Reviewer’s report

Title: Validation of the Imperial College Surgical Assessment Device for spinal anesthesia.

Version: 0 Date: 08 May 2017

Reviewer: Klaus Hofmann-Kiefer

Reviewer's report:

General considerations

Topic: To test whether an Imperial College Surgical Assessment Device (ICSAD) could be a useful assessment tool to evaluate technical proficiency of differently skilled operators practising spinal anesthesia in a simulated model. Concurrent validity of the ICSAD was investigated by comparison to a validated task specific checklist and a global rating scale (GRS).

The authors examine an interesting and actual topic, especially because there is an ongoing trend to establish more objective measures of technical skill in physicians of all disciplines rather than relying on direct observation by an experienced observer and/or using self-reported procedure lists or logbooks.

Novelty: Unfortunately, a very similar study concerning the applicability of ICSAD to discriminate technical skills of various operators to perform an epidural anesthesia in pregnant women has been published in 2009 by Hayter MA et al. (1). Both studies included the same number of subjects, used the same methods and statistics and reported nearly identical results (see below). They also present some similarities in the discussion section.

Methods:

Methods provided in the current study are principally adequate and sufficient. However, there are far reaching similarities to the study conducted by Hayter et al.: In both cases study groups consisted of a novice group, an experienced resident group and an expert group of attending anesthesiologists. In both studies technical skills were evaluated by the ICSAD system and through video tapes which were judged by two independent, blinded observers. In both of the studies the reliability of the ICSAD was compared to a GRS and a task specific checklist. Unfortunately the appendix containing the detailed GRS and the checklist of the current study were not included in the PDF sent to the reviewer. However, at least the GRS seems to be the same in both of the studies.
In contrast to Hayter, Corvetto used a simulated torso to test the skills of their subjects. In the eyes of the reviewer this facilitates comparability and can be seen as an advantage of the current investigation.

Statistics:

Both studies use the same statistic methods which are - as far as I can say - appropriate. Unfortunately Corvetto et al. do not provide a statistical analysis concerning demographic differences between their study groups (Table 1). On first sight, it seems possible that "intermediates" and "experts" do not differ significantly with regard to the number of spinals performed in the last month/last 6 month. This may have contributed to the fact that the ICSAD was not able to differentiate between these groups.

Both studies have a major statistical flaw concerning power. However, unlike Hayter et. al, who simply stated that their study might have been underpowered concerning the most important outcome parameters (ability of the ICSAD to differentiate between groups), Corvetto et al. provide a "post hoc" power analysis. Despite the well-known fact that post hoc power analyses are questionable in general (2), the calculations behind this power analysis are unclear (lines 250-254) and primary outcome parameters are not defined. According to Corvetto, an effect size of 1.5 SDs (of which parameter??) led to a power of 88% (?). However, SD is questionable for calculating power when comparing not-normally distributed data, what is obviously the case in the current study (see line 155). At least the authors used non-parametric statistical tests (3). The authors claim "that there are no previous data available to calculate a sample size on "a priori" basis for this kind of studies but others have been done with 20-30 participants". However, in the eyes of the reviewer it shouldn't have been too difficult to define a proper outcome parameter characteristically for the ICSAD (for example path length), then to define a proper effect size and to perform a valid power analysis.

Results

Corvetto (as well as Hayter) found that the ICSAD was able to differentiate between novices and experts, but not between intermediates and experts. However, both authors do not blame this fact on the method itself but speculate about the possibility of too far reaching conformities between the latter groups. In the eyes of the reviewer it would have been more useful to simply stay with the facts: Obviously the ICSAD is not suitable to find differences between trained and expertly trained anaesthesiologists, but only between novices and experts. This result questions the method in general and should have been properly discussed. If reasonably worked up by Corvetto, this could have been a possibility to demarcate the current study from the very similar findings of Hayter et al.
Corvetto also compared the ICSAD results (only path length) to a validated GRS and obtained a correlation coefficient of -0.467. In contrast to Hayter's study the other modalities (duration, number of movements) of the ICSAD were not correlated to the GRS. This should be made up in a revised version and the results should be discussed, too. The rather moderate correlation between the ICSAD (path length, and if provided the other modalities), which does not find further mentioning, should also be a topic of discussion. GRS-testing and checklist-testing are useless, if the results of these tests are neither correlated to the ICSAD nor properly worked up.

Summary

The current study has an interesting topic, but unfortunately a very similar study already exists. The reviewer doubts whether it is useful to publish a further study dealing with the topic, especially considering the fact that epidural anesthesia (Hayter) usually is regarded as the more challenging technique. Corvetto and colleagues missed to correct some of the disadvantages of Hayter's study, especially concerning the number of subjects included, power calculation, definition of outcome variables and a less speculative but fact-based discussion.

