

17th January 2020

Dear Professor Tampi,

Please find submitted the revised version of our paper. We have addressed each of the three reviewers' comments as outlined below in green font. This includes a shortening of the title to led "A comparison of two novel tools with traditional bedside cognitive tests in detecting delirium in older hospitalised medical patients". We believe that we have comprehensively addressed each of the comments and would like to thank the reviewers for their helpful comments and believe that the paper is significantly improved as a result.

We would be happy to further amend the report if that is deemed necessary.

Yours sincerely,

David Meagher

(on behalf of all the authors)

Reviewer #1:

- The title of the manuscript does not accurately reflect the main subject/hypothesis. Is the focus of the study to detect delirium or to identify comorbid delirium-dementia in elderly medical patients?

The Lighthouse and LSD were specifically designed with delirium detection in mind. The inclusion of 'comorbid delirium-dementia' in the title is a little distracting and was included because of the considerable interest in improving delirium diagnosis in subjects with pre-existing dementia, which is very poor. Therefore, we agree with the reviewer and have altered the title to the shorter "A comparison of novel tools with traditional cognitive tests in detecting delirium in elderly medical patients".

- In the introduction, it is important to mention that we already have a common diagnostic language for delirium that takes into account and recognizes patient diversity. This diagnostic language is based upon the Confusion Assessment Method (CAM) algorithm, which is the most widely used delirium assessment and consists of four features: 1) acute alteration/fluctuation from baseline mental status, 2) inattention, 3) disorganized thinking, and 4) acute altered level of consciousness.

The CAM algorithm has revolutionised delirium practice (both clinical and research) over recent decades. The CAM requires that cognitive impairment (inattention) is identified. This is best achieved using a standardised test. In recent years it has become apparent that, in

addition to attention, visuospatial function is disproportionately impaired in delirious patients and this is why we have focused upon attention AND visuospatial function as cognitive functions to test. Inevitably, single cognitive tests must be contextualised in terms of the broader clinical picture (e.g. to ascertain acuity of impairments) in order to determine whether they reflect a delirious process. The tests described herein can be integrated into CAM-based assessments and are not an alternative to an actual diagnostic process, which must include evaluating the context of symptoms. For this work, we used the DRS-R98 for delirium diagnosis because this allows for greater diagnostic specificity than the CAM, especially where there are substantial numbers of patients with comorbid dementia. Of note, the idea of these cognitive tests is not to replace robust diagnostic tools but to provide a user-friendly and delirium-accurate means of identifying the cognitive impairments that form part of a diagnosis of delirium.

To this end, we have addressed how these tools can be used in everyday practice – i.e. for efficient monitoring of cognitive function as well as how they relate to formal diagnosis, including using tools such as the CAM and DRS-R98. This has been added to the discussion just before the shortcomings section.

– Were the tests all administered in English? The majority of studies on tools for identifying delirium were conducted across a broad range of inpatient settings internationally in elderly inpatients, including patients with dementia but most excluded non-native language speakers.

Yes. Non-English speakers are uncommon in Ireland (although that is changing). We expect that the Lighthouse and LSD will be less subject to language-related inaccuracies than many other tests because they do not emphasise verbal skills, but this needs to be tested. We have added a comment on this to the discussion at the bottom of the penultimate paragraph before study limitations.

- Please provide the actual IRB study/approval number.

OK. This has been added to the text of the ethics section (i.e. REC 100/12).

- How was sample size determined? There is currently no evidence of power calculation. The present sample appears small and limited to a convenience sample.

A power calculation for the main research question (the ability of LSD-4 to detect delirium against no delirium, with $\alpha=0.05$, effect size=0.5, $df=3$, and sample size= 180 showed a power of nearly 1 (0.99) – see below. We have added this information to the text of the statistics section.

χ^2 tests - Goodness-of-fit tests: Contingency tables

Analysis: Post hoc: Compute achieved power

Input:	Effect size w	= 0.5
	α err prob	= 0.05
	Total sample size	= 180
	Df	= 3
Output:	Noncentrality parameter λ	= 45.0000000
	Critical χ^2	= 7.8147279
	Power (1- β err prob)	= 0.9999825

- Although there are an existing plethora of validated delirium screening tools, it is unclear which tool best suits particular populations, especially as this study utilized a rather undifferentiated population. This should be a proposed area for future work.

As suggested – we have added a piece to the discussion (at the end of the final paragraph before limitations): *“Psychometric data to guide the choice of test in particular settings is relatively lacking but ultimately the choice of cognitive testing tool is determined by a variety of factors that relate to patient, tester and other resource issues that are particular to the healthcare environment. Further work exploring the impact of these factors on the efficiency of providing cognitive-friendly healthcare is needed to guide choice of testing methods across settings”*.

- Suggest to replace Ref [14] with a published manuscript.

