may be used as... primary endpoint...’</p> <p>Page 17, lines 31-33 – again a mention of AD as defined by brain amyloid, yet no mention of amyloid imaging</p>
--	---

	<p>Page 18, line 51 – it is stated that the only data source for this study is the EPAD LCS, yet there are some data coming from community-based PCs, yes? At least for recruitment purposes.</p> <p>Page 19, line 3 – It is stated that there will be central neuroimaging reads, yet no location provided, while central CSF and genetics laboratories are provided. In addition, will there be a centralized data entry, or by site?</p> <p>Page 20, first part of line 14, 'As EPAD LCS is not a Clinical Trial of Investigational Medicinal Product (CTIMP), is not needed. Why even bring this up? Also, do you need to define an AE or SAE?</p> <p>General comments. It appears that ADNI has a similar goal with somewhat similar approach, albeit narrower in participant recruitment strategies and sample size than EPAD LCS. There is no mention of ADNI in the Discussion. As an example, the 2016 Nature Commun paper by Iturria-Medina Y et al., should be cited. In addition, there are other consortia attempting to ascertain this space using only existing cohort data, e.g., IALSA/Maelstrom, STROKOG, etc. It may be worth mentioning this in the Discussion as well. This is a very long paper. This reviewer wonders if all of the detail toward the end of the paper re: Human Subjects, Privacy, etc. is necessary and/or this lengthy. It reads very much like a grant proposal.</p> <p>A list of EPAD LCS would be helpful.</p>
--	---

REVIEWER	Stéphane Epelbaum University Hospital Pitié Salpêtrière, AP-HP, Paris, FRANCE
REVIEW RETURNED	13-Jan-2018

GENERAL COMMENTS	<p>The manuscript by Alina Solomon et al describes the EPAD Longitudinal Cohort Study (LCS) protocol. EPAD LCS is one of the most important effort worldwide to tackle the complex problematic of preclinical Alzheimer's disease diagnosis and care. This means that this manuscript is certainly worthwhile and should be published. I am however concerned about a few points that I think must be improved in a revised draft:</p> <p>1) The "flexible" recruitment procedure is not clear. Could the authors give some concrete examples of the balancing committee choices</p> <p>2) No data is provided on the starting date of inclusions, nor on the expected date of the end for EPAD LCS (at least, the duration of the grant by IMI should be mentionned). As some centers have started recruiting for EPAD, their curves of inclusions should be provided. How many centers are actively recruiting at this point is also lacking.</p> <p>3) considering merged protocols such as Aetionomy, which is in its ending phase, how do the investigators of EPAD see the sharing of data that have not yet been locked in the EPAD database to the collaborating investigators. -considering the latest news on the EPAD website, the recruiting centres are "ten centres in six European countries". Which can hardly be considered as "pan-european" -On what figures do the authors base their assumption that "A constant sample size of approximately 6,000 participants for the</p>
-------------------------	---

	<p>EPAD LCS is considered sufficient for a readiness cohort that should provide approximately 1,500 participants for the EPAD PoC trial." ? This should be backed by prior evidence of screening failure rates and referenced.</p> <p>Also, no reference is made to other projects worldwide (eg Brain Health registry) to identify putative research participants in the preclinical/prodromal phases of AD. This is a serious flaw in the manuscript.</p> <p>The endpoints of the study are defined but not the investigators hypothesis about them.</p> <p>In summary, I think EPAD LCS is a great study that is an example of the collaborative strength of european expert centres. This manuscript however falls short in demonstrating it. I would gladly read a revised version of the manuscript when the few points above have been adressed.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reply to reviewers' comments

We would like to thank the reviewers for their very helpful comments that have contributed to improving the manuscript. We have done an extensive revision of the manuscript to shorten it and make it clearer. We have also made changes to reflect the most recent amendment to the EPAD LCS protocol. Based on feed-back from research participants indicating their commitment to longer-term active contribution to the study, participants will not be deselected from EPAD LCS and will be able to remain in the study for as long as they wish to. Given the extensive assessments, we have also decided to reduce the burden for participants who during the course of LCS maintain a low likelihood of being invited to the PoC trial for various reasons (e.g. develop health conditions or risk factors precluding safe trial participation, or do not show any impairment/decline in cognition and AD biomarkers). Starting from their year 1 visit, such participants may have the possibility to opt out of the yearly MRI and CSF sampling.

All changes are tracked in the manuscript, and we have also provided a clean revised version.

Reviewer: 1

Reviewer Name: Betty Tijms

Institution and Country: Alzheimer Center and Department of Neurology, VUmc, Amsterdam Neuroscience, Amsterdam, The Netherlands.

Please state any competing interests or state 'None declared': My affiliation (VUmc Alzheimer center) is part of the EPAD consortium, I am personally not involved in this project.

This study protocol describes the protocol for a longitudinal study on predementia Alzheimer's disease, which is part of the European Prevention of Alzheimer's Dementia. This study will be very important for the field, and overall it is well described, but some issues require further clarification:

-The key concept of this project is the 'probability continuum spectrum' for AD dementia, but it remains unclear how this continuum is defined. Throughout the manuscript it remains vague whether this continuum is assumed and that the precise definition it is one of the objectives of the present study (in this case this should be part of the objective descriptions of the study at p6) or whether the project already has a working definition of this continuum (in which case this working definition should be clearly stated with a corresponding rationale), and that this will definition will further be refined in this project.