In addition, I cannot completely agree with the author's conclusion. ICSAD failed to discriminate between novices and intermediates as well as between intermediates and experts and the authors observed only a moderate correlation to validated instruments of measurement (GRS), which in addition was carried out only in one of three ICSAD modalities. It is easy and does not need a special device to discriminate between a highly trained expert and a novice. For clinical purposes it would be far more important to discriminate between beginners and intermediates in order to decide whether the (ex) beginner has gained enough knowledge and skills to perform neuroaxial blockades without aid and supervision.

The current study can perhaps be interpreted as a pilot study. On the basis of its results, a useful sample calculation should be possible. Including an adequate number of subjects might definitely allow deciding whether ICSAD can serve as a reliable assessment tool or not. This and a more data-related discussion may be considered as a clear improvement compared to Hayter's investigation and perhaps will make publication worthwhile.

Details:

Line 51: this study aimed to …..(Change to simple past)

Line 90 and line 98 (and others): The positioning of the literature cross references is inconstant: Please set all of them at the end of a sentence (after the dot) or in between, according to the author's instructions of the journal
…study was to determine ….

this aim was only partially fulfilled (see General considerations)

Did the participants wear gloves? I think in some cases it should be easy to differentiate between the hands of a young unexperienced novice and an elder expert

The reviewer does not understand why the authors undertook these complex preparations concerning GRS, if the results were hardly put into relationship with the ICSAD results.

Why were the results displayed as median and interquartile range? Normally this is the case if data are not-normally distributed. Which tests were used to verify data distribution? The post hoc power calculation and its methods/variables should also be displayed in the methods section.

According to the guidelines of Landis and Koch a kappa coefficient of 0.76 is "substantial", not "almost perfect" (only values > 0.8)

In line 168 to 170 the authors describe the inter-reliability of the GRS scores, but than do not describe the results of the measurements but suddenly change to the results of the ICSAD?? (Line 171). That's disturbing. Where are the GRS results of the preparation phase displayed?

Why were the results displayed as median and interquartile range? Normally this is the case if data are not-normally distributed. Which tests were used to verify data distribution? The post hoc power calculation and its methods/variables should also be displayed in the methods section.

According to the guidelines of Landis and Koch a kappa coefficient of 0.76 is "substantial", not "almost perfect" (only values > 0.8)

In line 168 to 170 the authors describe the inter-reliability of the GRS scores, but than do not describe the results of the measurements but suddenly change to the results of the ICSAD?? (Line 171). That's disturbing. Where are the GRS results of the preparation phase displayed?

Why were the results displayed as median and interquartile range? Normally this is the case if data are not-normally distributed. Which tests were used to verify data distribution? The post hoc power calculation and its methods/variables should also be displayed in the methods section.

According to the guidelines of Landis and Koch a kappa coefficient of 0.76 is "substantial", not "almost perfect" (only values > 0.8)

In line 168 to 170 the authors describe the inter-reliability of the GRS scores, but than do not describe the results of the measurements but suddenly change to the results of the ICSAD?? (Line 171). That's disturbing. Where are the GRS results of the preparation phase displayed?

Why were the results displayed as median and interquartile range? Normally this is the case if data are not-normally distributed. Which tests were used to verify data distribution? The post hoc power calculation and its methods/variables should also be displayed in the methods section.

According to the guidelines of Landis and Koch a kappa coefficient of 0.76 is "substantial", not "almost perfect" (only values > 0.8)

In line 168 to 170 the authors describe the inter-reliability of the GRS scores, but than do not describe the results of the measurements but suddenly change to the results of the ICSAD?? (Line 171). That's disturbing. Where are the GRS results of the preparation phase displayed?

According to the guidelines of Landis and Koch a kappa coefficient of 0.76 is "substantial", not "almost perfect" (only values > 0.8)

In line 168 to 170 the authors describe the inter-reliability of the GRS scores, but than do not describe the results of the measurements but suddenly change to the results of the ICSAD?? (Line 171). That's disturbing. Where are the GRS results of the preparation phase displayed?

According to the guidelines of Landis and Koch a kappa coefficient of 0.76 is "substantial", not "almost perfect" (only values > 0.8)

In line 168 to 170 the authors describe the inter-reliability of the GRS scores, but than do not describe the results of the measurements but suddenly change to the results of the ICSAD?? (Line 171). That's disturbing. Where are the GRS results of the preparation phase displayed?

According to the guidelines of Landis and Koch a kappa coefficient of 0.76 is "substantial", not "almost perfect" (only values > 0.8)

In line 168 to 170 the authors describe the inter-reliability of the GRS scores, but than do not describe the results of the measurements but suddenly change to the results of the ICSAD?? (Line 171). That's disturbing. Where are the GRS results of the preparation phase displayed?

According to the guidelines of Landis and Koch a kappa coefficient of 0.76 is "substantial", not "almost perfect" (only values > 0.8)

In line 168 to 170 the authors describe the inter-reliability of the GRS scores, but than do not describe the results of the measurements but suddenly change to the results of the ICSAD?? (Line 171). That's disturbing. Where are the GRS results of the preparation phase displayed?
Line 269-277: Concerning conclusion: see general considerations


Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal