Reviewer’s report

Title: Evaluation of a multicomponent pathway to address inpatient delirium on a neurosciences ward

Version: 0 Date: 05 Jul 2017

Reviewer: Long H Ngo

Reviewer’s report:

The authors reported findings from a single-center observational cohort study of 1597 admissions from neurology and neurosurgery ward in the time period between 11/01/2012 and 03/31/2015, 798 of which came from patients admitted prior to the start of the intervention on 11/01/2012, and 797 came from patients admitted after the intervention. The intervention was a multi-component approach designed to prevent and treat delirium. Specifically, there were nine outcomes hypothesized to be reduced by the intervention: delirium prevalence (yes/no), delirium incidence (yes/no), duration of delirium (days), length of stay (days), use of restraint (yes/no), days restraint (days), use of sitter (yes/no), readmission (yes/no), and discharge to skilled nursing facility (yes/no). Delirium incidence was the primary outcome, and the others were secondary outcomes. Of these 9 outcomes, the authors found 2 (length of stay, readmission) that were statistically significantly (at type-I error of 0.05) reduced after the intervention in those patients with prevalent and incident delirium (with 420 admissions out of 1597). The remaining seven other outcomes were not statistically significantly different between the before and after-intervention group.

1) Thank you the authors and the editor for the opportunity to review this important paper. Following are a number of issues that the authors may want to clarify.

2) Nowhere in the paper did the authors mention the number of unique patients who contributed the 798 admissions before the intervention, and 797 after. If these came from unique patients, then the authors should indicate in the study design or analysis that the data were restricted to the index admission or a randomly selected admission among multiple admissions, and the reasons for such a restriction. If no such restrictions were used, then a patient could have more than one admission within the time period (before, after intervention), or across time periods, and the data from within a patient would not be independent, and should not be assumed so in all the analyses. In other words, if multiple admissions within the same person were used in the analysis, and the independence assumption was invoked (which is indicated by the analysis method that the authors used), then within-patient correlation information should be used; otherwise, the standard errors as currently reported could be underestimated, and the reported p-values are exaggerated (i.e. reported to be statistically significant but they may not be) assuming that the point estimates are correctly modeled. The point is the key finding of reduced length of stay may have a p-value that is not as small as currently reported (0.022, Table 3) if within-patient correlation is properly accounted for.
3) Another important point on the finding of the statistically significant reduction of length of stay in those with delirium (Table 3) is that what if there is also a reduction in length of stay for those without delirium? Note that in Table 3 which has all 1597 admissions, only 26% (420/1597) were from admissions of delirium patients; so it seems that there could be a reduction in those non-delirium admissions as well. On page 10, line 256, the authors mentioned about "there was not a similar degree of reduction in length of stay for non-delirious patients", but how much was the reduction seen in these patients if not of similar degree? How would the authors interpret such a reduction in length of stay in non-delirious patients who did not receive the intervention? It could be due to some secular trend (e.g. improvement in hospital care in general, or specifically in the neurology ward, or possibly early discharge as discussed by the authors?) but this improvement, or reduction in length of stay, however small, should be subtracted out from the overall difference detected in the length of stay of delirious patients. In other words, the analysis from table 3 should account for the possible "secular" effect from the "control group" (the non-delirious patients with no intervention). So in short, it is possible that the length of stay effect reported in table 3 was over-estimated (larger than it should be).

4) Another point of consideration in term of statistical testing, specifically to the length of stay finding is that the authors did not discuss as to why there were no adjustments considered for multiple testing. Of the nine outcomes being tested, the overall type-I error is inflated (so it is bigger than 0.05 and thus could improperly lead to more statistically significant findings) unless the individual outcome-specific type-I error is set much smaller than 0.05, for example, at 0.05/9 = 0.006, using Bonferroni adjustment. But if this adjusted outcome-specific type-I error is used to control for the overall type-I error, then none of the reported findings, including length of stay, would be statistically significantly different between the two time periods. The authors may want to discuss this point and defend the rationale for not using multiple testing adjustment.

5) The authors mentioned on page 2 line 32 that the patient sample was largely neurology and neurosurgery patients; however, on Table 1, the proportion of non-neurology admissions is almost 30% (472/1597). What is the reason for including this large number of non-neurology admissions in the analysis if the intention is to demonstrate the effectiveness of the delirium care pathway intervention in the neurology patient population? Did the authors carry out a sensitivity analysis to see if the intervention effect is heterogeneous between the neurology and the non-neurology patients? Given the fact that other studies have demonstrated the efficacy of multicomponent delirium intervention and treatment in general medical patients (authors' references #1 and #7), including this group of patients here would possibly risk bias away from the null for the neurology patients?

