PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Validation of an automated delirium prediction model (DEMO Delirium Model): an observational study
AUTHORS	Mestres Gonzalvo, Carlota; de Wit, Hugo; van Oijen, Brigit; Debben, Debbie; Hurkens, Kim; Mulder, Wubbo; Janknegt, Rob; Schols, Jos; Verhey, Frans; Winkens, Bjorn; van der Kuy, Hugo

VERSION 1 – REVIEW

REVIEWER	Jin Han Vanderbilt University Medical Center USA
REVIEW RETURNED	23-Mar-2017

GENERAL COMMENTS	This study sought to validate DEMO to predict the development of delirium during hospitalization. An automated delirium prediction tool is an important endeavor. However, there are some critical limitations of your study: 1) Your gold standard for delirium threatens the validity of your study. The methods you used to determine the presence of delirium has little detail. It seems you relied upon the EPR, but it is well known that documentation of delirium is poor since the majority of delirium remains unrecognized by the clinical teams [1]. Did you also consider other forms of acute mental status such as “encephalopathy” since this term is interchangeably used with delirium. 2) Using your chart review method to ascertain delirium makes it difficult to determine when the delirium actually occurred. An episode of delirium that was documented at 48 hours could have actually occurred within 24 hours, but was not recognized by the clinical team until later. 3) Severity of illness is an important driver of delirium development, but is part of DEMO. Isn’t vital sign and lab abnormalities (e.g., APACHE) electronically available? Below are some specific comments: Page 3, lines 33 to 38, Introduction: There no universally accepted treatment for delirium after it has developed [2]. Though commonly used, there is little evidence that supports the use of antipsychotics in the treatment of delirium [3, 4]. The need for DEMO becomes more compelling in the absence of an effective delirium treatment; preventing delirium is more effective than treating delirium after it occurs.

	Page 4, lines 1 to 10, Introduction: Can you also provide the DEMO instrument in table format? In the table, make sure you place the cutoff used. Page 4, lines 33 to 39, Methods: Regarding the sample size calculations, I would mention that your sample size was based upon the requirement that the lower limit of your 95%CI would be 60%. Page 6, lines 38 to 39, Limitations: Another limitation is that this is a single center study located in the Netherlands and may not be generalizable in other settings. References 1. Collins N, Blanchard MR, Tookman A, Sampson EL: Detection of delirium in the acute hospital. Age Ageing 2010, 39(1):131-135. 2. AGS/NIA Delirium Conference Writing Group PC, Faculty: The American Geriatrics Society/National Institute on Aging Bedside-to-Bench Conference: Research Agenda on Delirium in Older Adults. Journal of the American Geriatrics Society 2015, 63(5):843-852. 3. Flaherty JH, Gonzales JP, Dong B: Antipsychotics in the treatment of delirium in older hospitalized adults: a systematic review. Journal of the American Geriatrics Society 2011, 59 Suppl 2:S269-276. 4. Neufeld KJ, Yue J, Robinson TN, Inouye SK, Needham DM: Antipsychotic Medication for Prevention and Treatment of Delirium in Hospitalized Adults: A Systematic Review and Meta-Analysis. Journal of the American Geriatrics Society 2016, 64(4):705-714.

REVIEWER	Vanja Douglas, MD University of California, San Francisco United Stated of America
REVIEW RETURNED	04-Apr-2017

GENERAL COMMENTS	The authors attempt to validate an electronic delirium prediction rule that uses age and medication information to predict delirium. The risk score (DEMO) is calculated automatically with electronic medical record data. The outcome of delirium for this study was assessed retrospectively by chart review. DEMO has good sensitivity and specificity, but the outcome assessment raises concerns. Overall, I think the authors should be commended for developing an electronic delirium prediction tool. The study has merits and would be a valuable contribution to the growing literature on delirium prediction. The problem, as is often the case with delirium studies, rests in the outcome ascertainment, which was done through chart review rather than by in person assessment. This is a reasonable approach, but it is incumbent on the authors to explain the approach more clearly, and to either use a validated method (or validate their own). Major comments: 1. In the introduction, the authors say "Treatment measures for delirium include pharmacological..." and also that "Haloperidol or atypical anti-psychotics are the standard pharmacological treatment for delirium...". This is not aligned with current evidence.

