PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	
AUTHORS	

VERSION 1 - REVIEW

REVIEWER	Dag Aarsland Karolinska Institutet, NVS, Alzheimer's Disease Research Centre
REVIEW RETURNED	21/10/2011

THE STUDY	Patients are selected from a multicentre trial population from various specialist centres in EU.
RESULTS & CONCLUSIONS	With a large number of centres and lack of inter-rater reliability scores or procedures to improve harmonization across centres, there is a high risk of measurment error which might influence the findings.
REPORTING & ETHICS	Attrition is not described; a flow.chart would be useful
GENERAL COMMENTS	The authors report additional clinical findings from a large Phase-III multicentre study of Dat scan as a diagnostic marker for probable DLB vs AD. There are few longitudinal studies of DLB and findings have been inconsistent. The authors compare changes in cognition (MMSE and CAMCOG), fluctuating cognition, and neuropsychiatric symptoms (NPI) in 58 DLB and 100 AD patients at the follow-up assessment 12 months after baseline. No significant differences in change were found. This is a clear report, and the original design is excellent. Strengths include the number of DLBs, the rigorous diagnostic procedures and the use of detailed cognitive measurement. There are however a some limitations which make the interpretation somewhat difficult: 1. As 40 centres were included, i.e. mean 4 patients per centre, there is a risk for measurement errors which might have disguised real differences. There is no report of inter-rater reliability, or whether attempts were made to harmonize procedures. 2. Since patients were recruited for a diagnostic trial, they may not be representative to the population of DLB patients (ie last sentence of first para of Discussion is not correct) 3. Furthermore, only those with complete follow-up data were included. We need to know the attrition rate and whether selective attrition might have influenced the findings. 4. As the authors acknowledge, the duration might have been too short to detect any difference in course between groups Minor issues: First para, Discussion; claims that there was “higher level of carer stress”; this is actually not quite correct; the NPI-carer distress score reports stress related to the specific symptoms, but not overall burden. Furthermore, with a higher NPI score in DLB, the carer distress score is expected to be higher in DLB than AD. A better strategy would be to have a general carer-burden scale, or alternatively, compare mean distress on items were both groups had

	positive scores. Methods, p8 (p9/22); 2nd para: “The consensus panel did not have access to scans”. This makes sense for the biomarker study, but seems a bit odd for the current objective/design. Next page: Last sentence before Statistics: 14 DLB and 3 AD were excluded since scan reading “was not in keeping with the consensus diagnosis”. It is not clear to me what this means. Is it related to the previous sentence above that DLB included only positive and AD only negative scans. A clarification would be helpful. The table is very informative. You might consider showing more clearly that Cornell and CDR were baseline values? P-values are given, but the tests should be given in footnote as well; assuming that they refer to the Statistics; with ANOVA for identifying groupxtime interaction? ADL was not measured It is appropriate to discuss a potential effect of antipsychotics. Similarly, given recent data suggesting antidepressants may also have negative effects in the elderly, an effect of antidepressant use on the clinical course is also relevant.

REVIEWER	Clive Ballard King's College London, Wolfson Centre for Age-Related Diseases Competing interest: Previous colleague of several of the authors. Currently co-hold a grant with the lead applicant
REVIEW RETURNED	28/10/2011

THE STUDY	The study is described by the authors as "large", but they do not describe the statistical power of the study. As this is driven by the smallest group (n=58) the power is probably modest. On the CAMCOG there is a 3 point numerically greater deterioration in the DLB than the AD group. A better understanding of the statistical power is needed to interpret these data - any limitations in statistical power need to be acknowledged.
RESULTS & CONCLUSIONS	The limitations of statistical power are again central to understanding and interpreting the data
GENERAL COMMENTS	This is an important study, which makes the diagnosis of DLB more rigorously than has been undertaken in previous longitudinalm studies. The paper is well written and the data are clearly presented. I do feel though that a comment regarding the statistical power and some discussion regarding whether limited power could be one interpretatiopn of the absence of significant differences between groups would be really helpful.

VERSION 1 – AUTHOR RESPONSE

Response to Reviewer 1: Dag Aarsland
Karolinska Institutet, NVS, Alzheimer's Disease Research Centre

1. Patients are selected from a multicentre trial population from various specialist centres in EU. With a large number of centres and lack of inter-rater reliability scores or procedures to improve harmonization across centres, there is a high risk of measurement error which might influence the findings.

All 40 centres were trained during 5 European workshops by ZW using identical training material and

exercises

2. Attrition is not described; a flowchart would be useful

A flow chart has been added detailing attrition

3. Since patients were recruited for a diagnostic trial, they may not be representative of the population of DLB patients (ie last sentence of first para of Discussion is not correct)

The sentence referred to has been edited out of the present draft.

4. Furthermore, only those with complete follow-up data were included. We need to know the attrition rate and whether selective attrition might have influenced the findings.

The characteristics of those lost to follow up have been included and discussed, as have the resultant limitations of the study

5. Minor issues: First para, Discussion; claims that there was “higher level of carer stress”; this is actually not quite correct; the NPI-carer distress score reports stress related to the specific symptoms, but not overall burden. Furthermore, with a higher NPI score in DLB, the carer distress score is expected to be higher in DLB than AD. A better strategy would be to have a general carer-burden scale, or alternatively, compare mean distress on items were both groups had positive scores.