Response: Thank you for pointing out that this may be confusing. We have edited the manuscript to improve clarity. We use “probability continuum” in its common statistical meaning, i.e. the continuous line that covers everything from low to high probability of developing AD dementia. By “probability-spectrum population” we mean a population that includes individuals with low probability, those with high probability, and everyone in between. This is important to distinguish the EPAD LCS population from other types of populations, e.g. general populations dominated by low-risk individuals, or clinical or trial populations dominated by high-risk individuals.

There is currently no disease model that would allow us to accurately determine exactly where along this continuum a certain individual is located at a certain point in time. Thus, we want to avoid constricting the cohort from the start into one of the current, limited disease models. As specified in Objective 2, we do aim to develop more accurate disease models, and we also aim to continuously improve them during LCS as more data accumulates over an increasing time period. To make this possible, the EPAD LCS selection process has a very wide range of recruitment sources and established mechanisms for active and continuous data monitoring and balancing of the cohort. This is different from traditional approaches in AD research, but after so many drug trial failures it is perhaps time to try something new.

-One of the main objectives of the present project is to define a probability continuum for AD dementia. The primary outcome measure (to determine this probability continuum?) will be composite scores of neuropsychological test. Clinically relevant measures, i.e., diagnosis of dementia, MMSE and CDR scores are listed as ‘additional assessments’ in tables 4 and 5. Progression to dementia however is the most relevant primary outcome measure, and CDR and MMSE also are clinically relevant measures, and so I find the strategy to list these outcomes as ‘additional’ measures amongst factors such as ‘handedness’ somewhat difficult to follow.

Response: The outcomes of EPAD LCS were carefully chosen to have the greatest possible sensitivity to the subtle changes that occur during the very early stages of AD. MMSE or CDR are notorious for lacking such sensitivity, but we have included them among LCS assessments because of their wide-scale use in clinical practice.

The primary outcome of EPAD LCS was also chosen to align with the primary outcome for the PoC trial. The problems with using progression to dementia as a primary outcome in clinical trials in early AD stages have been pointed out in recommendation documents from regulatory authorities (FDA, EMEA).

We have edited the table (now Table 3) and grouped the list of Other assessments into Clinical, Biomarkers and Other to make it easier to follow.

- A data analysis plan including more detailed planned statistical analysis is missing. This part is important, as one of the studies objectives is disease modelling, but no details of potential disease models that will be tested (and validated) are given. When the aim is to develop novel disease models, the authors should make this more clear in the text, as well as what the strategy to develop disease model adequate disease models will be. Such a data analyses plan should also more clearly describe the analysis strategy to define a probability continuum.

Response: We have edited the Statistical analysis section to include more details about the disease modelling work.

-Although this protocol is specific for the LCS substudy, the primary objective of this study is the creation of a ‘trial ready’ cohort for the PoC trial. However, the Bayesian adaptive aspect of this PoC is not explained, and so it is unclear how such alignment will be ensured, as well as how this could complement or rather be a potential limitation for the secondary objectives of the LCS.

Response: We have included an explanation in the manuscript section describing the LCS participant selection process (subtitle “Novel flexible approach to selection”). The adaptive PoC trial is designed to include the possibility to have multiple active experimental drugs assessed concurrently, with a shared placebo group. Pre-specified trial interim analyses may also affect participant accrual or stopping/continuing trial arms. Trial arms may be started/stopped at different times. The LCS participant selection process has a built-in flexibility that aligns with the adaptive design needs for the trial. The LCS primary outcome is the same as the trial primary outcome. Depending on the tested drugs, secondary and exploratory outcomes can also be shared.

The main limitation of trial-ready populations is that they end up being dominated by high-risk individuals, which limits disease modelling work. The data monitoring and cohort balancing procedures during LCS participant selection and cohort refilling were designed specifically to mitigate this limitation.

-p19 section ‘Sample Size’. In this paragraph it is unclear how many subjects are already eligible for the LCS study, and so it is unclear whether it is feasible to establish an ongoing cohort of n=6000 within time period of 4 years of this study. Similarly, subjects will be (temporarily) lost when they are selected and participating in the PoC trial, and so it is conceivable that this group will be the largest to be re-sampling (possibly this is part of the rationale for over- under sampling, but not this is not clear from the manuscript text). It would be helpful to provide estimations of the number of subjects expected to be included to maintain n=6000 [and whether these are selected specifically to replace lost subjects; or whether subjects will not be selected and continue to be recruited according to inclusion criteria].

Response: The Participant Register for EPAD (PrePAD) currently includes 10 different cohorts with a total of 17500 participants aged >50 years and without dementia (described in more detail in Vermunt et al, Alzheimer’s and dementia, in press). New cohorts are continuously added.

The participant selection procedure for refilling the cohort is the same as for establishing the cohort (now described in Table 2). This is a very dynamic cohort, i.e. the composition of the probability continuum may change over time because participants (i) move into the PoC trial; (ii) drop out; or (iii) their characteristics (e.g. cognition, biomarkers, risk factors) change. Depending on the composition of the probability continuum at any given time point in LCS, participants coming in may or may not need to match participants moving out. Data monitoring and cohort balancing procedures will indicate which approach is needed at a specific point in time.