	A recent meta-analysis by Neufeld et al, "Antipsychotic Medication for Prevention and Treatment of Delirium in Hospitalized Adults: A Systematic Review and Meta-Analysis," J Am Geriatr Soc 2016, concluded "Current evidence does not support the use of antipsychotics for prevention or treatment of delirium." Antipsychotics should be reserved for sedation when patients present a danger to themselves or staff, and should be recognized as sedatives and not delirium-specific treatment. I worry this sentence in the introduction would be taken literally and used by providers to justify over-medication of delirious patients, which is a problem in many hospitals. Prior to publication this sentence should be revised to say something like: "There is no evidence to support pharmacologic treatment or prevention of delirium. Antipsychotics are often used for sedation when delirious patients pose a danger to themselves or to hospital staff. On the other hand, non-pharmacological symptom-targeted measures have been shown to prevent delirium in 30-40% of cases..." 2. Methods have a number of concerns: 2a. There is missing data from the case ascertainment. The manuscript says patients over 60 admitted to the hospital were included. Were ALL patients over 60 included? That is what the text suggests in some places, but later it says DEMO patients were randomly selected. The authors need to include a flow diagram showing total number of patients over age 60 admitted during the study period, number excluded due to delirium at admission, number screened by DEMO, and then the number excluded due to unclear diagnosis, and the final number included. The authors can look at the STROBE statement for standard examples of this flow diagram. 2b. This validation study does not appear to be truly prospective. To be prospective, both the prediction score AND the outcome should be determined prospectively. In this study, DEMO was calculated prospectively, but the outcome was ascertained by chart review retrospectively. This design should be more clearly stated in the methods. 2c. The measurement of the outcome has limitations. The diagnosis of delirium by chart review is notoriously difficult. We have performed similar studies using a chart-review based outcome of delirium and have been roundly criticized for it; while it offers the advantage of being more feasible, it truly does not meet the gold standard of in-person evaluations. However, I think it is a reasonable approach when resources are limited. That said, it needs to be done in a validated way, and it is not clear that the method used here is validated. In fact it's not really clear how the outcome was assessed. A search in the EPR for "delirium" and "delirious" suggests it was simply an electronic search? (As an aside, if this is the case, why didn't the authors include more patients?). If this is the case, how do we know patients where the chart said "this patient is at high risk for delirium" (and were not yet delirious) were not classified as delirious??? How were the patients who were excluded for being delirious at admission identified??? Or was the search a manual review of medical records? If the latter, simply excluding patients because "patient seems confused" is odd; generally that would indicate likely delirium. Inouye et al (PMID: 15673358) report a chart review method for identifying delirium where the sensitivity is 74% and specificity is 84%; if the authors performed a chart review, they should use a cited, validated methodology such as this.

