Author's response to reviews

Title: The accuracy of the MMSE in detecting cognitive impairment when administered by general practitioners

Authors:

Patrizio Pezzotti (pezzotti@asplazio.it)
Silvia Scalmana (Scalmana@asplazio.it)
Antonio Mastromattei (mastromattei@asplazio.it)
Domenico Di Lallo (dilallo@asplazio.it)

Version: 2 Date: 25 January 2008

Author's response to reviews: see over
25 January 2008

Dr Lolu da-Silva
Senior Assistant Editor, BMC-series journals

RE: MS: 5841473815470627
The accuracy of the MMSE in detecting cognitive impairment when administered by general practitioners
Patrizio Pezzotti, Silvia Scalmana, Antonio Mastromattei and Domenico Di Lallo

Dear Dr. da-Silva,

The manuscript has been revised in accordance with the referees’ comments, as described in detail in the enclosure.

As per the check-list for re-submissions: there are no conflicts of interest and this was specified in the text; the informed consent was documented, a section about the author’s contribution was added; the references have been checked for accuracy and completeness and they were revised according to the requested standard format of the journal.
I, Dr. Pezzotti, will act as guarantor for the paper; and the manuscript has been checked by a native English speaker.

Thank you for your consideration.

Sincerely,

Patrizio Pezzotti
List of answers and changes made to the manuscript and responses to the referees’ comments

Reviewer: Hein van Hout

Only suspected previously undiagnosed patients with MMSE <25 were referred. Was there one UVA or more? If there were more do they function the same? What did they do to standardise the scores? In the Netherlands a consensus version of the MMSE was made including the interpretation of the scores. The total score can differ easily 3 or 4 points if the assessors use different interpretations. For example, if you are strict on the date, one day wrong is a fault (while in our consensus up to three days is considered a good answer).

There were four UVAs involved in our research (as we have already reported on line 4, 3rd paragraph on page 4) and they function the same. In particular, as we have already written in the methods section they followed the same standardized protocol for the clinical evaluation (see below)

“The clinical evaluation at the UVA was based on a standardized protocol and included cognitive and familial anamnestic, neurological examination, blood examination, neuroimaging (i.e., computed tomography or magnetic resonance), neuropsychological evaluation (i.e., MMSE, the Mental Deterioration Battery14, the Stroop Test27), functional evaluation [i.e., Activities of Daily living (ADL)28 and Instrumental Activities of Daily Living (IADL)29] and behavioural evaluation through the Neuropsychiatric Inventory (NPI).30 The diagnosis of any form of dementia was made according to DSM IV criteria31; the diagnosis of AD was made according to NINCS-ADRDA criteria;32 that of vascular dementia according to NINDS-AIREN criteria,33 and that of frontotemporal dementia according to the Lund and Manchester criteria.34 Mixed dementia was defined as the presence of both AD and vascular dementia. Mild cognitive impairment (MCI) was diagnosed according to Petersen et al. and Ritchie & Touchon criteria.35-38 “

Thus, we do not feel that text revisions are needed for this point.

For the MMSE scores both GPs and UVA used the same interpretation (for example, for what concern date, one day wrong is a fault). The GPs attended to a course on dementia with a specific training in the use of MMSE made by neuropsychologist of the four UVA’s, so the assessors used the same interpretation. We added a further specification about this point in the methods section.

UVA scorers were systematically higher. The authors mention this and find several associations but do not really explain this difference. Was this the case for all participating UVA’s?
We revised the discussion section to try to explain the reasons for higher UVA scorers compared to GP's. The UVA mean scorers were higher also for each UVA. This was added in the results section.

Strangely the difference between the standardised scores were much smaller, how do the authors explain this?

We have not understood this point because the difference between the standardized scores are exactly the same of the non-standardized scores (as expected). This is also reported at page 8 of the manuscript (“The mean and the median differences between paired scores (both crude and adjusted) were 1.58 (SD: 4.28) and 1 (IQR=-1; 4), respectively”).

No revision of the text for this comment was done.

There is likely to be verification bias:
The difference was (partly?) explained by ‘better patients’. Borderline scoring patients (23 or 24) according to the GPs have more chance to receive higher UVA scores than patients scoring far under the cut off, simply due to measurement error and daily variation. To check this I would suggest that the authors do a sensitivity analysis excluding patients scoring around the cut off, for example by taking the standard error of the mean of the GPs scores under 24 as extra limit. In fact the difference was largely explained by patients with MCI and patient with no impairment, which confirms this explanation.