The reviewer is correct that the largest group moving out of LCS will most likely be the group going into the PoC trial.

minor:

-start and end dates of the project are missing

Response: we have added these. Recruitment started in May 2016, and the current IMI funding will end in December 2019. Our aim is to keep the cohort active afterwards as well.

-p7 “(Table 1) to ensure fast recruitment of a probability-spectrum population covering the entire continuum of probability for AD dementia development’ It is not evident that the use of parent cohorts will ensure fast recruitment, and/or that this will cover the entire continuum.

Response: We have rephrased as “The involvement of existing PCs and clinics where some data is already available on potential participants will facilitate fast recruitment. In addition, the variety of recruitment sources (from general populations to memory clinics) will provide a probability-spectrum population covering the entire continuum of probability for AD dementia development”. Table 1 has been removed and the entire section edited.

-p7 paragraph starting with “To ease the search process, “ a ‘flexible search algorithm’ is stated as a tool to select subjects for the LCS study, but, at this point it is unclear what inclusion criteria for subjects and so this part is difficult to understand. The authors may consider to move this part of the manuscript after subject inclusion/exclusion criteria are discussed, as well as providing a clearer explanation of the ‘flexible algorithm’ (is this like a flow chart?).

Response: We have moved this after inclusion/exclusion criteria and restructured the entire Methods section so that the flexible approach to selection is more clearly explained (including a new Table 2).

-p8 : the PrePAD velocity section is unclear, it seems that in this case only amyloid positive subjects will be eligible for LCS? It is also unclear why this approach would ‘optimise the balance in the LCS towards as large a proportion as possible to be amyloid positive by existing thresholds’ (this sentence seems to imply that not a balance, but a bias is the purpose. If an as ‘large a proportion as possible to be amyloid positive’ is important, then it would make more sense to use amyloid positivity as an inclusion criterium for LCS?).

Response: We have edited this section to improve clarity. The new Table 2 also explains how PrePAD Velocity is used. Amyloid positivity is not an inclusion criterium for LCS because we do not aim for an amyloid-positive cohort but one that covers the entire continuum of probability for subsequent dementia (including the entire continuum of amyloid levels). PrePAD Velocity is one way of ensuring that the cohort will not be overwhelmingly low-probability (including biomarker-negative).

-p10 “Population size will be maintained over time by continuous refilling from the PCs or via PrePAD Velocity” please clarify whether it is meant that all PCs will be continuously screened for eligible participants in an un-selective manner, or, whether subjects will be selected to match ‘lost/to be replaced’ participants in their subject characteristics.

Response: This is now explained in Table 2.

-p10 section “EPAD LCS participant selection process” the first part of this section discusses generically the need for disease modelling, this should be moved and more explicitly explained as part of the analysis plan, and this paragraph should focus on clearly explaining to the selection process. Here a working definition of the probability spectrum is necessary, as it reads now it as if estimated disease probabilities from to-be-developed models are already used it as a selection criterium at the same time.

Response: We have restructured and rephrased the text and included a new Table 2 to clarify the selection process.

-p16 Explanation of primary, secondary and exploratory outcome measures should be moved to the general section that gives an overview of all outcomes, as this explanation is not specific for cognitive outcomes.

Response: We have shortened this section to avoid an excessive increase in manuscript length. Given the link between LCS and the PoC trial and the shared primary outcome, this text is now focused on the cognitive outcomes taking into account the current regulatory perspective for AD trials.

-p17 Sudden change of future to past tense in Neuroimaging paragraph (this paragraph is difficult to follow, mostly due to the lack of data analysis plan. This section seems to imply that subjects will be compared on amyloid / APOE status, it is unclear what how this relates to the primary/secondary objectives of the present project).

Response: We have edited this paragraph to improve clarity. It does not refer to LCS data analysis but to the selection process for the neuroimaging assessments included in the LCS protocol.

-p19 last sentence: 'after 500, 1000, 2000' participants?

Response: Thank you, we have corrected this.

-p21 section 'informed consent' (and or p22 privacy of personal data; and or dissemination plan p23): the authors stated that data will be shared with 'researchers', please clarify whether 'researchers' indicates the global scientific community, or will this be restricted to EPAD consortium members; in the first case does the informed consent specifically mentions such wide data sharing?

Response: In the informed consent form participants are asked if they provide consent for data to be shared anonymously with other researchers.

EPAD has procedures in place for handling research proposals based on EPAD LCS data/samples coming from within or outside the EPAD Consortium (this is what we mean by "researchers everywhere"). All sharing of data/samples will of course have to follow the appropriate regulations and consider changes in such regulations over time.

We have shortened the Ethical issues section to address other comments regarding its excessive length.

- p25. "because different contributions from various components in each dimension may results in a similar overall probability, flexible algorithms are more suitable than simple cut-offs" this statement is inaccurate, as the use of cut-offs also allows to combine various components as part of a continuum (which in a more simple way is already done in research criteria when defining multiple disease stages (0, 1, 2) in preclinical/prodromal AD).

- p25 "... and for providing a sufficient number of potential trial participants" it is not evident why a probability continuum to estimate a subject's dementia risk necessarily results in 'sufficient number of potential trial participants'.

Response: We have edited the entire Discussion section to address several comments referring to it. This paragraph is now rephrased.