	If they can't go back and do this, then they need to much more clearly explain how the diagnosis was ascertained using their methods. The quality of this paper really hinges on the identification of subjects and the outcome. 2d. Of note, the Inouye chart review method does not describe how to differentiate between prevalent (present on admission) and incident (developing after admisison) delirium. I am not aware of a validated chart review method for doing this. Therefore the authors need to be especially clear about how they did this. 3. This study begs for a validated, in person outcome assessment. In fact, a better design would have been to validate it in a cohort where the outcome of delirium was prospectively assessed in person, and the DEMO model was measured retrospectively (since it's a chart-based electronic method, all the data should be the same whether examined prospectively or retrospectively). I would change all of the conclusion to state this as the next step instead of saying DEMO is ready for real time use. Minor comments: Abstract: 1. Primary outcome measure is not sensitivity and specificity of DEMO - rather, it was development of delirium. The objective of the study was to measure the sensitivity and specificity of DEMO. But it would be more helpful here to explain that the outcome was delirium and how it was measured. Introduction: 1. How is polypharmacy defined in DEMO? Methods: 1. Page 5, line 22. What is DD delirium? What does DD stand for here? (differential diagnosis? please define) 2. Please explain what constitutes a positive DEMO screen vs. negative DEMO screen. It appears the DEMO is either positive or negative based on the text available and based on the way the tables are constructed. However, in other places it refers to DEMO predicting a specific % risk, and then being cut into high/low risk groups based on 14.1% being the cut-point. Why was 14.1% chosen? 3. Was consent obtained or waived? Results: 1. Table 3: does not need a separate row for man and woman (the numbers for one can be easily calculated from the other). Discussion: 1. First sentence is not technically true (not prospective) - see comments above. 2. Second paragraph - problem with outcome ascertainment is discussed in detail above. 3. Third paragraph - needs to be greatly expanded given the limitation of outcome ascertainment.

	4. Fourth paragraph - I would emphasize that DEMO uses a clever way of identifying cognitive impairment by including medications used for dementia. For example, donepezil is unlikely to be prescribed except in a patient with cognitive impairment - this is a nice way of getting at this risk factor electronically. 5. Another strength of the DEMO model is that it predicts delirium on a daily basis. This is a novel concept (most delirium prediction rules work at admission but not daily). It is not clear to me whether there is a definite advantage to predicting delirium on a daily basis (it could end up being information overload for nurses and physicians) but it also could eventually be something that is tracked along with the vital signs and intake/output. Overall: 1. English grammar is poor in places. (For example, delirium is not preceded by the article "a"; it's just "delirium"). It should be proofread by an English language editor prior to re-submission.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Jin Han
Vanderbilt University Medical Center, USA
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
This study sought to validate DEMO to predict the development of delirium during hospitalization. An automated delirium prediction tool is an important endeavor. However, there are some critical limitations of your study:

Comment 1) Your gold standard for delirium threatens the validity of your study. The methods you used to determine the presence of delirium has little detail. It seems you relied upon the EPR, but it is well known that documentation of delirium is poor since the majority of delirium remains unrecognized by the clinical teams [1]. Did you also consider other forms of acute mental status such as “encephalopathy” since this term is interchangeably used with delirium.

Response 1) In order to determine the presence of delirium we relied on the EPR.
As mentioned in the limitations we are aware of the fact that delirium documentation might be on the lower side.
DEMO is a simple aid of use to detect delirium, not a diagnostic tool in itself.
A second analysis which was performed using more search terms has been added into the article as it seems it could be of interest.

Comment 2) Using your chart review method to ascertain delirium makes it difficult to determine when the delirium actually occurred. An episode of delirium that was documented at 48 hours could have actually occurred within 24 hours, but was not recognized by the clinical team until later.

Response 2) These issue is mentioned as a limitation. We are dependent on how and when delirium is documented. As mentioned before we aim at detecting delirium, not at making a diagnostic delirium.

Comment 3) Severity of illness is an important driver of delirium development, but is part of DEMO. Isn't vital sign and lab abnormalities (e.g., APACHE) electronically available?

Response 3) In the article describing the development of the DEMO a full model including lab values was taken into account but the differences with the medication model were almost non-existing in terms of prediction value. Vital signs are electronically available but at this moment it isn't possible to extract such data. The idea of DEMO is to keep it simple, not to go for the an extensive model.

Below are some specific comments:
Page 3, lines 33 to 38, Introduction: There no universally accepted treatment for delirium after it has developed [2]. Though commonly used, there is little evidence that supports the use of antipsychotics in the treatment of delirium [3, 4].
The need for DEMO becomes more compelling in the absence of an effective delirium treatment; preventing delirium is more effective than treating delirium after it occurs.

