Author's response to reviews

Title: The overall impairment of core executive function components in patients with amnestic mild cognitive impairment: a cross-sectional study

Authors:

Dongming Zheng (zhengdm@sj-hospital.org)
Xiaoyu Dong (dongxy@sj-hospital.org)
Hongzan Sun (sunhongzan@sina.com)
Yongchuan Xu (xuyc@sj-hospital.org)
Ying Ma (mayingwfd@yahoo.com.cn)
Xiaoming Wang (wangxm1@sj-hospital.org)

Version: 2 Date: 24 October 2012

Author's response to reviews: see over
Dear Dr. Marwan Sabbagh:

Thank you very much for your attention and the referee’s comments and suggestions on our paper (MS: 1149589283690759). These comments and suggestions helped us a lot. We have revised the manuscript according to referee’s detailed suggestions. Here below are our responses to the referee’s comments. We sincerely hope this manuscript will be finally acceptable to be published on BMC Neurology. Thank you very much for all your help and looking forward to hearing from you soon.

Should you have any questions, please contact us without hesitate.

Sincerely yours,

Dongming Zheng, MD
Department of Neurology
China Medical University

Reviewer 1: Michael Malek-Ahmadi

Major Compulsory Revisions

Reviewer: Background: 2nd paragraph, line 15 “...are under the umbrella of EF” need to cite this information. Recommend using Chan et al (2008) Arch Clin Neuropsych

Answer: Thanks for the referee’s suggestion. We have cited this important article in our revised manuscript. (Page 4, line 21)

Reviewer: Statistical Analysis

1. The section on the z-score transformations of the executive tests is unclear. The authors state that because the measures were correlated that a direct comparison of the aMCI and NC groups could not be carried out. I disagree with this assertion. Although the correlations were statistically significant, the strength of these correlations was weak to moderate. This does not justify the rationale for not doing the aMCI vs NC comparison. In addition, these same executive measures are compared in a more specific sub-group analysis (Table 4) which also contradicts the rationale of not doing the group comparison. The subsequent table (Table 5) showing the paired z-score comparisons is not informative as it appears to simply pool the aMCI and NC individuals together, which contradicts the overall rationale of the study. There appears to be no methodological value in using the z-score analyses.

Answer: We apologize for the confusion and have rewritten the section for clarification (Page 12, Paragraph 2). The goal of this comparison is to examine whether the three core components of EF were impaired to the same extent. To ensure the comparability of the scores from different components of EF we used normalized Z-score of each executive test. Supposed we are interested in comparing two components. Using controls one can get the distribution of the difference in Z-scores for the two components. Both components may have been impaired in the aMCI patients but if the extent of
impairment was the same for both component (the null hypothesis) one should expect the differences in Z-scores in aMCI patients be similar to those in normal controls. Hence, one can perform a t-test to compare the differences in Z-scores in aMCI patients and those in normal controls. The added advantage of this approach is that by considering the difference between scores of two components correlation between components is no longer an issue.

An alternative way of this analysis is to first calculate the mean differences of Z-scores (D1, and D2) and their variances (V1 and V2) between cases and controls for two tasks and then compare the mean differences assuming independence (i.e., assuming the variance of D1-D2 is the sum of the variances of the mean V1+V2). In practice this alternative way of analysis does not make much difference and we got to the same conclusion. The drawback of this way of analysis is the independence assumption since scores for different tasks are actually correlated (Table 2). So in our paper we didn’t choose this way of analysis.

Reviewer: 2. When describing the binary logistic regression analysis, please state which group will be used as the reference and which group will be used as the outcome. A listing of the predictor variables would also be helpful.

Answer: According to reviewer’s suggestion, we have revised our description of the binary regression analysis (Page 12, Paragraph 3). The healthy control group was used as the reference and the aMCI group was used as the outcome. The five dependent measures from the 5 EF tasks were used as predictors.

Reviewer: Results and Tables

1. In general, all group comparisons for the t-tests should include measures of effect size. In this case, the authors will want to use Cohen’s d. Given the relatively small sample, reporting the magnitude of these effects will be provide a more meaningful interpretation that goes beyond simply reporting results that showed a statistically significant effect. Cohen’s d values can be calculated very easily using the means and standard deviations for the variables for each group. There is a very reliable Cohen’s d calculator which can be found at http://www.uccs.edu/~faculty/lbecker/. The Cohen’s d values can be reported next to the p-values in the tables (Tables 1, 3, 4.). For Table 1, the authors need only to report effect sizes for the neuropsychological tests. The interpretive scheme for Cohen’s d is: .00-.49 = small effect, .50-.79 = medium effect, #.80 = large effect.