We agree with the reviewer that borderline scoring patients (23 or 24) according to the GPs have more chance to receive higher UVA scores than patients scoring far under the cut off, simply due to measurement error and daily variation. The sensitivity analysis, as suggested by the reviewer, confirmed this issue. Discussion was revised to take into account of this point.

Remarkably the population had an average score of about 16. I assume that the diagnoses were seriously delayed otherwise I would not expect such low average scores. Can the authors comment on this?

We added a comment in the discussion

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
ROC score for UVA has no value as their MMSE assessment was used to make the clinical diagnosis. It was part of the assessment and interpretation, consequently it has a good ROC score. The same is true for table 2 with the accuracy measures, it may hold for the GPs but it is ridiculous for the UVA.
We disagree with the reviewer about this comment. The clinical diagnosis was essentially based on the complete clinical evaluation including cognitive and familial anamnesis, neurological examination, blood examination, neuroimaging, neuropsychological evaluation (the Mental Deterioration Battery, the Stroop Test), functional evaluation [i.e., Activities of Daily living (ADL), Instrumental Activities of Daily Living (IADL)] and behavioural evaluation through the Neuropsychiatric Inventory (NPI).

We feel that no revision of the text about this issue is needed.

In addition, we do not know whether the UVA was blinded for the GPs MMSE score, that would additionally compromise the accuracy measures on the GP. The ROC value for the GPs may make sense although there is the problem of the selected sample causing verification bias, and whether the GPs’ MMSE scores were blinded to the UVA team.

The GP’s MMSE score were not blinded to the UVA team, because patients were invited to go to UVA based on the MMSE score. However, it is unlikely that the UVA team were conditioned by the score given by the GP. For clarity of the reader, we added as a potential limit, in the discussion section, the fact that UVAs were not blinded to GP’s MMSE score.

Except for these limitations which should be discussed, for the Discussion section I miss a wider perspective in what the role of GPs can be in timely diagnostics.

Both in the introduction (page 3) and in the discussion section (page 12), we provided a perspective in what the role of GPs can be in timely diagnostics.

Also it would be interested to see a the problem of taboo being mentioned. I understand that, as in most countries (including Italy?) Alzheimer disease is still taboo for many patients, families and professionals. If you decide to screen and refer all suspected patients, as the authors did, it means that the GPs did not have trouble to test their patients and confront them with a referral to an Alzheimer unit. This is truly remarkable and the numbers of successful referral are indeed impressive as only 28 persons out of 397 were not tested by the GP, and only 15 out of 397 were not referred. Do the authors have any (anecdotal) information about this? See dfor further reading papers of H. van Hout or S. Iliffe

We revised the discussion to take into account of this point

Another point for the discussion is the timely diagnosis and ditto screening. In my opinion MMSE screening would makes sense especially for borderline persons, as for persons
below 18 it is clear for anybody that cognitive impairment is a serious problem. So it raises the question whether these persons were in fact detected timely.

*We added a specific comment in the discussion section*

**Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)**

Another discussion point is that the MMSE is a little long for busy general practices (at least in the Netherlands, Australia and the UK). The authors could comment on a variety of briefer tests (i.e. MIS, MISplus, MINICOG, GPCOG) which are probably more attractive for use general practice (see Brodaty 2006).

*We added a specific comment in the discussion section*

**Discretionary Revisions (which the author can choose to ignore)**

Instead of their figure 1 the authors may consider a bland-altman plot to picture the MMSE differences between GPs and UVA. Such a plot could additionally show whether the difference was systematic or in certain score ranges only.

*We feel that the reviewer has not correctly interpreted the figure 1, because it represents a Bland-Altman plot! To better clarify, we simply specified in the methods section (page 6) adding the specification that the figure is a Bland-Altman plot.*
Reviewer: Barton Palmer

Major Compulsory Revisions: None

Minor Essential Revisions:
It should be noted that the GP’s were selected from those willing to attend training on MMSE administration and all in fact attended that training. Such GP’s may be more attentive to standardized administration and scoring rules than the general population of GP’s, e.g., those who may administer the MMSE solely after reading the original Folstein et al. article and or the article describing the Italian translation. The possible effects of this potential sampling bias should be addressed in the Discussion section.