Response: The text has been adjusted in the manuscript, references are added

Page 4, lines 1 to 10, Introduction: Can you also provide the DEMO instrument in table format? In the table, make sure you place the cutoff used.

Response: Table 1 has been added to the manuscript, explaining also the cut-off point.

Page 4, lines 33 to 39, Methods: Regarding the sample size calculations, I would mention that your sample size was based upon the requirement that the lower limit of your 95%CI would be 60%.

Response: The sentence has been adjusted in the text

Page 6, lines 38 to 39, Limitations: Another limitation is that this is a single center study located in the Netherlands and may not be generalizable in other settings.

Response: The sentence has been added in the text

References
1. Collins N, Blanchard MR, Tookman A, Sampson EL: Detection of delirium in the acute hospital. Age Ageing 2010, 39(1):131-135.
2. AGS/NIA Delirium Conference Writing Group PC, Faculty: The American Geriatrics Society/National Institute on Aging Bedside-to-Bench Conference: Research Agenda on Delirium in Older Adults. Journal of the American Geriatrics Society 2015, 63(5):843-852.
3. Flaherty JH, Gonzales JP, Dong B: Antipsychotics in the treatment of delirium in older hospitalized adults: a systematic review. Journal of the American Geriatrics Society 2011, 59 Suppl 2:S269-276.
4. Neufeld KJ, Yue J, Robinson TN, Inouye SK, Needham DM: Antipsychotic Medication for Prevention and Treatment of Delirium in Hospitalized Adults: A Systematic Review and Meta-Analysis. Journal of the American Geriatrics Society 2016, 64(4):705-714.

Reviewer: 2

Vanja Douglas, MD
University of California, San Francisco, United States of America
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
The authors attempt to validate an electronic delirium prediction rule that uses age and medication information to predict delirium. The risk score (DEMO) is calculated automatically with electronic medical record data. The outcome of delirium for this study was assessed retrospectively by chart review. DEMO has good sensitivity and specificity, but the outcome assessment raises concerns.

Overall, I think the authors should be commended for developing an electronic delirium prediction tool. The study has merits and would be a valuable contribution to the growing literature on delirium prediction. The problem, as is often the case with delirium studies, rests in the outcome ascertainment, which was done through chart review rather than by in person assessment. This is a reasonable approach, but it is incumbent on the authors to explain the approach more clearly, and to either use a validated method (or validate their own).
Thank you for the comments!
The physician diagnosed the delirium, we performed the search for delirium in the EPR. In addition, and as mentioned above, a second analysis using more search terms was performed and has now been added in the manuscript.

Major comments:
Comment 1. In the introduction, the authors say "Treatment measures for delirium include pharmacological..." and also that "Haloperidol or atypical anti-psychotics are the standard pharmacological treatment for delirium...". This is not aligned with current evidence. A recent meta-analysis by Neufeld et al, "Antipsychotic Medication for Prevention and Treatment of Delirium in Hospitalized Adults: A Systematic Review and Meta-Analysis," J Am Geriatr Soc 2016, concluded "Current evidence does not support the use of antipsychotics for prevention or treatment of delirium." Antipsychotics should be reserved for sedation when patients present a danger to themselves or staff, and should be recognized as sedatives and not delirium-specific treatment. I worry this sentence in the introduction would be taken literally and used by providers to justify over-medication of delirious patients, which is a problem in many hospitals. Prior to publication this sentence should be revised to say something like: "There is no evidence to support pharmacologic treatment or prevention of delirium. Antipsychotics are often used for sedation when delirious patients pose a danger to themselves or to hospital staff. On the other hand, non-pharmacological symptom-targeted measures have been shown to prevent delirium in 30-40% of cases..."