In the edited Statistical analysis section, we have now mentioned that many of the factors contributing to an individual's overall probability of subsequent dementia (cognition, AD biomarkers, various risk factors, and their changes over time) are continuous in nature. From a statistics perspective, treating them as continuous rather than dichotomizing or categorizing them may result in substantial gains in efficiency and avoidance of information loss when deciding where and why a participant falls in the overall probability continuum. Also, cut-offs can have an inherent arbitrariness that can be problematic, e.g. it's not that easy to draw an exact line separating preclinical from prodromal AD, or prodromal AD from dementia.

Table 1: Is this table necessary? Possibly, it would be more informative to add/merge this information to figure 1, as there is detail lacking in this table (e.g., the numbers of cohorts, references to the studies referred to), as well as the mixed presentation of types of cohorts with the cohort eligibility criteria.

Response: We have removed table 1 and included the information in the text after restructuring the Methods section. Figure 1 is meant to give a brief overview of EPAD LCS and EPAD in general. We have not listed specific parent cohorts because they join EPAD continuously, and we are open to considering any cohort that fits the criteria and has a PI willing to provide participants for EPAD. The

parent cohorts currently linked to EPAD are described in more detail in Vermunt et al, Alzheimer's and dementia, in press (reference 12).

Table 3:

Please provide in the manuscript text a rationale for under- and over sampling (to ensure high amyloid positive proportion ...?), as well as the corresponding proportions.

it is confusing to list factors that will be considered to be important for the development of the probability continuum in the same table as the selection process for participants.

Response: Table 3 has been removed and replaced with a new Table 2 including this information. Over/under-sampling is one of the tools we may use as part of the flexible approach to participant selection, depending on the structure of the LCS population and the needs of the PoC trial at each time point during the course of the study.

Reviewer: 2

Reviewer Name: Deborah Gustafson

Institution and Country: State University of New York - Downstate Medical Center

Please state any competing interests or state 'None declared': None declared

This paper describes the Protocol development and implementation for EPAD LCS, a cohort designed to characterize adults age 50 years and older in a rigorous manner, consistent with that for RCT enrollment. The goal is to identify candidates for secondary prevention trials of AD dementia, defined as AD neuropathology and clinical, symptomatic AD.

Some comments are as follows:

Page 4 – Strengths and Limitations – There are no limitations listed.

Response: We have indicated the main limitation, i.e. LCS will not be a traditional epidemiologically selected real-life population.

Secondary prevention is used too broadly in this manuscript. It is used to describe both preclinical biomarker only-based as well as clinical symptom-based intervention windows. This term should be defined and used consistently. Then, seemingly in contrast, at the end of the paper, Page 23, line 48, re: Dissemination, it is stated that EPAD LCS will help researchers, '... improve their understanding of the early, pre-dementia phase of AD, and facilitate collaborations.' Note: 'early, pre-dementia phase of AD'. Again, consistency regarding the exposure and outcomes windows needs to be enhanced.

Response: There is currently no generally accepted definition of 'secondary prevention' given ongoing discussions about when AD actually starts. Both broader biomarker-based and narrower symptom-based definitions are in use. We have chosen a broader definition for pragmatic reasons, ie this covers the entire intervention window targeted by current AD drug trials, from asymptomatic amyloid-positivity to prodromal AD. Our definition is stated in the Introduction as "evidence of AD pathology through relevant biomarker abnormalities, but without a clinical diagnosis of dementia".

We have checked that our definition is consistent throughout the manuscript, and removed the phrasing indicated by the reviewer as confusing.

Page 7, line 51 - Flexible search algorithm is not defined here or later in the manuscript.

Page 8, line 22 – without knowing the details of the flexible search algorithm, and noting Table 2, Inclusion and exclusion criteria, mention of amyloid positivity is a surprise. In addition, on page 10, line 41, there is mention of limiting selection bias by not overspecifying inclusion criteria. Some idea of the algorithm is finally provided in Table 3, but remains a black box.

Response: We have restructured and edited the entire Methods section to address this.

Table 2 – Eligibility criteria: it is unclear as to how the second bullet point is an eligibility criterion, rather it describes an aspect of protocol. In addition, the fifth bullet point is not only eligibility but reasons for the eligibility criterion. Reasons should be in the text, not in the table.

Response: We have edited the second bullet point and moved the definition of informant (5th bullet point) to the table legend.

Fulfill - check spelling

Response: Thank you, done.

Exclusion criteria - are there definitions for these criteria, e.g., uncontrolled diabetes or hypertension, 'severe' vision impairment, etc.

Response: These are defined according to the clinical judgement of each site investigator/study physician, as per local/national clinical praxis. The common denominator is "might make the subject's participation in an investigational trial unsafe", and "preventing cooperation or completion of the required assessments" (table 2). If a participant is later invited to an EPAD trial, there can of course be more specific definitions for these exclusion criteria, depending on the needs of each trial.

Page 10, line 15 and elsewhere, there is reference to 'continuous refilling of the cohort' to maintain N=6000. What does this mean?

Response: The number can decrease as participants move out of the cohort by going into the PoC trial or dropping out, hence the need to continuously refill the cohort. We have included a clarification in the text.

Page 10, last sentence of the first paragraph, beginning line 24 is very difficult to understand. Please clarify.

Response: Thank you, we have rephrased to make it clearer.

Page 12, Table 3 – it is interesting to note that only BMI is listed as a risk phenotype and not blood pressure or lipid levels or other risk factors for that matter, including control of vascular risk. Instead several of these are listed as Outcomes in Table 4. In addition, Secondary neuroimaging outcomes listed in Table 4 are only volumetric. There is no mention of amyloid except for CSF levels. This also raises the question of whether all enrolled in the EPAD LCS will be imaged and experience LP.

Response: Table 3 was indeed somewhat confusing, we have removed it as part of the extensive restructuring and editing of the Methods section. Vascular and lifestyle factors will indeed be assessed yearly and considered in the disease modelling work.

Outcomes listed in Table 4 (now table 3) include both secondary and exploratory MRI measures that go beyond volumetrics. The Data collection schedule (now table 4) specifies which assessments are planned at which visits, including yearly MRI and CSF. All sites will do the core MRI sequences, but not all sites will have the infrastructure for the advanced sequences, so these will only be done at a subset of the sites.

Page 16, line 14 – modifiable is a word and does not need to be in “ “

Page 16, lines 40-42 – be careful of grammar, consistent use of singular, plurals e.g., ‘... outcome measures may be used as... primary endpoint...’

Response: Thank you, done.

Page 17, lines 31-33 – again a mention of AD as defined by brain amyloid, yet no mention of amyloid imaging

Response: We have edited this paragraph to improve clarity. EPAD LCS does not include amyloid imaging (AD biomarker status is assessed using CSF). We are however planning to collaborate with the AMYPAD project focused on amyloid imaging, as mentioned in the Discussion.

Page 18, line 51 – it is stated that the only data source for this study is the EPAD LCS, yet there are some data coming from community-based PCs, yes? At least for recruitment purposes.

Response: There is no need for PCs to share their individual-level data with EPAD LCS. We are using a data discovery software tool that only allows us to see counts of subjects meeting certain criteria (e.g. age range, no dementia diagnosis, or other parameters available in each PC). Only the PI and team of each PC have access to individual-level data from their own cohorts. After obtaining the counts, a list of encrypted subject IDs is sent from EPAD LCS to each PC team, and they use their own decryption keys to identify exactly which subjects should be contacted for recruitment. This process is managed by the EPAD LCS Balancing Committee and Algorithm Running Committee.

Page 19, line 3 – It is stated that there will be central neuroimaging reads, yet no location provided, while central CSF and genetics laboratories are provided. In addition, will there be a centralized data entry, or by site?

Response: MRI images are uploaded from each site on a common platform. We have added the central reading location.

Page 20, first part of line 14, ‘As EPAD LCS is not a Clinical Trial of Investigational Medicinal Product (CTIMP), is not needed. Why even bring this up? Also, do you need to define an AE or SAE?’

Response: We have removed this from the manuscript.

General comments.

It appears that ADNI has a similar goal with somewhat similar approach, albeit narrower in participant recruitment strategies and sample size than EPAD LCS. There is no mention of ADNI in the Discussion. As an example, the 2016 Nature Commun paper by Iturria-Medina Y et al., should be cited. In addition, there are other consortia attempting to ascertain this space using only existing cohort data, e.g., IALSA/Maelstrom, STROKOG, etc. It may be worth mentioning this in the Discussion as well.

Response: We have included them in the Discussion.

This is a very long paper. This reviewer wonders if all of the detail toward the end of the paper re: Human Subjects, Privacy, etc. is necessary and/or this lengthy. It reads very much like a grant proposal.

Response: We have shortened the manuscript including these sections. Given the complexity of EPAD and the links between LCS and the PoC trial, we wanted to include at least some key points regarding ethical issues.

A list of EPAD LCS would be helpful.

Response: We are not sure what the reviewer means. EPAD partners? We have included a list in the Acknowledgements section. LCS sites? This is a very dynamic cohort, with a continuously increasing number of sites and countries. We have specified in the manuscript that the current status of the cohort and site locations are continuously updated on the EPAD website (<http://ep-ad.org/>). We are currently working on an updated version of the website with more detailed recruitment and cohort status information (this will be launched soon). Clinicaltrials.gov also shows detailed information about site locations and their status (<https://clinicaltrials.gov/ct2/show/NCT02804789?term=02804789&rank=1>).

Reviewer: 3

Reviewer Name: Stéphane Epelbaum

Institution and Country: University Hospital Pitié Salpêtrière, AP-HP, Paris, FRANCE

Please state any competing interests or state 'None declared': None declared

The manuscript by Alina Solomon et al describes the EPAD Longitudinal Cohort Study (LCS) protocol. EPAD LCS is one of the most important effort worldwide to tackle the complex problematic of preclinical Alzheimer's disease diagnosis and care. This means that this manuscript is certainly worthwhile and should be published. I am however concerned about a few points that I think must be improved in a revised draft:

1) The "flexible" recruitment procedure is not clear. Could the authors give some concrete examples of the balancing committee choices

Response: We have edited the Methods section extensively to clarify this, and examples are provided in the new Table 2.

2) No data is provided on the starting date of inclusions, nor on the expected date of the end for EPAD LCS (at least, the duration of the grant by IMI should be mentioned). As some centers have started recruiting for EPAD, their curves of inclusions should be provided. How many centers are actively recruiting at this point is also lacking.

