Author’s response to reviews

Title: The 10/66 Dementia Research Group's fully operationalised DSM IV dementia computerized diagnostic algorithm, compared with the 10/66 dementia algorithm and a clinician diagnosis: a population validation study.

Authors:

Martin J Prince (m.prince@iop.kcl.ac.uk)
Juan Llibre de Rodriguez (mguerra@infomed.sld.cu)
L Noriega (mguerra@infomed.sld.cu)
A Lopez (mguerra@infomed.sld.cu)
Daisy Acosta (daisyacosta@verizon.net.do)
Emiliano Albanese (e.albanese@iop.kcl.ac.uk)
Raul Arizaga (raularizaga@fibertel.com.ar)
John RM Copeland (jrmcop@btinternet.com)
Michael Dewey (m.dewey@iop.kcl.ac.uk)
Cleusa P Ferri (cleusa.ferri@iop.kcl.ac.uk)
Mariella Guerra (magasc@terra.com.pe)
Yueqin Huang (dengy@mail.tsinghua.edu.cn)
K S Jacob (ksjacob@cmdvellore.ac.in)
E S Krishnamoorthy (krish@neurokrish.com)
Paul McKeigue (paul.mckeigue@ucd.ie)
Renata Sousa (r.sousa@iop.kcl.ac.uk)
Robert J Stewart (spjurjs@iop.kcl.ac.uk)
Aquiles Salas (aquiles@cantv.net)
Ana Luisa Sosa (drasosa@hotmail.com)
Richard Uwakwe (ruwakwe2001@yahoo.com)

Version: 2 Date: 24 March 2008

Author’s response to reviews: see over
Thanks to both referees for their insightful comments. We hope that the revised version is now improved, with respect to description of methodology, identification of limitations, and discussion of findings. In the text below we have reprinted the reviewers comments, and inserted our responses in italics detailing the changes made.

Review 1

Minor Essential Revisions
A number of times throughout the manuscript the authors make reference to concepts that are not fully explained.

For example, in the measures section the authors talk about a weighted score from the cognitive scale of the CSI D and a discriminant function score from a combination of two other scores. What do the authors mean by these terms?
This is all detailed in the original CSI-D development/validation reference provided (Hall et al). ‘Discriminant function score’ is the term used by Hall et al to describe the summary score derived from the informant interview and cognitive test. We have now tried to clarify this further in the method section (page 9), and have provided the formula proposed by the instrument developers, and used by ourselves to calculate the DFSCORE.

The authors talk about the CERAD without defining this acronym.
The authors talk about the CERAD without defining this acronym. It is ‘Consortium to Establish a Registry for Alzheimer’s Disease’. It was in the list of abbreviations, but has now been clarified when first used in the text also (page 9).

The authors also talk about "indices of definition" when referring to the geriatric Mental State interview. What does this mean?
We had clarified this as ‘levels of psychopathology’. It is perhaps, though, an unhelpfully technical term, familiar only to those with knowledge of systems used in structured mental state interviews. We have therefore removed ‘indices of definition’ and replaced this term with a description of these levels – 0= non-case, 1 and 2 = subcase, 3, 4 and 5 = mild moderate and severe cases (page 10)

How exactly is this interview used to generate diagnoses of psychopathology.
The computerized clinical algorithm is extremely complex. General principles are outlined in the Copeland 1986 reference provided. We have also now provided an additional reference to our validation of the GMS/AGECAT algorithm in our earlier 10/66 pilot studies (page 10)

In the section that defines memory impairment the authors talk about a threshold of 1.5 standard deviations below the mean. What are they finding the mean of?
We have added the clarification that this is 1.5 SD below the mean for those with no dementia (page 12). As already explained, we used age and education-specific norms.

Discretionary Revisions
It is perhaps not surprising that the 10/66 dementia algorithm is more concordant with the clinician diagnoses given the 10/66 algorithm was initially derived based on correspondence with a clinician. This raises the issue of what is the gold standard against which to compare any new algorithm? Should the gold standard be the clinician diagnosis? It is obvious from this study (and the authors mention in themselves) that the clinicians did not fully operationalise the DSM-IV criteria for dementia when they were making their diagnosis. Therefore, it seems unfair to compare a strict fully operationalized DSM-IV definition against such a clinician. It will always come up short. When discussing the operationalization of the DSM-IV criteria the authors themselves say "we sought to operationalize the decision making process of a competent clinician". What does this say about their clinicians? Further discussion of the issue of appropriate gold standards is warranted.

We did not intend to suggest that any of the three outcomes, the 10/66 or DSM survey diagnoses, or the clinician diagnoses were definitive ‘gold standards’. In that respect, use of the term ‘criterion validation’ was perhaps unhelpful, and has been removed. DSM IV criteria have been critiqued for their lack of conformity to clinical consensus (see additional text on p22, and additional reference 16), so competent clinicians might well be correct in identifying cases of dementia beyond DSM IV. Even if the clinicians were meant to be applying DSM IV the lack of operationalisation of the DSM IV criteria may account in part for differences between the results from our computerized DSM IV algorithm and their judgment. Alternatively, we acknowledge that there could have been an issue with the reliability and/or validity of the clinician diagnoses. We wanted to try to understand more about the differences between the three diagnostic approaches, and how they might map on to an underlying construct of dementia – this is therefore closer to a construct validation in the absence of a gold standard as originally described by Cronbach, rather than a criterion validation. Hopefully the inserted text will help to get this point across.

A limitation of the methodology of the current study is that all diagnostic constructs were derived from the same pool of assessment data. Given this data was collected via clinician assessment there is a chance that method variance is reduced resulting in an increase in concordance amongst the different operationalizations of dementia. This would make the results seem more optimistic than they actually are. The manuscript would be enhanced with a discussion of these issues.

We have now acknowledged this important issue in new text on pp 21-22

Review 2
Major compulsory revisions
The methodology is far from clear. After a half day puzzeling with the data, I still don't understand major elements in the methodology. Important questions remain. Until these are not clarified, a proper evaluation is impossible.
- If I understand well, in the described group of patients, three elements are compared to each other: the clinical Cuban diagnosis, based on the 10/66
algorithm, DSM-IV computerized algorithm and 10/66 algorithm. Is here not a bias because you compare an algorithm with a clinical diagnosis based on the same algorithm?

As per the responses to the previous reviewer, we now acknowledge this issue in new text on pp 21-22

- Word fluency test: why taking a part of a validated instrument, and use it in a non validated way?

The animal naming verbal fluency test was first developed as an individual cognitive test in the early 1980’s (see ref provided). It was later used as part of the CERAD test battery (note, battery, not instrument), and is a standard component of multi-domain cognitive assessment in those with dementia, widely used in research and clinical practice. We have used it in the way that it was always intended to be used. Regarding validation, a link is now provided to the reference for the cross-cultural validation of this and other 10/66 cognitive tests in our earlier 10/66 pilot study.(page 8)

- Idem for the adapted Cerad: adapted how, based on what? Described where? Validated how?

The modification of the CERAD 10 word list learning task is described in full in the reference already provided (Ganguli et al). A clarification has been inserted in the text (page 9) – “The adapted CERAD ten word list learning task was found in the Indo-US Ballabgarh dementia study to improve the discrimination of the Hindi Mental State Examination. Six words were taken from the original CERAD battery English language list; butter, arm, letter, queen, ticket, and grass. Pole, shore, cabin, and engine were replaced with corner, stone, book and stick, which were deemed more culturally appropriate."

- Idem for the modification of the HAS.

The HAS was originally intended to provided information from an informant with relevance to the full range of mental disorders in older people, i.e depression, psychosis, alcohol use disorders etc, as well as dementia. We abbreviated it by omitting these sections and focusing on those directly relevant to the onset and course of dementia. This is implicit in the description of the modified HAS that we have already provided. However, the coverage of the HAS was insufficient to permit some dementia sub-types (e.g Lewy Body dementia and Frontotemporal dementia) that were first described and defined after the original HAS was developed. We therefore also updated it to make it consistent with modern nosology. However, since dementia subtypes are not the topic of this paper, we feel that this information need not be included in the paper.

- Description of the contente of the physical and neurological examination should be desirable.