Response: The sentence has been adjusted and literature has been added

Comment 2. Methods have a number of concerns:
2a. There is missing data from the case ascertainment. The manuscript says patients over 60 admitted to the hospital were included. Were ALL patients over 60 included? That is what the text suggests in some places, but later it says DEMO patients were randomly selected. The authors need to include a flow diagram showing total number of patients over age 60 admitted during the study period, number excluded due to delirium at admission, number screened by DEMO, and then the number excluded due to unclear diagnosis, and the final number included. The authors can look at the STROBE statement for standard examples of this flow diagram.

Response: A flow diagram (figure 1) has been added to make the inclusion more clear.

Comment 2b. This validation study does not appear to be truly prospective. To be prospective, both the prediction score AND the outcome should be determined prospectively. In this study, DEMO was calculated prospectively, but the outcome was ascertained by chart review retrospectively. This design should be more clearly stated in the methods.

Response: Methods section has been adjusted

Comment 2c. The measurement of the outcome has limitations. The diagnosis of delirium by chart review is notoriously difficult. We have performed similar studies using a chart-review based outcome of delirium and have been roundly criticized for it; while it offers the advantage of being more feasible, it truly does not meet the gold standard of in-person evaluations. However, I think it is a reasonable approach when resources are limited. That said, it needs to be done in a validated way, and it is not clear that the method used here is validated. In fact it's not really clear how the outcome was assessed. A search in the EPR for "delirium" and "delirious" suggests it was simply an electronic search? (As an aside, if this is the case, why didn't the authors include more patients?). If this is the case, how do we know patients where the chart said "this patient is at high risk for delirium" (and were not yet delirious) were not classified as delirious??? How were the patients who were excluded for being delirious at admission identified??? Or was the search a manual review of medical records? If the latter, simply excluding patients because "patient seems confused" is odd; generally that would indicate likely delirium. Inouye et al (PMID: 15673358) report a chart review method for identifying delirium where the sensitivity is 74% and specificity is 84%; if the authors performed a chart review, they should use a cited, validated methodology such as this. If they can't go back and do this, then they need to much more clearly explain how the diagnosis was ascertained using their methods. The quality of this paper really hinges on the identification of subjects and the outcome.
The search was performed by first identifying where in the EPR the words “delirium” and/or “delirious” appeared and then the whole text was read and interpreted to make sure that it was truly delirium diagnosis. Two authors checked and interpreted the charts.
Patients who were excluded because presented with delirium at admission were identified in two ways: by reading the reason for admission which is stated in the EPR and by checking the admission date and the delirium diagnosis; in addition there was no DEMO score for these patients before the diagnosis was documented.

Response: As mentioned above a second analysis using more search terms had now been included in the manuscript.

Comment 2d. Of note, the Inouye chart review method does not describe how to differentiate between prevalent (present on admission) and incident (developing after admisison) delirium. I am not aware of a validated chart review method for doing this. Therefore the authors need to be especially clear about how they did this.

Response: Patients who were excluded because presented with delirium at admission were identified in two ways: by reading the reason for admission which is stated in the EPR and by checking the admission date and the delirium diagnosis; in addition there was no DEMO score for these patients before the diagnosis was documented.

Comment 3. This study begs for a validated, in person outcome assessment. In fact, a better design would have been to validate it in a cohort where the outcome of delirium was prospectively assessed in person, and the DEMO model was measured retrospectively (since it's a chart-based electronic method, all the data should be the same whether examined prospectively or retrospectively). I would change all of the conclusion to state this as the next step instead of saying DEMO is ready for real time use.

Response: Thank you for the comment. We understand it would have been better to do it the other way around. The conclusion section has been adjusted an a new study is being set up!

Minor comments:
Abstract:
Comment 1. Primary outcome measure is not sensitivity and specificity of DEMO - rather, it was development of delirium. The objective of the study was to measure the sensitivity and specificity of DEMO. But it would be more helpful here to explain that the outcome was delirium and how it was measured.