Answer: First of all, we would like to express our sincere gratitude to the reviewer for the detailed guidance on how to calculate Cohen's d values. Based on the reviewer’s suggestion, we have calculated Cohen's d values and listed them next to p values in tables (See Tables 1, 3, 4).

Reviewer: 2. Table 5 – Significant differences between the respective groups are noted, but it is unclear what the p-values for these are. The last column of the table appears to report a p-value for the overall effect, but the specific groupwise p-values (NC vs. EF intact aMCI, NC vs. EF deficit aMCI, EF intact aMCI vs. EF deficit aMCI) do not appear to be reported. In addition, Cohen’s d values should also be reported for the respective group comparisons (NC vs. EF intact aMCI, NC vs. EF deficit aMCI, EF intact aMCI vs. EF deficit aMCI). Table 5 could be separated into 2 tables (NC vs. EF intact aMCI and NC vs. EF deficit aMCI) to facilitate the reporting of p-values and effect sizes for the group comparisons.
Answer: According to reviewer’s advice, we revised Table 4 to report the respective group comparisons (NC vs. EF intact aMCI, NC vs. EF deficit aMCI, EF intact aMCI vs. EF deficit aMCI). p-values and effect sizes for the group comparisons were all included in this table.

Reviewer: 3. In reporting the logistic regression results in Table 6, there are some serious errors. First, logistic regression analyses typically report odds ratios (OR) and not betas. Although SPSS does report betas for logistic regression, my experience is that odds ratios are also given in the SPSS output. Given that the authors are using the EF measures to classify aMCI and NC, odds ratios are more appropriate in this context. Regardless of whether odds ratios or betas are reported, they MUST be accompanied by 95% confidence intervals. If these are not automatically reported in the SPSS output, there is an option to have 95% CI’s reported, if my memory serves me correctly. In addition, the authors must indicate whether they are reporting unstandardized or standardized betas. This is usually indicated in the SPSS output.

Answer: Thanks to the reviewer for pointing out this error. Based on the reviewer’s feedback, we have reported OR values and their associated 95% confidence intervals for our logistic regression results. (Page 14, line 8)

Reviewer: 4. Page 12, 5th line: “…but the other three tests did not…” There are only two p-values reported after this statement.

Answer: We have corrected the error. (Page 13, line 10)

Reviewer: 5. ANOVA – Is there a rationale for using the LSD post-hoc test? Tukey HSD and Bonferroni corrections are more common. In truth, there is not much difference between the various post-hoc tests, but my preference is to use post-hoc correction methods that are consistent with similar studies.

Answer: According to reviewer’s suggestion we used Bonferroni corrections in the post-hoc test. (Page 13, line 19)

Reviewer 2: Christine Belden

Discretionary Revisions

Reviewer: Background: I appreciate that the ambiguous nature of EF and need for a more standardized definition and measures are addressed in this paper. The purpose of the paper (identifying a model of EF) is clear; however, the choice of “core” components is not as clear to me. In reading the background section, I found myself questioning why the Miyake model was chosen. The comment is made that EF is not a single-component brain function but only certain EF components were chosen for this study. This information regarding choice of components is eventually provided in the discussion section, but it would have been helpful in my understanding of the study methods/rationale if some discussion had occurred earlier in the paper.

Answer: We fully understand the reviewer’s opinion and have added a more detailed explanation in the background section, in accordance with the recommendations by the reviewer, on why the Miyake model was chosen (Page 5, Paragraph 2). Our purpose in choosing an executive function (EF) model to determine and detect the core components is to make selected EF core components more representative and more reasonable instead of using several subjectively designated executive function components as the research objects as some previous studies did. Among various EF models, there is
a category of EF models built on the basis of factor analyses. This category of model conducts specialized reasoning on which the combination of EF core components is a better representative of EF, therefore it is suitable for clinical studies. These models have their individual pros and cons, and it is difficult to claim that one model is the best. We have chosen the Miyake model for two main reasons: (1) the Miyake model is the most influential (the original paper proposing the Miyake model has been quoted more than 2400 times), there have been more clinical studies using this model, and its rationality has been validated by other research studies. (2) EF in this model only contains three core components; therefore, this model is more simple and feasible to be applied in clinical studies compared to those models containing more (four to seven) core components.