We have now included a reference to our protocol paper (page 7), which contains full details of these and all other measures used in the survey). These examinations do not contribute to the diagnosis of dementia, only to that of dementia subtype, which are not the focus of this paper.

- The interview at the end of GMS, was it additionally or a part of GMS?
We make no mention of an interview at the end of the GMS, and we are not sure what this refers to?

- The 10/66 algorithm: is it described in the first section under the heading measures?
  Yes
- Operationalisation of DMS IV algorithm: I suppose this is the computerized algorithm?
  Correct. Now clarified at the bottom of page 11
- Criterion A 2: Informant opinion about decline: it has been proven that the informant opinion is far from reliable.
  We recognize the concern. However, with respect to dementia diagnosis, we demonstrated in our earlier 10/66 pilot study
  that informant opinion alone provided better discrimination than cognitive testing alone, and that inclusion of informant opinion in the 10/66 algorithm was the essential component for avoiding educational bias. Others have made similar observations, for example for the other two widely used informant-based screening measures for dementia, the IQCODE and the DECO.
- Did the clinician take all the information? In one session? What was the time needed for this session? If so, wasn’t there an experience of tiredness causing unreliable data? If it was done at several moments for each participant, what was the time delay between different moments? Was it then done by the same investigator?
  Yes, all data was gathered by a clinician, in one session. However, as some assessments were administered to the informant, and others to the older person, in practice fatigue was not an issue. Also cognitive tests were administered first to minimize this problem. The participant assessments took around 90 minutes, and the informant assessments between 20 minutes and just over an hour. More information is provided in the protocol paper, which has now been referenced.
- How many different physicians were involved?
  We have now clarified (page 8) that there was one per catchment area, i.e six in all for those six catchment areas where all interviews were carried out by experienced clinicians
  What was the basis of their expertise in dementia diagnosis?
  All polyclinics participate in the Cuban national program for older people. In the Cuban health system the routine diagnosis and management of dementia is devolved to specialists attached to polyclinics. The clinicians involved in the 10/66 study were all lead clinicians for the detection and management of dementia in their respective polyclinics. As explained they all had specialty training in psychiatry or geriatric medicine.
  What was the interrater reliability between the different physicians?
  In the training for the study co-rating of videotaped interviews was continued until adequate reliability was achieved. The Cuban PI Prof Rodriguez supervised initial interviews to ensure that ratings were correctly applied. Inter-rater reliability was not formally tested.
- Was the clinical diagnosis made at the same moment by the same person as the DSM IV computerized diagnosis?
  No. The clinical diagnosis was made at the end of each interview. The computerized 10/66 and DSMIV algorithms were run at the end of the survey period on the entire data file, by the London coordinating centre. The algorithms are relatively complex and were not provided to the interviewers. We can be reasonably confident that these two processes were as independent as they could be. A clarification on these important points has now been inserted in the design sub-section of the methods section (page 8).

- In table 6 NPI, CDR and Whodas are mentioned. Where in the methodology section is described that these instruments were performed? At what moment by who?
  These assessments, with the relevant references were described in the analysis sub-section of the methods section. This was because the measures section focused exclusively on those assessments that were incorporated into the diagnostic algorithms. We have now listed them under the measures subsection (pp10-11), which is probably more appropriate.

- In the results section it is mentioned that the group consisted of 1887 participants while later on the full sample has 2909 persons? Where are the 1022 persons left?
  This does seem reasonably clear in the text originally provided. Clinical diagnoses were obtained in the subset of 6 out of 8 catchment areas where the interviewer was an experienced dementia diagnostician. 1887 participants were interviewed in these 6 catchment areas. The remaining two catchment areas accounted for the other 1022 persons. We have now tried to clarify this further by mentioning the full 2909 sample and response rate (see below) before going on to describe the 1887 sub-sample from the 6 catchment areas.

Minor comments
- lack of numbering the pages
  Page numbering has now been provided
- Is there as to dementia an important difference between DSM-IV en IV-R?
  No. Only minor alterations to subtype classification.
- All eligible persons were included: no refusals? No dropouts? I nearly can not believe it.
  The overall response proportion was 96.4%. A note to this effect has now been provided at the beginning of the results section (p18).