PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Retrospective delirium ascertainment from case notes: a retrospective cohort study
AUTHORS	Geriatric Medicine Research Collaborative, ; Welch, Carly

VERSION 1 – REVIEW

REVIEWER	Valerie Page Watford General Hospital UK I was the PI on a multi-centre delirium project run by Geriatric Medicine Research collaborative in March 2018.
REVIEW RETURNED	13-Aug-2020

GENERAL COMMENTS	This is a solid project of data collection and review from a trainee-led collaborative. It is a multi-centre retrospective review of notes looking for documentation of delirium recognition, or symptoms of delirium by clinicians during acute hospital admission. Fundamentally it suffers from the key limitation common to all such studies, which is trying to diagnose delirium from notes when we know delirium is often difficult to detect even when you are looking for it. Given that delirium recognition requires a screening tool at the very least it is likely that a retrospective review of notes will miss significant numbers of patients having a delirious episode. The conclusions are necessarily limited. That said there are 'validated' methods of note screening used in the literature, usually as a supplement to screening rather than the primary method of detection. It would be helpful to know how the criteria for retrospective diagnosis of delirium were developed from the Kuhn paper. The Meagher project (ref 14) is not entirely correctly referenced. His group showed a change in arousal or alertness was not needed to fulfil criteria 1 rather inattention was sufficient. Please comment. Abstract The conclusions are more comments and while important points were not related to the results. Results As the data was collected in 27 different centres, which is to the groups credit, it does create additional confounders - what was the variability of patient numbers contributed by different hospitals, was there a difference in delirium diagnosis and screening between centres particularly if they had an active delirium prevention programme? It is a relatively small study so it is likely there were insufficient numbers to be able to demonstrate whether there is an effect of recognition of delirium upon adverse outcomes. I think that point needs to be made in the discussion.

	Why was mortality at one month used as an outcome? How many patients were still in hospital after a month? I think the point needs to be discussed earlier in the discussion that if a clinician uses the word confusion and/or agitation is that equivalent to missed delirium, or just a failure to use the right nomenclature? Is that what the use of record review is really testing? Consider whether the conclusion might be clinicians need to use the word delirium rather than the words confusion or agitation and how to action that? The conclusion as it currently stands is over stated. Minor points Abstract The primary and secondary outcome measures needs to be reworded. Is one month mortality the primary outcome? Or are they all primary outcomes? Introduction First sentence needs rewording - delirum can also be caused by drugs. Method Page 17 Line 53, consciousness is no longer a term used in the delirium diagnosis DSM-5. Page 22 line 29. What screening tools were used for delirium? The discussion first paragraph relates to the prevalence of delirium in this group rather than starting with the outcomes of the trial performed. A more balanced critique of the use of retrospective note reviewing as a means to determine prevalence of delirium use for research projects would be useful.

REVIEWER	Gen Shinozaki University of Iowa, USA
REVIEW RETURNED	02-Sep-2020

GENERAL COMMENTS	This manuscript from Geriatric Medicine Research Collaborative reported retrospective study investigating the impact of delirium recognition by comparing documented delirium cases versus additional cases identified by retrospective chart review by the study team. This is an important topic of interest for wide healthcare community, and study design is appropriate in general, but some limitation with sample size for the comparison for main aim between documented delirium versus unrecognized delirium, which might have limited the difference between the two as reported. Overall this (“the effect of recognition of delirium upon clinical outcomes”) is an important topic for investigation, and it is worth publishing with additional clarification and modification. First, as mentioned above, I wonder power analysis for the comparison between documented delirium versus unrecognized cases were done. If so, please add that in this report. If not, please do so. Some additional comments below. Cohort Identification section, multi-center, how many hospital, and how many are recruited from each would be worth describing. Also, I do not recall seeing analysis based on cite in the analysis.