Response: This section has been adjusted

Introduction:
Comment 1. How is polypharmacy defined in DEMO?

Response: Polypharmacy is defined as the number of drugs a patient uses

Methods:
Comment 1. Page 5, line It is 22. What is DD delirium? What does DD stand for here? (differential diagnosis? please define)

Response: The abbreviation has been fully written, DD = differential diagnosis

Comment 2. Please explain what constitutes a positive DEMO screen vs. negative DEMO screen. It appears the DEMO is either positive or negative based on the text available and based on the way the tables are constructed. However, in other places it refers to DEMO predicting a specific % risk, and then being cut into high/low risk groups based on 14.1% being the cut-point. Why was 14.1% chosen?

Response: A DEMO screening is positive when the risk is above 14.1% and negative when it's equal or lower than 14,1%.
This cut-off point is the optimum point of the AUROC of 0,77 that was established during the development of the DEMO. (Hugo AJM de Wit, Bjorn Winkens, Carlota Mestres Gonzalvo, Kim PGM Hurkens, Wubbo J Mulder, Rob Janknegt, Frans R Verhey, Paul-Hugo M van der Kuy, and Jos MGA Schols. The development of an automated ward independent delirium risk prediction model. Int J Clin Pharm (2016) 38:915–923)
The DEMO has been placed into a table and the cut-off point has been better explained in the text.

Comment 3. Was consent obtained or waived?

Response: No consent was needed as the present study had no consequences for the patient. The medical ethical committee agreed on that point. The letter from the medical ethical committee from the 9th of November 2015 states that no informed consent is needed as "there is no question of subjecting persons to treatment or imposing on persons of a particular behavior"

Results:
Comment 1. Table 3: does not need a separate row for man and woman (the numbers for one can be easily calculated from the other).
The table has been adjusted

Discussion:
Comment 1. First sentence is not technically true (not prospective) - see comments above.

Response: The text had been adjusted

Comment 2. Second paragraph - problem with outcome ascertainment is discussed in detail above.

Response: The ascertainment using chart review is further explained as a limitation

Comment 3. Third paragraph - needs to be greatly expanded given the limitation of outcome ascertainment.

Response: The text has been adjusted
Physicians would diagnose delirium and to validate the DEMO we performed a search on different terms; we aim at detecting delirium, not at making a diagnostic delirium

Comment 4. Fourth paragraph - I would emphasize that DEMO uses a clever way of identifying cognitive impairment by including medications used for dementia. For example, donepezil is unlikely to be prescribed except in a patient with cognitive impairment - this is a nice way of getting at this risk factor electronically.

Response: Thank you! It has been added in the manuscript.

Comment 5. Another strength of the DEMO model is that it predicts delirium on a daily basis. This is a novel concept (most delirium prediction rules work at admission but not daily). It is not clear to me whether there is a definite advantage to predicting delirium on a daily basis (it could end up being information overload for nurses and physicians) but it also could eventually be something that is tracked along with the vital signs and intake/output.

Response: Thank you! It has been added in the manuscript.

Overall:
1. English grammar is poor in places. (For example, delirium is not preceded by the article "a"; it's just "delirium"). It should be proofread by an English language editor prior to re-submission.
The manuscript is currently being copyedited, the manuscript will be uploaded asap.

VERSION 2 – REVIEW

REVIEWER	Jin Han Vanderbilt University Medical Center USA No Competing Interest
REVIEW RETURNED	13-Jun-2017

