PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Aging Stereotypes and Prodromal Alzheimer's Disease (AGING): Study protocol for an ongoing randomised clinical study
AUTHORS	Gauthier, Kim; Morand, Alexandrine; Dutheil, Frederic; Alescio-Lautier, Béatrice; Boucraut, José; Clarys, David; Eustache, Francis; Girard, Nadine; Guedj, Eric; Mazerolle, Marie; Paccalin, Marc; de la Sayette, Vincent; Zaréa, Aline; Huguet, Pascal; Michel, Bernard; Desgranges, Béatrice; Régner, Isabelle

VERSION 1 – REVIEW

REVIEWER	Bin Cheng Columbia University USA
REVIEW RETURNED	16-Jul-2019

GENERAL COMMENTS	1. Page 5 of 35, line 23, N=290. This is inconsistent with N=260 claimed earlier. 2. There is no mentioning how the missing data issue will be handled. 3. Please provide more detail about the analysis of the primary endpoint.

REVIEWER	Jim Burke University of Michigan, US
REVIEW RETURNED	30-Jul-2019

GENERAL COMMENTS	Gauthier et al present a study protocol for a randomized controlled trial to explore a simple video-based intervention to reduce stereotype threat (ST) at the time of initial neuropsychological evaluation and to measure the effect of this intervention on a variety of outcomes. This is an interesting idea and it is valuable to pre-specify study such study protocols, particularly i the context of multiple measured outcomes. The theory underlying the study is well-described and study measures and interventions are described clear and in detail. However, there are several parts of the protocol that seem to be under-developed in this manuscript. Major Issues: 1. Statistical analysis is under-developed — This is the single greatest weakness of this paper and there are several important omissions here. First, the statistical analysis likely merits a section in the methods all to itself. What is the primary hypothesis to test? What is the outcome that will go with that hypothesis? How will that outcome be tested, statistically? What assumptions will be

	evaluated prior to that test and, if assumptions are violated, what alternate approaches might be used? Will the primary analysis be adjusted or unadjusted? If adjusted, how will adjustment variables be selected. Similar details should follow for all secondary outcomes. How will multiplicity be accounted for? How will the study be interpreted if a single (or multiple) secondary outcomes are "positive", but the primary outcome is not? 2. Power calculation requires more detail. What was the targeted effect size? How was that effect size identified? What is the primary outcome that will be measured? 3. Ethical issues should be expanded. In particular, what is the justification for exposing patients to the harms associated with CSF collection? How were those harms weighed against the potential benefits of the study? 4. CSF collection is likely to induce substantial selection bias amongst study participants? Can that bias be measured or mitigated? How likely is it that the selection bias induced may influence the estimates of the effect of the intervention? Is it plausible that factors that lead to selection into the study (e.g. traits of low anxiety, fear) might also lead to differences in the estimates of the treatment effect? Minor Issues 5. How will randomization be performed? Will randomization be stratified? Should it be? 6. How will blinding of the neuropsychological evaluation be insured? How will blinding of the assignment of diagnostic categories (e.g. aMCI vs. AD) be ensured? 7. The introduction is a nice summary on ST. As a non-expert on that literature, though, I wonder about the strength of the evidence, particularly in light of the recent awareness of suboptimal reliability throughout science, and particularly in experimental psychology. I'm somewhat skepticvla given some of the enormous effect sizes presented. Some comment on the overall strength of the evidence is warranted. Also, when presenting effect sizes from prior work, would report 95% CI so that readers can get a sense of the strength of the statistical evidence. 8. How will the inclusion criterion of "must report cognitive complaints but should not present any sign of dementia." be operationalized? This is a confusing criterion. 9. How feasible are the recruitment goals within the specified time window? They seem ambitious. How will the study (and study analysis) be altered if those goals are not met? 10. The authors might consider mentioning in the introduction that the implications of this study may be similar for patients diagnosed with "presymptomatic AD" (e.g. positive PET scan with no cognitive complaints) 11. There are several parts that may benefit from some language editing.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Bin Cheng
Institution and Country:
Columbia University
USA
Please state any competing interests or state 'None declared': None declared.

1. Page 5 of 35, line 23, N=290. This is inconsistent with N=260 claimed earlier.
Answer: The N=290 referred to the word count of the abstract. Sorry for that, we have removed this information from the ms.

2. There is no mentioning how the missing data issue will be handled.
Answer: We have planned to handle missing data with the "missing at random" hypothesis (multiple imputation, longitudinal mixed effects models) as recommended for missing data in clinical trials involving patients with potential neurodegenerative disease (Coley et al., 2011, Curr Alzheimer Res). This information was in fact provided at the end of the ms (in the data management and monitoring section). To make this information more visible, we have moved it to the new section entitled "statistical analyses" (page 18).