	Is it based on admission cases only from one day on Sept 14th, 2018? If so, please make it clear that it is all from the same day across multiple hospitals. And I would like to know why it was decided to pick that specific day, and instead of multiple days, why only one day? What was the rational for excluding ICU? Other variable recorded section, local “trust”, I am sorry but I am not familiar with this term. Would that mean local “hospital”? Dementia status was based on documentation or “high clinical probability”. Could you add more detailed definition of “high clinical probability”? Statistical analysis section, for multivariable analysis, I did not see baseline disease severity and/or comorbidities, which can affect mortality and LOS. I would like to see those information adjusted, such as CCI. Patient and public involvement section, I am not clear if that would add value for this report. If you want to keep it, I noted that line 45, I wonder it is missing “how” between “about” and “long”. Discussion section, authors mentioned that lack of severity measure might have influenced their results, indicating that severer cases were more likely to be detected and thus documented, but less severe cases were more likely to be unrecognized and identified only with their study process. If that is true, I assume those with severer and detected delirium would have higher mortality, then those with less severe and unrecognized delirium would have less mortality, then we should see the difference. Those argument were not clear and should require more convincing description of the points to be made. Are you planning to indicated that severer and detected delirium receive prompt and appropriate care, and thus it ameliorate the risk of poor outcomes compared to those with unrecognized delirium who do not receive appropriate intervention, and then their outcomes are “balanced”? It was mentioned about their previous prospective study. Could you add a little more about the detail for comparison? Conclusion section, “clinicians identified delirium associated with worth prognosis; the most severe”. Could you reward this sentence? Not clear what you tried to say. One last major point. This study compared delirium patients both detected and unrecognized. But those are based on definition of delirium per DSM, which has very loose association with underline biology and pathophysiological mechanisms as authors pointed out the importance of better understanding of it. Recent works have started showing the importance of more biological basis approach to identify patients at high risk for poor outcomes including mortality. One is using EEG technology. Please refer to recent works (PMID: 31467255, PMID: 31483958) with that regard, and I would like to know authors perspective on it in

	discussion. I wonder our “chase” of delirium may not be enough to capture subjects with real high risk for poor outcomes including mortality, and some paradigm shifts with our approach may be needed. Ultimately, the goal of diagnosis is to use that information to tailor made our treatment and intervention to improve outcomes to better serve our patients. Thank you again for allowing me to review this important work.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

General comments: This is a solid project of data collection and review from a trainee-led collaborative. It is a multi-centre retrospective review of notes looking for documentation of delirium recognition, or symptoms of delirium by clinicians during acute hospital admission. Fundamentally it suffers from the key limitation common to all such studies, which is trying to diagnose delirium from notes when we know delirium is often difficult to detect even when you are looking for it. Given that delirium recognition requires a screening tool at the very least it is likely that a retrospective review of notes will miss significant numbers of patients having a delirious episode. The conclusions are necessarily limited. That said there are 'validated' methods of note screening used in the literature, usually as a supplement to screening rather than the primary method of detection.

1) It would be helpful to know how the criteria for retrospective diagnosis of delirium were developed from the Kuhn paper.

Thank you. We have amended our text to explain that the original validation study showed high sensitivity for prospectively identified delirium using case vignettes to diagnose probable delirium retrospectively. Our approach was previously piloted in a single site as part of another study, with excellent agreement with multiple data collectors.

Location: Methods, Retrospective delirium ascertainment

2) The Meagher project (ref 14) is not entirely correctly referenced. His group showed a change in arousal or alertness was not needed to fulfil criteria 1 rather inattention was sufficient. Please comment.

Thank you we have amended our text to clarify this. Change in awareness is not required as part of the relaxed DSM-5 definition. As inattention is more difficult to identify retrospectively if screening has not been performed, we used a relaxed definition requiring the presence of disturbances in either attention or awareness.

Location: Methods, Retrospective delirium ascertainment

3) Abstract: The conclusions are more comments and while important points were not related to the results.

Thank you. We have amended our conclusions section within our abstract to include points specifically relating our study.

Location: Abstract, Page 6

4) Results: As the data was collected in 27 different centres, which is to the groups credit, it does create additional confounders - what was the variability of patient numbers contributed by different hospitals, was there a difference in delirium diagnosis and screening between centres particularly if they had an active delirium prevention programme?

Thank you. We have included delirium prevalence and recognition rates separated by individual sites within our online supplement for transparency to the readers. Individual analysis by sites was not possible given the sample sizes at individual hospitals.

Location: Table S1, Additional file 1

5) It is a relatively small study so it is likely there were insufficient numbers to be able to demonstrate whether there is an effect of recognition of delirium upon adverse outcomes. I think that point needs to be made in the discussion.

Thank you. We agree that this is a very important point, and we agree that our study may have been underpowered. We encourage the development of further studies to assess if these results are duplicated in larger powered studies, and in other settings, and the incorporation of our results into future systematic reviews on this subject.

Location: Discussion

6) Why was mortality at one month used as an outcome? How many patients were still in hospital after a month?

Thank you very much for this comment. In our prospective data collection rounds, outcome data was collected up until 30 days as a pragmatic time cut-off for data collection. However, in this retrospective study we were able to collect mortality censored at the time of discharge. We have therefore, corrected our manuscript to ensure that we refer to inpatient mortality and repeated our analysis to ensure these are correct. We have rechecked all of our data, and have excluded a further 30 patients who had lengths of stay of less than two days, and we have proceeded to repeat all of our analysis for this reason. Length of stay ranged from 2 to 95 days; 34/626 (5.4%) were still inpatients beyond 30 days.

Location: Methods, Statistical analysis

7) I think the point needs to be discussed earlier in the discussion that if a clinician uses the word confusion and/or agitation is that equivalent to missed delirium, or just a failure to use the right nomenclature? Is that what the use of record review is really testing?

Thank you. We have restructured our discussion section so that this aspect is discussed earlier.

Location: Discussion

8) Consider whether the conclusion might be clinicians need to use the word delirium rather than the words confusion or agitation and how to action that? The conclusion as it currently stands is over stated.

Thank you. We have amended our conclusion and incorporated your recommendation within this.

Location: Conclusion

9) Abstract: The primary and secondary outcome measures need to be reworded. Is one-month mortality the primary outcome? Or are they all primary outcomes?

Thank you. We have amended this to state that 30-day mortality was our primary outcome.

Location: Abstract

10) Introduction: First sentence needs rewording - delirium can also be caused by drugs.

Thank you – we have reworded this to physical precipitants rather than physical illness, with examples.

Location: Introduction

11) Method: Page 17 Line 53, consciousness is no longer a term used in the delirium diagnosis DSM-5.

Thank you – we have reworded this.

Location: Methods, Retrospective delirium ascertainment

12) Page 22 line 29. What screening tools were used for delirium?

We did not collect details of the screening tools used. We have added this information to our methods for transparency.

Location: Methods, Retrospective delirium ascertainment

13) The discussion first paragraph relates to the prevalence of delirium in this group rather than starting with the outcomes of the trial performed.

Thank you. We have restructured our discussion section and moved this paragraph to later in the article.

Location: Discussion

14) A more balanced critique of the use of retrospective note reviewing as a means to determine prevalence of delirium use for research projects would be useful.

Thank you. We have incorporated this into our discussion.

Location: Discussion

Reviewer: 2

General comments: This manuscript from Geriatric Medicine Research Collaborative reported retrospective study investigating the impact of delirium recognition by comparing documented delirium cases versus additional cases identified by retrospective chart review by the study team. This is an important topic of interest for wide healthcare community, and study design is appropriate in general, but some limitation with sample size for the comparison for main aim between documented delirium versus unrecognized delirium, which might have limited the difference between the two as reported. Overall, this ("the effect of recognition of delirium upon clinical outcomes") is an important topic for investigation, and it is worth publishing with additional clarification and modification.

1) First, as mentioned above, I wonder power analysis for the comparison between documented delirium versus unrecognized cases were done. If so, please add that in this report. If not, please do so.

Thank you. We agree that this is an important point, and we agree that our study may have been underpowered. Our post-hoc power calculation is included within the discussion. However, it is important to consider that our results did not suggest any trend towards improved outcomes with recognition. In fact, if anything, there was a trend towards greater risk of adverse outcomes with recognised delirium.

Location: Discussion

2) Cohort identification section, multi-center, how many hospital, and how many are recruited from each would be worth describing. Also, I do not recall seeing analysis based on cite in the analysis.