Response: Recruitment started in May 2016, and the current IMI funding will end in December 2019. Our aim is to keep the cohort active afterwards as well. We have included this information in the manuscript.

This is a very dynamic cohort with a rapidly increasing number of sites and countries. We have specified in the manuscript that the current status of the cohort and site locations are continuously updated on the EPAD website (<http://ep-ad.org/>). We are working on an updated version of the website with more detailed recruitment and cohort status information that will better reflect the dynamic character of the study (this will be launched soon). Clinicaltrials.gov also shows detailed information about site locations and their status (<https://clinicaltrials.gov/ct2/show/NCT02804789?term=02804789&rank=1>).

3) considering merged protocols such as Aetionomy, which is in its ending phase, how do the investigators of EPAD see the sharing of data that have not yet been locked in the EPAD database to the collaborating investigators.

Response: Collaboration with other IMI projects, such as AETIONOMY, would have specific project agreements, which could allow sharing of data as described in the specific project agreement.

-considering the latest news on the EPAD website, the recruiting centres are "ten centres in six European countries". Which can hardly be considered as "pan-european"

Response: These numbers are continuously increasing as new sites and countries join EPAD. We hope that EPAD LCS will eventually come as close to "pan-European" as possible in the longer term.

-On what figures do the authors base their assumption that "A constant sample size of approximately 6,000 participants for the EPAD LCS is considered sufficient for a readiness cohort that should provide approximately 1,500 participants for the EPAD PoC trial." ? This should be backed by prior evidence of screening failure rates and referenced.

Response: Screening failures in previous trials have often been due to 'blinded' recruitment, i.e. individuals whose biomarker status was unknown prior to screening. As EPAD LCS includes comprehensive longitudinal assessments including AD biomarkers, this will facilitate a more targeted recruitment tailored to the specific trial profile, and thus minimal screening failures.

Also, no reference is made to other projects worldwide (eg Brain Health registry) to identify putative research participants in the preclinical/prodromal phases of AD. This is a serious flaw in the manuscript.

Response: We have included this reference in the Discussion section of the manuscript.

The endpoints of the study are defined but not the investigators hypothesis about them.

Response: Our starting point is that AD is a complex condition, and an individual's probability of developing dementia is most likely the result of multiple contributing factors. In EPAD LCS, participants may fall on a continuum of overall probability for subsequent dementia driven by several underlying dimensions: cognition; AD-related biomarkers; traditional risk factors (genetic and environmental); and their longitudinal changes. We also hypothesize that participants with similar overall probability may have different contributions from the various dimensions. Interrogating the underlying dimensions in addition to the overall predicted probability would be expected to facilitate participant stratification for identifying potential interventions, the size of a potential intervention effect, and directing participants to the most appropriate interventions. We have provided a more detailed description in the Disease modelling section.

Also, as mentioned in the EPAD LCS outcomes and other assessments section, an important factor driving the choice of outcomes in EPAD LCS was sensitivity to capturing the subtle changes that may occur during the very early stages of AD.

In summary, I think EPAD LCS is a great study that is an example of the collaborative strength of european expert centres. This manuscript however falls short in demonstrating it. I would gladly read a revised version of the manuscript when the few points above have been adressed.

Response: We would like to thank the reviewer for supporting our work. We hope that our revisions have addressed these issues and improved the manuscript.

VERSION 2 – REVIEW

REVIEWER	Betty Tijms Alzheimer Center and Department of Neurology, VUmc, Amsterdam, Neuroscience, Amsterdam, The Netherlands
REVIEW RETURNED	14-May-2018

GENERAL COMMENTS	<p>Some of my previous questions remain unclear:</p> <ul style="list-style-type: none"> - Probability spectrum: please provide a clear definition of low risk and high risk: Is this based age? amyloid status? combination of factors? The - cognitive tests as primary outcome: The authors state at p9 "EPAD LCS will provide a probability-spectrum population, i.e. where the entire continuum from low to high probability of subsequent dementia is represented at any time during the study." <p>It is unclear how the primary end-points align with this objective: A decrease on a neuropsychological test does not necessarily mean that an individual will develop dementia.</p> <p>Novel point: Since one of the major objectives is to model disease progression.</p>
-------------------------	---

REVIEWER	<p>Epelbaum Stéphane Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne Université, and Department of Neurology, Institute for Memory and Alzheimer's Disease, Pitié Salpêtrière Hospital, AP-HP, F-75013, Paris, France</p>
REVIEW RETURNED	30-Apr-2018