3. Please provide more detail about the analysis of the primary endpoint.
Answer: We completely agree with this important point. We intentionally summarized analytical strategy in the previous version of this ms to comply with words limits of the submission. We have now added a new section (pp. 18-19) that describes the statistical analyses used for testing the primary as well as secondary hypotheses.

Reviewer: 2
Reviewer Name: Jim Burke
Institution and Country: University of Michigan, US
Please state any competing interests or state 'None declared': None declared

Gauthier et al present a study protocol for a randomized controlled trial to explore a simple video-based intervention to reduce stereotype threat (ST) at the time of initial neuropsychological evaluation and to measure the effect of this intervention on a variety of outcomes.

This is an interesting idea and it is valuable to pre-specify study such study protocols, particularly i the context of multiple measured outcomes. The theory underlying the study is well-described and study measures and interventions are described clear and in detail. However, there are several parts of the protocol that seem to be under-developed in this manuscript.

Major Issues:
1. Statistical analysis is under-developed — This is the single greatest weakness of this paper and there are several important omissions here. First, the statistical analysis likely merits a section in the methods all to itself. What is the primary hypothesis to test? What is the outcome that will go with that hypothesis? How will that outcome be tested, statistically? What assumptions will be evaluated prior to that test and, if assumptions are violated, what alternate approaches might be used? Will the primary analysis be adjusted or unadjusted? If adjusted, how will adjustment variables be selected. Similar details should follow for all secondary outcomes. How will multiplicity be accounted for? How will the study be interpreted if a single (or multiple) secondary outcomes are "positive", but the primary outcome is not?

Answer: We completely agree with this important point. We intentionally summarized analytical strategy in the previous version of this ms to comply with words limits of the submission. We have now added a new section (pp. 18-19) that describes the statistical analyses used for testing the primary as well as secondary hypotheses. Additionally, we also clarified the primary versus secondary objectives on page 7.

2. Power calculation requires more detail. What was the targeted effect size? How was that effect size identified? What is the primary outcome that will be measured?
Answer: As now described at the end of the introduction (see also our answer to your point #7), Lamont et al. (2015) conducted a meta-analysis of age-based stereotype threat and found an effect size of d = .52 (95%CI [.248,.717]; corresponding to a f2 = .15), when using stereotype-based manipulations as we do here. However, because we are conducting the first study on age-based ST in the clinical setting, we decided to use a lower effect size (f2 = .07) to determine our target sample size in order to have sufficient power to accurately detect a probably smaller-sized effect in an ecologically valid field experiment compared with previous lab studies. Using this smaller effect size, the error rate set to 0.05, and the power set to 0.80, the power analysis indicated that a sample of 250 participants would be sufficient to detect the critical effects of the conditions and their potential interactions with several moderators with a multiple regression analysis (11 predictors: conditions, age, physiological stress, 4 vulnerability factors and the interaction terms between condition and these vulnerability factors). These details are now added in the ms (pp. 11-12).

3. Ethical issues should be expanded. In particular, what is the justification for exposing patients to the harms associated with CSF collection? How were those harms weighed against the potential benefits of the study?
Answer: The hospitals involved in the present study are using CSF biomarkers beta-amyloid and tau as a routine procedure for the high diagnostic accuracy of these biomarkers in detecting early or even prodromal AD (Ewers et al., 2012, Neurobiol Aging). Therefore, the CSF puncture is proposed as the routine procedure to all patients categorized as aMCI on the basis of their neuropsychological performances at visit 2 (whatever their experimental condition), and we ask for their agreement in the consent form to use their CSF data within the AGING protocol. In other words, patients are not exposed to the harms associated with CSF due to the AGING protocol. Rather, AGING is benefiting from CSF biomarkers data that are already available. We expect that the combination of neuropsychological performances and biomarkers of neurodegeneration will contribute to improve the accuracy of the diagnosis. We have clarified this point (pp. 15-16).

4. CSF collection is likely to induce substantial selection bias amongst study participants? Can that bias be measured or mitigated? How likely is it that the selection bias induced may influence the estimates of the effect of the intervention? Is it plausible that factors that lead to selection into the study (e.g. traits of low anxiety, fear) might also lead to differences in the estimates of the treatment effect?
Answer: As explained in our answer to the point above, CSF puncture will be proposed as a routine procedure only to patients categorized as aMCI on the basis of their neuropsychological performances at visit 2, whatever the experimental condition under which they took neuropsychological testing. Patients who will refuse CSF puncture will not be removed from the AGING protocol (CSF data will be simply missing data for these patients). There is therefore no specific reason that CSF collection would induce any selection bias.
We completely agree with the reviewer that factors like anxiety, fear, or other individual characteristics could make patients more or less susceptible to ST effects and thus to the experimental treatment. This is why we have included a questionnaire measuring several potential vulnerability factors (stereotypical perceptions of aging, memory complaints, anxiety about aging, subjective age) as well as measures of physiological stress during neuropsychological testing. These factors will be treated, respectively, as potential moderators and mediators of the treatment effect on cognitive

performances. These points are now made more salient in the ms (pp. 18-19), where we have specified the main and secondary objectives and related statistical analyses.