3) Is it based on admission cases only from one day on Sept 14th, 2018? If so, please make it clear that it is all from the same day across multiple hospitals. And I would like to know why it was decided to pick that specific day, and instead of multiple days, why only one day?

Thank you. We have added further clarification about this. This project was performed as a sub-study within a larger quality improvement project; the date was chosen as it was six months before and after the dates of separate prospective data collection.

Location: Methods, Cohort identification

4) What was the rational for excluding ICU?

Delirium is known to be common in patients admitted to critical care but requires a separate screening process and our retrospective ascertainment has not been validated in this group. We have added this to our manuscript.

Location: Methods, Cohort identification

5) Other variable recorded section, local "trust", I am sorry but I am not familiar with this term. Would that mean local "hospital"?

Apologies, this is a UK-specific term. We have removed this and replaced with hospital as per your suggestion.

Location: Methods, Other variables recorded

6) Dementia status was based on documentation or "high clinical probability". Could you add more detailed definition of "high clinical probability"?

Data collectors made a clinical diagnosis of probable dementia if there was documentation of pre-existent cognitive decline affecting the patient's activities of daily living, but a formal diagnosis had not been made. We have added this to our manuscript

Location: Methods, Other variables recorded

7) Statistical analysis section, for multivariable analysis, I did not see baseline disease severity and/or comorbidities, which can affect mortality and LOS. I would like to see those information adjusted, such as CCI.

Unfortunately, this information was not collected and we are unable to adjust for it. We adjusted for the variables we collected – age, gender, dementia status, frailty, specialty, delirium subtype, and delirium duration.

Location: Methods, Statistical analysis

8) Patient and public involvement section, I am not clear if that would add value for this report. If you want to keep it, I noted that line 45, I wonder it is missing "how" between "about" and "long".

This section is necessary as part of the requirements for BMJ publications. We have amended the minor grammatical error – thank you.

Location: Methods, Patient and public involvement

9) Discussion section, authors mentioned that lack of severity measure might have influenced their results, indicating that severer cases were more likely to be detected and thus documented, but less severe cases were more likely to be unrecognized and identified only with their study process. If that is true, I assume those with severer and detected delirium would have higher mortality, then those with less severe and unrecognized delirium would

have less mortality, then we should see the difference. Those argument were not clear and should require more convincing description of the points to be made. Are you planning to indicated that severer and detected delirium receive prompt and appropriate care, and thus it ameliorate the risk of poor outcomes compared to those with unrecognized delirium who do not receive appropriate intervention, and then their outcomes are "balanced"?

Thank you, yes this is what we have suggested. We have further elaborated this section to add clarity.

Location: Discussion

10) It was mentioned about their previous prospective study. Could you add a little more about the detail for comparison?

We have added more detail about this study as requested. Our previous manuscript is available open access and fully cited for readers who wish to know more.

Location: Introduction

11) Conclusion section, "clinicians identified delirium associated with worth prognosis; the most severe". Could you reward this sentence? Not clear what you tried to say.

We have deleted this phrase from our conclusion section.

Location: Conclusion

12) This study compared delirium patients both detected and unrecognized. But those are based on definition of delirium per DSM, which has very loose association with underline biology and pathophysiological mechanisms as authors pointed out the importance of better understanding of it. Recent works have started showing the importance of more biological basis approach to identify patients at high risk for poor outcomes including mortality. One is using EEG technology. Please refer to recent works (PMID: 31467255, PMID: 31483958) with that regard, and I would like to know authors perspective on it in discussion. I wonder our "chase" of delirium may not be enough to capture subjects with real high risk for poor outcomes including mortality, and some paradigm shifts with our approach may be needed. Ultimately, the goal of diagnosis is to use that information to tailor made our treatment and intervention to improve outcomes to better serve our patients.

Thank you. We have added reference to this in our discussion.

Location: Discussion

VERSION 2 – REVIEW

REVIEWER	Gen Shinozaki University of Iowa USA
REVIEW RETURNED	06-Nov-2020

GENERAL COMMENTS	Thank you for the opportunity to review this nice work.