We are currently responding to reviewers’ comments regarding reference [14] and are hopeful that it will be accepted and progress to a full reference in the very near future. If needed we can substitute a much older and less detailed reference that relates to this issue (The performance of the Clock Drawing Test in elderly medical inpatients: does it have utility in the identification of delirium? Adamis D, Morrison C, Treloar A, Macdonald AJ, Martin FC. J Geriatr Psychiatry Neurol. 2005;18(3):129-33.) but our sense is that the currently cited reference will be formally published in time for the publication of this manuscript and is a much more useful reference for the reader.

- The underlying data should be made available to the readers (if this is not possible, please state why).

This is a difficult issue for us at present as the interpretation of EU GDPR regulations in Ireland is currently at the most restrictive end of the spectrum! We would need to revert to the ethics committee in order to do this and our sense is that we may not be granted approval despite the obvious value of providing data and the simple remedy of removing any potential identifying data. If it is felt that this is essential we would be happy to engage with the Ethics Committee but this is likely to cause substantial delays.

Reviewer #2:

In this article, the author attempted to apply a couple of new cognitive tests for detection of delirium. The study design seems to have been well constructed. Besides some limitations the author also mentioned, the conclusion looks acceptable.

I am concerned about the circumstance the cognitive test battery was performed, because most patients with delirium, as well as patients with Levy-body type dementia, alter their level of consciousness in a day. When did the participants take the test? Were the starting time same (e.g. ten o'clock, or early afternoon) among the participants? If so, the author should describe that. If not, are there any proof that each participants could perform their best effort to take the exam?

We have added a piece to the methods under 'Subjects and Design' as follows: *"Patients were assessed during the usual working day and in the majority of cases the process of receiving referrals and responding meant that this occurred in the early afternoon when the anchors of the day are thought to be optimally active"*

Where was the place the test performed, an examination room or the bed room of the participant?

We have added a piece to the methods under 'Subjects and Design' as follows: *"The assessments were conducted at the bedside to mimic real world practice"*.

Are there any effort performed to uniform the condition of the examination? These information will be helpful for other researchers to evaluate the result of this study later.

Yes. The methods section has been enhanced to provide a more detailed account of how we tried to optimise the level of standardisation in terms of timing and location of assessments as well as skillsets of the assessors and ordering of testing procedures etc.

Reviewer #3: 51749

A comparison of novel tools with traditional cognitive tests in detecting delirium and comorbid delirium-dementia in elderly medical patients, by Meagher et al., 2019.

This is a study to investigate the accuracy of bedside tests of attention, vigilance and visuospatial ability. For 180 consecutive elderly medical inpatients (around age 80; 51% female) referred to a psychiatry for later life consultation-liaison service, the final samples consisted of following: with delirium (n =44), dementia (n =30), comorbid delirium–dementia (n = 60) and cognitively intact controls (n = 46). Participants were assessed cross-sectionally with conventional bedside cognitive tests (WORLD, Months Backward, Spatial span, Vigilance A and B, CDT and Pentagons) and two novel cognitive tests (Lighthouse test, LSD-4). All testes showed high sensitivity (>70%). Authors suggested that these tests can distinguish neurocognitive disorders, including delirium, from other presentations. The Lighthouse Test and the LSD-4 are novel tests with high accuracy for detecting delirium.

This is a useful study investigating the performance of different bedside neurocognitive tests. The sensitivity and specificity of each tests were presented. I would ask the authors to

make clear which is the gold standard comparison for each test. This should be done across the text and the Tables.

OK. We have highlighted this in the methods section. *“Delirium was diagnosed according to a cut-off score of ≥ 15 on the severity scale of the DRS-R98 [18] and / or presence of DSM IV criteria [19] based upon a full clinical assessment. This approach was used for this work, because this allows for high diagnostic specificity in a population that includes substantial numbers of patients with comorbid dementia”.*

I would improve the section on statistics.

We have substantially reworded the statistics section as follows: *“Statistical analysis was conducted using SPSS-19 [28]. Continuous data are presenting as means plus standard deviation. Categorical data are presented as counts and percentages. . When multiple comparisons were conducted (ANOVA) the Bonferroni correction for multiple comparisons was used. The accuracy of tests of cognition (and their combinations), sensitivity and specificity as well as positive and negative likelihood ratio, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) were calculated, with confidence intervals testing significance at 95%. Post hoc power calculation for the main research question (the ability of LSD-4 to detect delirium against no delirium) was performed by using the G*Power v3.1.2. software. With $\alpha=0.05$, effect size=0.5 and $df=3$, a sample size of 180 indicated power of almost 1 (0.99).”*

The ROC curve analyses is essential for this type of study. I recommend plotting the ROC curve of all tests in a single additional Figure.

We have added this as a new figure - Figure 3. ROC analyses depicting the accuracy of conventional bedside tests of cognition for delirium diagnosis in the overall population (n=180)

On the limitation: currently, it is listed several limitations. I think that not all of them could have biased your findings. Please edit this and only declare those core limitations. Otherwise, readers will see a paper with so many errors that will not believe on your findings.

We have reduced the seven previously stated limitations to the four that appear most important and that can guide further research efforts.

Minor English double check is advised.

Yes, Thankyou – we have done this and tightened up the language throughout.