_We agree with the reviewer. In the revised version this limitation was addressed in the discussion section._

Discretionary Revisions:
There is a sentence beginning line 5, 3rd paragraph on page 3 regarding the effectiveness of existing pharmacological interventions that seems to overstate the empirical data on the utility/effectiveness of existing pharmacologic interventions in retarding disease progression in early AD. Although I agree with the authors’ general point that early diagnosis will be important in development of more effective interventions, the assertion of the effectiveness of existing medications should probably be worded a bit less emphatically.

_We agree with the reviewer. In the revised version the sentence was rewritten in a less emphatic way._

The GP administration of the MMSE always occurred first, i.e., two to six months prior to the administration by a neuropsychologist at the UVAs. Was time between assessments related to degree of discrepancy in the GP versus UVA scores? How much deterioration in MMSE scores would be expected over this time period? Also, is there any basis (in this study or prior reports) by which to estimate to what degree the slightly higher scores at the UVAs might reflect practice effects? These issues should be addressed in the Discussion section if possible.

_We did some other analyses to explore if time between assessments was related to the degree of discrepancy between the paired MMSE score. The median time between the two examinations was of 43 days and we revised the text at page 5 correcting “two to six months” in “within six months”. Thus, to evaluate if time between assessments was related to degree of discrepancy in the GP versus UVA scores, we evaluated the differences between paired MMSE using the quartiles of the days between the two MMSE administrations (i.e., ≤16, 17-42, 43-118, ≥119 days.) We found that there was not an association (the mean differences were –2.8, -1.0, -2.3, -0.2, respectively). This was_
added in the results section. We also expanded in the discussion section the paragraph about possible explanations regarding the difference between the paired MMSE score.
Reviewer: howard fillit

I believe there are some problems with the methods. While the authors state they are testing the accuracy of general practitioners to conduct an MMSE in practice, in fact, the GPs in this study underwent a training session. Thus the GPs in this study do not represent the general community. In addition, there is potential selection bias in the way that the patients were chosen for the study (page 5 second paragraph).

We agree with the reviewer’s issues. In the revised version these issues were addressed in the discussion section.

The GPs also used a "standardized questionnaire" (not identified) to collect data. It is not clear the issue posed by the reviewer. Did he refer to the fact that is a form and not a questionnaire? We revised the text substituting the word “questionnaire” with “form”.

Finally, the follow up visit at the UVA was 2 to 6 months later. Thus, patients may have significantly changed in the mental status in this interim.

We explored if the elapsed time between the two MMSE score is related to the degree of discrepancy between the two measurements. We created a categorized variable based on the quartiles of the days between the two MMSE administrations (i.e., ≤16, 17-42, 43-118, ≥119 days) We found that there was not an association (the mean differences were -2.8, -1.0, -2.3, -0.2, respectively). and this was reported in the results section. The problem of the time elapsed between the two measurements was also reported in the discussion as a possible explanation of this difference.

Are the data sound and well controlled?
Generally yes. It is a bit confusing that 26.5% of cases with suspected cognitive impairment were not confirmed by the UVA.

We do not have a specific answer to this consideration. We feel that this point can be left out.

Are the discussion and conclusions well balanced and adequately supported by the data?
Paragraph 2 in discussion, bottom page 10: the MMSE was not designed to detect patients with MCI.

We modified the following sentence: “However, there were significant differences when considering the diagnosis made by the UVA, specifically, for individuals with no cognitive impairment and those with MCI” in this way: “There were significant differences when considering the diagnosis made by the UVA, specifically, for individuals with no cognitive
impairment and those with MCI; however, it is of note that the MMSE was not designed to detect patients with MCI”

There is no discussion of WHY there were differences in the scores of the GPs and the UVA. What are the GPs doing right, or wrong? No speculation on the impact of the difference in timing, site of administration, person of administration, time of administration and other variables that might account for the findings.

We agree with the reviewer. Discussion was revised to take into account of this point.

Do the title and abstract accurately convey what has been found?
I might change the word "detecting" in the title to: "confirming"

We feel that the suggestion of the reviewer could be confusing and we confirm our title.