Minor Issues

5. How will randomization be performed? Will randomization be stratified? Should it be?
Answer: The randomization is computer generated based on an excel file (created by an independent coworker) and is performed by using permuted blocks (size=4). Within each hospital, randomization is made according to patients' age, sex, and socio-economic status in order to ensure balance between groups in size and patient characteristics that are of potential importance to investigate susceptibility to age-based ST effects. This information is added on page 10.

6. How will blinding of the neuropsychological evaluation be insured? How will blinding of the assignment of diagnostic categories (e.g. aMCI vs. AD) be ensured?
Answer: As already indicated in the previous version of this ms, the neuropsychologists in charge of tests administration (not the diagnosis) cannot be blinded to group due to the nature of the intervention (the use of a video to reduce stereotype threat in one of the two experimental conditions). However, following Gamerman et al. (2019)'s recommendations to minimize potential subjective bias, patients are blinded to the group and study hypotheses (p. 10), and neurologists, who are in charge of the diagnosis, are also blinded to the conditions (pp. 10, 16, 17). To ensure the blinding of the diagnosis, neuropsychologists who were in charge of the neuropsychological evaluation will not participate in the diagnosis decision-making (p. 10). Finally, the neuroimaging and CSF data will be analyzed by neurologists and neuroradiologists who are both blinded to experimental conditions. Analyses will be done by code (conditions labelled A versus B, and diagnostic categories coded with numbers), and the code will be revealed after all analyses are complete (Gamerman et al., 2019). This last point is now added in the ms in the new section "Statistical analyses".

7. The introduction is a nice summary on ST. As a non-expert on that literature, though, I wonder about the strength of the evidence, particularly in light of the recent awareness of suboptimal reliability throughout science, and particularly in experimental psychology. I'm somewhat skepticvla given some of the enormous effect sizes presented. Some comment on the overall strength of the evidence is warranted. Also, when presenting effect sizes from prior work, would report 95% CI so that readers can get a sense of the strength of the statistical evidence.
Answer: We agree with the Reviewer that cautious is warranted in the light of the crisis of replication. Lamont, Swift, and Abrams (2015, Psychol and Aging) published a meta-analysis of age-based stereotype threat and found a significant effect size on older adults' memory (d = .21, 95%CI [.020,.385]) and cognitive performance (d = .68, 95%CI [.399,.845]). Moreover, they found age-based ST effect to be stronger when the threat is subtly induced by the situation (d = .52, 95%CI [.248,.717]), as we do in the present study. This information is now added at the end of the introduction about age-based ST effects (page 6).

8. How will the inclusion criterion of "must report cognitive complaints but should not present any sign of dementia." be operationalized? This is a confusing criterion.
Answer: The present study focuses on the potential impact of negative aging stereotypes on MCI diagnosis, not on AD diagnosis whose accuracy is less debated. Thus, patients presenting signs of probable AD will not be enrolled in the study. The patients targeted by the study are SCI (cognitive complaints without cognitive decline) and/or MCI patients (cognitive complaints with cognitive decline but without AD dementia). The corresponding section in the ms is now clarified, with a clearer presentation of inclusion as well as non-inclusion criteria (page 9).

9. How feasible are the recruitment goals within the specified time window? They seem ambitious. How will the study (and study analysis) be altered if those goals are not met?

Answer: Strategies have been used since February 2019 to enhance patients' recruitment: communication on AGING protocol has been improved (e.g., poster displays in hospitals and fliers in general practices, call for participation to AGING on hospitals website), and appointment scheduling and reminders via email, mail, and/or telephone are used to retain enrolled patients in the trial. This is now added in the ms page 12. We agree with the reviewer that our sample size is ambitious. In fact, as noted in our answer to point #2 above, this sample size was aimed at guarantying enough statistical power for a field experiment.
In the case we do not achieve the initial number of inclusions, main analyses (test of the impact of instructions on diagnosis) will not be changed; only the number of moderators entered in the analyses will be reduced.

10. The authors might consider mentioning in the introduction that the implications of this study may be similar for patients diagnosed with "presymptomatic AD" (e.g. positive PET scan with no cognitive complaints).
Answer: Thank you for this suggestion. These patients will not be part of our study since we enroll only patients with memory complaints. However, we will take into account this suggestion in our future publication when discussing the findings of AGING.

11. There are several parts that may benefit from some language editing.
Answer: The English has been proof edited by a native speaker.

VERSION 2 – REVIEW

REVIEWER	Bin Cheng Columbia University
REVIEW RETURNED	30-Aug-2019

GENERAL COMMENTS	No further comments.