GENERAL COMMENTS	<p>The manuscript has improved since the original submission however some points remain problematic:</p> <ol style="list-style-type: none"> 1) In my opinion, the main problem concerns the sample size that EPAD aims to attain. On the dedicated EPAD website, it is written that 640 out of 6000 (just above 10%) of the participants have been included yet. This is troubling as two of the 4 and a half years financed by IMI have already passed. Even if the enrollment is, as the authors state, "very dynamic", I seriously doubt that the 6000 figure can be reached. Do the EPAD PIs have a contingency plan ? 2) In the response to reviewers, the authors state "the current IMI funding will end in December 2019. Our aim is to keep the cohort active afterwards as well". Once again, I fail to see how this can be achieved. ADNI has over 25 funding sources from academia and industrial partners alike and there are less than 2000 participants in all of ADNI. Can the authors elaborate on this crucial point and maybe refrain stating that EPAD is a "perpetual" cohort in their manuscript or at least write that the aim is to build a perpetual "ready for trial registry". What about EPAD Poc ? Eventhough it is clearly a manuscript on the LCS, the two parts of EPAD are very much related to one another. Is a funding secured for the LCS and the PoC after december 2019 ? 3) it is great that the authors "are working on an updated version of the website with more detailed recruitment and cohort status information that will better reflect the dynamic character of the study (this will be launched soon)." but it is really important to include a graphical representation of the inclusions center by center in the published manuscript. It is not the same if two centers have mainly contributed to the cohort as of now or if all of the centers contribute evenly.
-------------------------	--

	Thanks again to the authors for the upgraded version of the manuscript.
--	---

VERSION 2 – AUTHOR RESPONSE

Response to the reviewers' comments

*We would like to thank the reviewers for their comments and thorough assessment of our manuscript. Our point-by-point response is provided in *Italic font* below. We have also made changes in the manuscript to address the comments. All changes are tracked for easy identification and comparison with previous version.*

Reviewer: 3

Reviewer Name: Epelbaum Stéphane

Institution and Country: Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne Université, and Department of Neurology, Institute for Memory and Alzheimer's Disease, Pitié Salpêtrière Hospital, AP-HP, F-75013, Paris, France

Please state any competing interests or state 'None declared': None declared

The manuscript has improved since the original submission however some points remain problematic:

1) In my opinion, the main problem concerns the sample size that EPAD aims to attain. On the dedicated EPAD website, it is written that 640 out of 6000 (just above 10%) of the participants have been included yet. This is troubling as two of the 4 and a half years financed by IMI have already passed. Even if the enrollment is, as the authors state, "very dynamic", I seriously doubt that the 6000 figure can be reached. Do the EPAD PIs have a contingency plan?

Response: To achieve our objective of running a platform trial, we anticipate needing a readiness cohort of several thousand people, i.e. the number will be determined by trial-related needs. As of July 4, there are 809 participants. The current recruitment rate of about 100 participants/month is expected to increase as recently opened sites reach their full capacity, and new sites/countries also start recruiting. Thus, we estimate that a sufficient number of participants can be recruited even if N=6,000 is not reached before the end of 2019.

We have edited the manuscript to clarify this issue.

2) In the response to reviewers, the authors state "the current IMI funding will end in December 2019. Our aim is to keep the cohort active afterwards as well". Once again, I fail to see how this can be achieved. ADNI has over 25 funding sources from academia and industrial partners alike and there are less than 2000 participants in all of ADNI. Can the authors elaborate on this crucial point and maybe refrain stating that EPAD is a "perpetual" cohort in their manuscript or at least write that the aim is to build a perpetual "ready for trial registry". What about EPAD PoC? Eventhough it is clearly a manuscript on the LCS, the two parts of EPAD are very much related to one another. Is a funding secured for the LCS and the PoC after december 2019?

Response: We have removed "perpetual" from the manuscript.

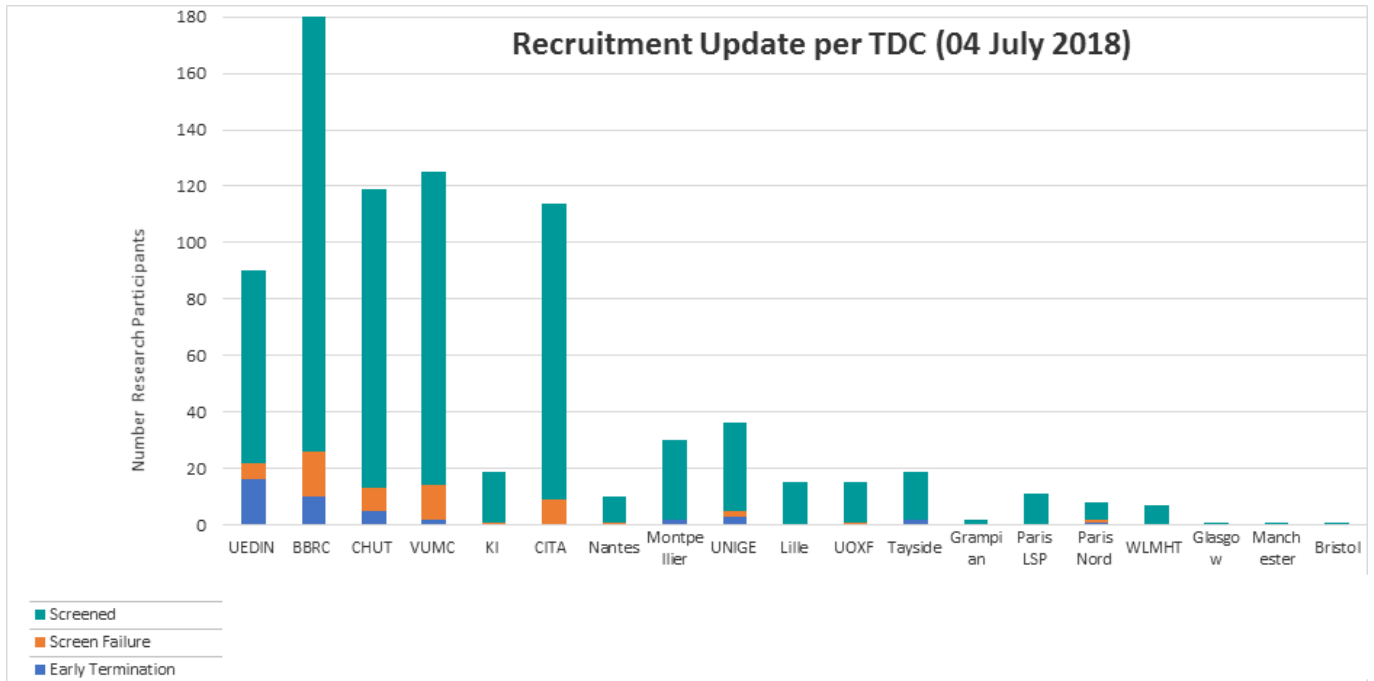
We are fully aware of the longer-term funding challenges. The EPAD Consortium is already working on a longer-term plan to make sure that the cohort and trial platform can remain active beyond December 2019. The primary focus is not on running the same protocol ad infinitum, but on innovative development of the cohort and trial platform over time. The very large Consortium with both academia and industry partners, and the dynamic, "open-ended" design that is fundamentally different from traditional cohorts will most likely increase the chances of securing longer-term funding. The openness to collaborations will also help.