6) Delirium imposed on dementia (DSD) is known to be much worse than delirium on the non-dementia patient population. Was dementia an exclusion criterion here or there were patients with dementia in the sample? If so, that was the percentage, and if the authors attempted to look at the intervention effect in this group of patients?
7) The team of nurses and staff who conducted the CAM assessment, and the chart review, was the same team used for both the before and after intervention time period? If there was a substantial change in the team composition, how would that affect the findings?

8) The AWOL prediction score was used to stratify the patients; however, the AWOL score was developed and validated in hospitalized medical patients, and the 4 main predictors used for AWOL score derivation (>80 years, failure to spell "World" backward, disorientation to place, high nurse-rated illness severity) were specific to this population of medical patients. In neurology patients, the prediction model for delirium could be different with predictors possibly are related more to neurology conditions and patients; therefore, there could be misclassification issues here. The authors may want to discuss this point further and how this could affect the findings. Also, how is the threshold of 2 for AWOL chosen? Why not at 3?

9) Page 5, line 97-99, the authors mentioned about the intervention for those with AWOL score exceeding 2. Using evidence-based non-pharmacologic prevention measures, and others including re-orientation, regulating sleep-wake cycles, reducing restraints, are part of the NICE guideline recommendation for general medicine patients. For neurology patients, were there other measures used in the intervention that were specific to delirium patients?

10) The mean length of stay for delirious patients is about 5 days (Table 2), the nurses screened for delirium once every 12 hours using the CAM instrument (page 5, line 100-101). This means each patient has on the average up to 10 delirium assessments during their stay per visit. How is the overall definition of delirium prevalence and incidence defined? Is it at any time when positive for CAM? In the before-intervention time period, how was delirium incidence defined? Was the chart review documentation good enough even with adjudication?

11) The primary outcome in this study was delirium incidence. Thus the measurement of this variable is critical. The method used to define delirium incidence is chart review, which the authors stated, to have a sensitivity of 74% and specificity of 83%. However, these estimates were based on a study of general medicine elderly patients (author reference #11). In this study, the authors could actually assess how sensitive or specific the chart review method is in the after-intervention period, because both the CAM data, and the chart interview data were available. Thus using the CAM as "gold standard" it would be useful to see an informal diagnostic accuracy analysis of the chart review method. The estimates could be over-estimated due to the fact that the CAM results could not be blinded from the chart review.

12) Another issue regarding the quality of the chart review data for delirium incidence is, as correctly pointed out by the authors, may be affected by the fact that screening through the CAM might have "naturally increased the detection rate of delirium, leading to over-diagnosis" in the after-intervention period. What this problem implies is that the sensitivity for chart review for delirium incidence could be higher in the after-intervention period compared to the before-intervention period. This means that there
could be more false negatives in the before-intervention period, so more people with delirium and possibly having longer length of stay being excluded from the before-intervention group, which could lead to underestimation of the currently reported mean of 9.6 days. In other words, the authors could have underestimated the mean difference in length of stay between the two groups for delirious patients in Table 3.

13) On page 9, line 165-168, the power analysis was done based on the assumption that the 800 for each group consisted of independent observations. The author may want to replace the word "case" there with a more precise description of either "admission" or "patient," and clarify if there were repeated visits per patients, and if the same patients were allowed to appear in both time periods, before and after the intervention. If not, please add more text to clarify this section.

14) On page 9, line 179-181, the authors mentioned the use of multivariable modeling with adjustment covariates of age, sex, prior cognitive impairment, and Charlson comorbidity index; however, these variables were shown on Table 1 to be non-statistically significantly different between the two groups, so why was the need to adjust for them? If clinical rationale was the reason, please explain and give appropriate references. Notice that no such Table 1 was done for the subset of delirious patients only, so for the analysis of Table 3, we do not know if these 4 variables were significant. Please discuss if or why an analysis for confounder identification was not done for the results of delirious patients only from Table 3.

15) In the Results section, mainly Table 2, and Table 3, all the secondary outcomes go in the right direction. That is the intervention appears to be effective; although, not statistically significant other than length of stay and readmission. Are these findings clinically significant, or important, and could it be that for these secondary outcomes, the sample size is simply not large enough to detect statistically significant differences, since the study was powered only for delirium incidence? Notably, for example, a reduction of more than 2 fold for readmission (from 11% down to 5.4%), and a cut of more than 38% in the mean of delirium days (from 5.3 days to 3.5). These findings are remarkable for the intervention, and they do deserve some discussion. Were these expected and are these findings of this magnitude seen in other studies? And what could explain these large observed differences in this study?

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

Yes

**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.
Yes

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

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

Yes