**Reviewer: Methods:** Inclusion/exclusion criteria are clearly defined. Assessments/procedures undergone are clearly defined. As the EF measures utilized were not all entirely familiar to me, I appreciate the thoughtful description of the EF tasks and procedures. Were vision and fine motor functioning taken into account? It is not clear whether all testing (cognitive evaluation and EF measures) occurred on the same day.

**Answer:** We accept the reviewer’s recommendation and have conducted a more detailed description of EF tasks and procedures (Page 7, Line 4; Page 8, last paragraph). All experimental subjects had normal vision or corrected normal vision. Cognitive evaluation and EF measures were not carried out on the same day, and we have fabricated supplemental descriptions of it. For neurological examination, none of the subjects had obvious fine motor disorders. Even so, taking into account that the elderly might have slow movement or might not be accustomed to using the computer keyboard or mouse to respond, we used a vocal key in the design of the Stroop task and the more-odd shifting task. The subjects only needed to speak into a microphone to record the reaction time, and no physical action was required. The design for the keep-track task did not require an action response. For the 2-back and stop-signal tasks, we also simplified the task design so that the subjects only needed to push one button on the response box, which was very simple and convenient. All of these designs helped to avoid experimental errors caused by elderly’s decline in fine motor ability.

**Reviewer:** It is not clear whether aMCI patient included “pure” aMCI subjects. Later in the paper mention is made that subjects are divided into EF intact and EF impaired groups, but it is not clear that difficulties in other cognitive areas (for instance, attention or processing speed) are not potentially affecting EF. The EF measures used to determine EF intact versus impaired appear minimal as only portions of the MoCA were utilized. Is it useful to divide the aMCI group based on limited EF information?

**Answer:** As indicated by the reviewer, the clinical assessment of executive function (EF) was relatively simple in this paper, and we have already mentioned this point in the “limitations” section in the paper. EF components for the Montreal Cognitive Assessment (MoCA) screening test, such as Alternating Trail Making, the Clock Drawing Test (CDT), Abstraction and Verbal fluency, are the most commonly used EF test items for clinical assessment. Some research studies have verified that the MoCA’s EF components showed a satisfactory power for discriminating EF in normal people. The purpose of subject grouping is to pick out aMCI patients with seemingly good EF based on routine clinical assessment to further examine their EF in detail. It is indeed insufficient to use only the above EF test items; thus, we have adopted the high standard of achieving perfect scores on these components as our
grouping criteria, by which we should be able to achieve the objective of screening out aMCI patients without obvious clinical EF defects. In our revised manuscript, we have added more explanations about this grouping (Page 7, Paragraph 2). We fully agree with the reviewer’s point related to the necessity of considering the impact of other cognitive functional difficulties on EF, as this is indeed an important direction of research in this field. Various cognitive functions have a very complex interaction mechanism, worthy of further thorough studies. This study does not yet cover this topic, which is another limitation of this study. We have added this limitation in the revised manuscript (Page 19, Line 13).

**Reviewer:** Results/Statistical analysis: Please see information provided by Michael Malek-Ahmadi.  
**Answer:** Please see our answers for reviewer 1.

**Reviewer:** Discussion: The discussion section is quite dense. As noted above, it would have been helpful for some of the information provided in the discussion section to have been introduced earlier to clarify study purpose and methods.  
**Answer:** As we stated in the previous explanations, we have made relevant modifications based on the reviewer’s comment.

**Reviewer:** In the first paragraph the comment is made that “all three components of EF declined significantly in aMCI patients.....” This gives the impression of change over time in this study, though subjects were assessed only once. It seems that stating that EF was significantly “impaired” and the degree of “impairment” did not differ significantly would be more accurate.  
**Answer:** We completely agree with reviewer’s suggestion and have modified the manuscript accordingly.

**Reviewer:** As noted above, it seems strong to state that “Patients who did not show executive dysfunction in clinical assessments exhibited significant deficits in computerized tasks” when minimal measures were administered to determine EF intact or impaired status. Ewers et al. are cited as reporting that “Trail Making Test B test, a commonly used method to assess task switching, was one of the best predictors of conversion from MCI to Alzheimer’s disease dementia.” I wonder then, why this measure was not included in the study (in it’s entirety rather than the abbreviated version on the MoCA) as this is a very commonly used clinical instrument.  
**Answer:** We wholly agree with the reviewer’s viewpoint that the Trail Making Test B should be used in its entirety, rather than the abbreviated version. However, the Trail Making Test B uses the English alphabet, and the majority of Chinese elderly are not familiar with the English alphabet. Therefore, the Trail Making Test B cannot be used directly. Because it is difficult to find twelve commonly used sequential Chinese characters, there is no available standardized Chinese version of the Trail Making Test B. Therefore, our study did not use the Trail Making Test B. The Alternating Trail Making of Chinese version of MoCA uses five commonly used Chinese words that follow a sequential order, and it shows an acceptable feasibility in clinical application.

**Reviewer:** From a clinical usefulness standpoint, it may be more helpful to have information on a measure that is very likely to be used, than a computerized measure that many may not be able to access easily. In your version of the Stroop task, were word-reading and color-naming trials
administered as well? Might a split between performance on these trials versus performance on the color/word task have revealed any useful information? I think the potential limitation of using computers in an aged population merits inclusion in the discussion. Though patient burden is always in important consideration, it is unclear why, in a study of EF, more comprehensive EF measures to classify subjects were not utilized.

**Answer:** Because some research studies have already shown a decrease in executive function in aMCI patients, the main focus of our study was not to prove this conclusion, but to evaluate the changes in every EF core component. The majority of EF tests used for clinical assessment cover a variety of EF components at the same time; therefore, we have used computer-programmed tasks extensively to evaluate single EF components. This method fits our research purposes, but it is indeed inconvenient for regular clinical use. There are many versions of the Stroop task, and most have used word-reading and color-naming trials as control conditions and 100% of incongruent trials in experimental conditions, as mentioned by the reviewer. The dependent measure for this type of Stroop task is the RT difference between the incongruent trials and the control trials. The Stroop task we used included a large proportion of congruent trials (75%) and a small proportion of incongruent trials (25%) in a mixed block, which can produce a larger Stroop effect on incongruent trials than pure blocks consisting of incongruent trials only. The dependent measure of this design is the mean RT to congruent stimuli subtracted from the mean RT to incongruent trials, and it is a sensitive indicator of the anti-distraction capability. Under the premise that this indicator can be calculated to further reduce the duration of testing time and decrease the patient’s burden, we did not add word-reading and color-naming trials. Belanger et al. conducted detailed studies related to the Stroop effect among MCI patients, AD patients and normal elderly people, and their results showed that there were no differences among these groups in terms of naming control speed[1]. Therefore we speculated that our conclusion would not be significantly altered by adding a naming control speed.

**Reviewer:** Conclusion: Very sparse section. Again, it seems that stating the EF components declined significantly is not entirely accurate as only one evaluation was conducted.

**Answer:** We have modified the conclusion section.

**Reviewer 3:** Alberto Costa

**Major Compulsory Revisions**

**Reviewer:** 1) More information about the characteristics of the aMCI sample (i.e., how many single domain and multiple domain subjects? Among multiple domain subjects, which cognitive domain plus memory was impaired?) should be given. Which is the rationale to include in the aMCI sample subjects who present with executive deficits on routine neuropsychological tests? Indeed, these subjects, that could be defined as aMCI multiple domain (plus executive deficits), are reasonably expected to perform worse than healthy controls on tasks tapping executive capacities. Rather, the most interesting findings refer to aMCI individuals who do not present with obvious executive deficits, but, nevertheless, show significant reduced executive functioning in respect to healthy controls. Moreover, were individual with depression excluded (did the authors administer some depression scale?)?

**Answer:** We completely agree with the reviewer’s comment that “Rather, the most interesting findings refer to aMCI individuals who do not present with obvious executive deficits, but, nevertheless, show
significant reduced executive functioning in respect to healthy controls.” This is one of the focused aspects of our research, but was not expressed clearly in the original text, and we have improved the description (Page 7, Last Paragraph and Table 4). In this paper, we used the full score in the EF tests of MoCA (Alternating Trail Making, the Clock Drawing Test (CDT), Abstraction and Verbal fluency) as the standard to identify “aMCI individuals who do not present with obvious executive deficits.” Even though this EF evaluation standard is simple, these test tasks are all commonly used tests for clinical assessment and have showed good discriminating power for executive function in normal people. Those aMCI patients, who did not lose any points on these tests, can be regarded as having no obvious executive dysfunctions.

In this study, we did not divide aMCI into single-domain aMCI and multiple-domain aMCI, but processed aMCI as a whole. As long as they met the diagnostic criteria of aMCI, they were accepted into the grouping, regardless of whether they did or did not clinically show a reduction in executive function or other non-memory cognitive functions. We did this based on the following considerations: (1) it would be ideal if we divide aMCI into a single-domain aMCI group, a multiple-domain aMCI EF normal group, and a multiple-domain aMCI EF abnormal group, but that would require a larger sample size. The sample size for the aMCI group in this study was relatively small, and it would be improper to divide them into three groups. Therefore, we chose to process aMCI as a whole, and certainly that is one of the limitations of this study. We agree with the reviewer’s point that “aMCI multiple domains (plus executive deficits) are reasonably expected to perform worse than healthy controls on tasks tapping executive capacities.” Due to the concern that this situation might impact the research results, we selected aMCI patients with no obvious dysfunctions to specifically study whether the EFs of this group of patients were significantly decreased compared to those of normal people. (2) Some research studies showed that patients who were clinically evaluated as single-domain aMCI could display cognitive dysfunctions in non-memory domains, when executing some non-conventional tests [3,4]. Single-domain and multiple-domain aMCI could likely be a continuum between normal aging and AD [5] and there is no absolute dividing line between them. Based on the above considerations, we did not divide aMCI into single-domain and multiple-domain aMCI, but treated aMCI as a whole, and evaluated its EF based on an EF model.

We used a 30-item Geriatric Depression Scale (GDS) to assess participants' depression, which was described in the text.

2) In the introduction, the discussion on the contrasting results reported in previous studies should be more explicative. Indeed, this is an important point to justify the present research and appears to be quite generic.

Answer: We have modified the introduction section based on the reviewer’s suggestion (Page 4, Paragraph 1).

3) More-odd shifting task. More explanations should be given for the way used to compute the switch cost. Indeed, it seems to me that in the "shifting" block both non-switch and switch trials are presented (is this right?). If it is so, why is the switch cost computed by using non-switch trials of a different block (i.e., the control one)?

Answer: Based on the reviewer’s suggestion, we added a more detailed description of the
more-odd shifting task (Page 10, Paragraph 2). The reviewer’s understanding is correct. The shifting block had both non-switch trials and switch trials. Non-switch trials in the shifting block usually have longer RT (response time) than the non-switch trials in the control block. That is mainly because non-switch trials in the shifting block could be disrupted by neighboring switch trials, so that they were not in the real non-switch state. Most studies in literature used non-switch trials in the control block to calculate the switch cost.

4) Statistical analysis. Given the high number of measures used in a relatively low sample, some measure should be adopted to avoid alpha inflation particularly for post-hoc comparisons.
Answer: We accepted reviewer’s suggestion and used Bonferroni correction to correct for multiple comparisons in the post-hoc comparison for five tasks between the aMCI group and the control group (Page 12, Line 1). \( P < 0.01(0.05/5) \) was used to make a claim about significance. The conclusion of this analysis was not changed after Bonferroni correction.

5) Z-scores. In calculating Z-scores, it is not clear which scores were used (“entire sample” at p. 11 line 8, includes MCI and healthy subjects?). More in general, I wonder whether the analysis on Z-score actually adds significant valuable information to the manuscript.
Answer: Yes, the entire samples include both cases and controls. The goal of calculating Z-score is to examine whether the three core components of EF were impaired to the same extent and the result was negative.

6) The discussion is somehow unfocused. It should be revised in order to make it more tight to data.
Answer: We have revised the discussion part to make it more focused on data. Some parts that are not closely related to our data have been deleted.

7) In the discussion, the authors often refer to the finding showing that aMCI subjects exhibit worse executive performance than healthy controls as if it would sustain the presence of a deficit. Indeed, this conclusion is not fully warranted as aMCI subjects’ performance is not compared with standardized performance of a normative population. In this regard, I would, thus, suggest to avoid the term “deficit”.
Answer: We agree with the reviewer’s suggestion to avoid using the term “deficit” to describe EF changes in aMCI patients. We have made the relevant modifications based on the reviewer’s suggestion.

Discretionary Revisions
1) In tables, the tests should be named at length to make the tables themselves more easy readable.
Answer: We have revised the tables and tests are all named at length in the revised manuscript.

References