3) it is great that the authors "are working on an updated version of the website with more detailed recruitment and cohort status information that will better reflect the dynamic character of the study (this

will be launched soon)." but it is really important to include a graphical representation of the inclusions center by center in the published manuscript. It is not the same if two centers have mainly contributed to the cohort as of now or if all of the centers contribute evenly.

Response: Contributions per site/country are not static but change over time. The rate of recruitment per site is usually somewhat lower when the site is first opened and increases gradually until the maximum capacity is reached. The maximum capacity will also vary between sites, depending on local logistics and size of parent cohort/clinic. Different sites are also opened at different times.

The figure below shows the current recruitment status (as of July 4) per site. We have included this in the manuscript as Figure 2.



Therapies for Alzheimer's disease Foundation (Spain); Nantes- Centre Hospitalier Universitaire de Nantes (France); Montpellier- Centre Hospitalier Universitaire de Montpellier, Gui de Chauliac (France); UNIGE- Geneva University Hospitals (Switzerland); Lille-Centre Hospitalier Régional Universitaire de Lille, Hôpital Roger Salengro (France); UOXF-University of Oxford (UK); Tayside- NHS Tayside, Dundee (UK); Grampian- NHS Grampian, Aberdeen (UK); Paris LSP- Hôpital Universitaire de la Pitié Salpêtrière (France); Paris Nord- Groupe Hospitalier Saint Louis - Lariboisière - Fernand Widal (France); WLMHT- West London Mental Health NHS Trust (UK); Glasgow- Glasgow Clinical Research Facility, NHS Greater Glasgow and Clyde (UK); Manchester- Greater Manchester Clinical Research Network (UK); Bristol- North Bristol NHS Trust (UK).

Reviewer: 1

Reviewer Name: Betty Tijms

Institution and Country: Alzheimer Center and Department of Neurology, VUmc, Amsterdam, Neuroscience, Amsterdam, The Netherlands

Please state any competing interests or state 'None declared': 'None declared': My affiliation (VUmc, Alzheimer center) is part of the EPAD consortium, I am personally not involved in this project.

Some of my previous questions remain unclear:

- Probability spectrum: please provide a clear definition of low risk and high risk: Is this based age? amyloid status? combination of factors? The

Response: As described in the manuscript section "Novel flexible approach to selection", we are deliberately moving away from the traditional approach based on a single definition with rigid cut-offs or pre-set categories. While we prefer combinations of factors to single factors, the exact combinations

may vary over time as more data accumulates and disease models are updated. Risk algorithms may also differ depending on their practical purpose, e.g. for LCS participants who are invited to participate in a trial, "high risk" may also be defined based on the specific characteristics of the drug to be tested, and the requirements of the trial.

Table 2 includes several examples of factors on which initial risk algorithms are based on. We have provided this at the beginning of the table to make it easier to find.

- cognitive tests as primary outcome: The authors state at p9 "EPAD LCS will provide a probability-spectrum population, i.e. where the entire continuum from low to high probability of subsequent dementia is represented at any time during the study. " It is unclear how the primary end-points align with this objective: A decrease on a neuropsychological test does not necessarily mean that an individual will develop dementia. Novel point: Since one of the major objectives is to model disease progression.

Response: While dementia is the most familiar outcome, modelling disease progression does not need to be restricted to dementia. As described in the "Disease modelling" section, we are also considering other meaningful intermediate disease states. Examples could be progression to preclinical AD, or progression to prodromal AD, or from CDR 0 to CDR 0.5 or 1, modelling trajectories of cognitive and functional decline in relation to various biomarker and risk factor trajectories and so on. Cognition is crucial to enable modelling across the entire disease continuum. It is especially important in the context of early treatments, when dementia may occur later, beyond the limited time frame of a clinical trial. We need to have disease models that also allow us to meaningfully assess treatment effects in the absence of dementia.