With hindsight it would have been better to have an additional general carer-burden scale but only NPI-carer distress scores were recorded. Individual scores for each item on NPI were not recorded on the database only total scores. The reference to ‘higher level of carer distress’ has been clarified to refer to patients’ symptoms and not general carer stress.

6. Methods, p8 (p9/22); 2nd para: “The consensus panel did not have access to scans.” This makes sense for the biomarker study, but seems a bit odd for the current objective/design.

We feel that by the independent panel rating patients as probable DLB or AD purely on clinical grounds and the scan also being in-keeping with the diagnosis, this makes the diagnostic process even more robust than if the panel knew the result of the scan and was possibly influenced by the result of it.

7. Minor issue: Next page: Last sentence before Statistics: 14 DLB and 3 AD were excluded since scan reading “was not in keeping with the consensus diagnosis”. It is not clear to me what this means. Is it related to the previous sentence above that DLB included only positive and AD only negative scans. A clarification would be helpful.

Yes, only patients with a scan supporting clinical diagnosis were included. 14 DLB and 3 AD cases were excluded from the original cohort since the scan rating was not in keeping with the consensus diagnosis. We hope that the flow chart makes it clearer.

8. The table is very informative. You might consider showing more clearly that Cornell and CDR were baseline values? P-values are given, but the tests should be given in footnote as well; assuming that they refer to the Statistics; with ANOVA for identifying group x time interaction?

Table: Rows with scores for Cornell Scale for Depression in Dementia and Clinical Dementia Rating have been relabelled as ‘baseline.’ The statistical tests have been footnoted.

9. It is appropriate to discuss a potential effect of antipsychotics.

Significantly more patients in the DLB group were on neuroleptics but overall the numbers were small and if relevant one would expect to make the progression faster in DLB group. We have discussed this point in limitations.

10. Similarly, given recent data suggesting antidepressants may also have negative effects in the elderly, an effect of antidepressant use on the clinical course is also relevant.

At the time of setting-up the study the information regarding the negative effect of antidepressants was not yet published; as we were not aware of this association we did not record on the database antidepressants use. However, DLB patients had a higher depression score and we have included this in the limitations.

Reviewer 2: Clive Ballard, King's College London, Wolfson Centre for Age-Related Diseases Previous colleague of several of the authors. Currently co-hold a grant with the lead applicant

1. The study is described by the authors as "large", but they do not describe the statistical power of the study. As this is driven by the smallest group (n=58) the power is probably modest. On the CAMCOG there is a 3 point numerically greater deterioration in the DLB than the AD group. A better understanding of the statistical power is needed to interpret these data - any limitations in statistical power needs to be acknowledged.

The limitations of statistical power have now been explained in more detail.

2. This is an important study, which makes the diagnosis of DLB more rigorously than has been undertaken in previous longitudinal studies. The paper is well written and the data are clearly presented. I do feel though that a comment regarding the statistical power and some discussion regarding whether limited power could be one of the interpretation of the absence of significant differences between groups would be really helpful.

We are aware that this is an important point and we have therefore given numbers to illustrate how large the sample would have to be to be able to detect a statistical difference.

VERSION 2 – REVIEW

REVIEWER	Dag Aarsland Karolinska Institutet, NVS, Alzheimer's Disease Research Centre
REVIEW RETURNED	06/12/2011

RESULTS & CONCLUSIONS	I may have misunderstood, but with reference to the new statement re attrition (page 11), it seems that 25 DLBs did not come to follow-up. Are these 25 of the original 58 patients; or did you have 83 and 25 dropped-out giving 58 in total? This is a substantial drop-out; (if the first alternative applies, more than 40%). The drop-outs also differ from the completers. Thus, there is a clear risk of attrition bias. This needs to be highlighted more, both in the flow chart, the Results, and in the Discussion.
GENERAL COMMENTS	As stated in the last sentence, "other impairments may be more important and disabling". In this regard it is a limitation that ADL was not measures, which should be acknowledged.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: Dag Aarsland, professor, Karolinska Institutet, Sweden

I may have misunderstood, but with reference to the new statement re attrition (page 11), it seems that 25 DLBs did not come to follow-up. Are these 25 of the original 58 patients; or did you have 83 and 25 dropped-out giving 58 in total? This is a substantial drop-out; (if the first alternative applies, more than 40%). The drop-outs also differ from the completers. Thus, there is a clear risk of attrition bias. This needs to be highlighted more, both in the flow chart, the Results, and in the Discussion.

We had 58 DLB cases that fulfilled all the criteria as shown in flowchart. The 25 DLB cases that did not return were additional cases not included in the flowchart as only cases that returned for follow-up were considered for analysis. This was because the cohort was derived from the second and final consensus diagnosis, which patients lost to follow-up did not receive. Although the characteristics of patients lost to follow-up are important, their diagnoses were subject to change over the one year follow-up period, however this information was unavailable. We have amended our flowchart with a footnote, as well as added extra information to the 'Results' and 'Discussion' sections. We hope things are now clearer.

As stated in the last sentence, "other impairments may be more important and disabling". In this regard it is a limitation that ADL was not measures, which should be acknowledged.

We have acknowledged this point in our discussion.

VERSION 2 – REVIEW

REVIEWER	Dag Aarsland, MD Professor Karolinska Institutet, Sweden
REVIEW RETURNED	16/12/2011

GENERAL COMMENTS	I am pleased with these responses. I still think though that the added information re patient flow should be incorporated in the flow chart itself rather than as a footnote. Since this is a follow-up study. it is reasonable to include the base population in the chart, and then report proportion of completers and drop-outs.