GENERAL COMMENTS	I am still not fully convinced that using the EPR is an accurate enough method to identify patients who are delirious. I fully understand that DEMO tool was not meant to be a diagnostic tool, but "an aid to detect delirium." But even with this goal in mind, your reference standard should be reasonably accurate especially when the title of your manuscript includes the word "validation". Inouye et al. validated a chart based method of delirium identification and reported a sensitivity of 74% and specificity of 83% [1]. But this chart based method used the medical record from the entire hospitalization, and was never meant to identify delirium within a specified time window as you performed. They also used search for terms of an acute confusional state (e.g., delirium, mental status change, inattention, disorientation), hallucinations, agitation, inappropriate behavior, etc.). As a result, I suspect that your chart based delirium ascertainment method has a higher proportion of misclassification especially for your day 1 and 3 analyses. To address this issue, my suggestions would be to: 1) In the methods, mention if your emergency department and inpatient clinicians routinely screened for delirium. If so, what delirium assessment did they use? This would provide some reassurance that your method of delirium ascertainment is reasonable. 2) If they do not routinely screen for delirium, I would consider just presenting Set 4 where you search for all the terms related to an acute confusional state for the first 5 days of the DEMO analysis. 3) In the methods, please specify where you searched for your delirium-related terms. Did you also look at the nurses, consultants, physical therapy, occupational therapy, or speech therapy notes? 4) I think it is important for the readers that you are more transparent about how well your chart based method reflects the patient's delirium status in the methods and the limitations. I would cite Inouye's study [1]. Below are some additional comments: Page 3, lines 37 to 49, Intro: You mention DEMO for the first time, but you don't explain what DEMO is until much later. You may want to consider shortening this section and simply state: (1) There is no effective treatment for delirium; (2) Preventing delirium is by far a more effective strategy to improve patient outcomes, (3) Risk models have been used to identify patients at higher risk for delirium development as these patients would most benefit from delirium prevention, (4) These models are based on manual evaluation of individual risk factors and may be difficult to implement, and (5) automated models are preferable and more feasible.

	Page 4, lines 38 to 39, Methods: Patients with delirium were excluded. How were they excluded (e.g., chart review?). If they were excluded by chart review, I would state that it is possible that patients with delirium may have been included in your non-delirious cohort in the limitations. Table 1: I would consider changing the headings to “Ever Delirious” and “Never Delirious” to indicate that you classified these patients based upon the first 5 day of the hospitalization. If not, References 1. Inouye SK, Leo-Summers L, Zhang Y, Bogardus ST, Jr., Leslie DL, Agostini JV: A chart-based method for identification of delirium: validation compared with interviewer ratings using the confusion assessment method. Journal of the American Geriatrics Society 2005, 53(2):312-318.

REVIEWER	Vanja Douglas University of California, San Francisco, USA
REVIEW RETURNED	14-Jun-2017

GENERAL COMMENTS	The authors have addressed nearly all of the points from the first review, but there is a still a very major concern about the outcome assessment. The authors have clarified this in some regard, but confused the issue further by introducing a new search strategy in the EPR to identify delirium, which led to a different number of delirious patients and different values for the test characteristics of DEMO, which calls into question their entire outcome ascertainment methodology. I think this can all be rectified by going through each chart to identify delirium, rather than performing an electronic search. See further details below. Major points: A) Introduction: Both reviewers asked you to change the part of the introduction mentioning pharmacological treatment of delirium; this was major comment #1 from my initial review. While some sentences have been added, the statement that “Treatment measures for delirium include both pharmacological and non-pharmacological symptom-targeted measures” remains misleading, and I am going to insist that it be removed. As I mentioned in my first review, I think it is misleading and potentially dangerous to say pharmacological measures have any role in the treatment of delirium. B) Methods: 1. In my original review I asked for a “flow diagram showing total number of patients over age 60 admitted during the study period, number excluded due to delirium at admission, number screened by DEMO, and then the number excluded due to unclear diagnosis, and the final number included.” The total number of patients over age 60 admitted during the study period was not included in the flow diagram; please include. 2. Regarding the diagnosis of delirium, thank you for your clarification. This is helpful, but the manuscript text is still unclear.

	The paragraph that begins “A search in the EPR was performed per patient and date…” should explain the chart review aspect of the outcome assessment earlier. After “…professor of geriatric psychiatry,” please move the sentence explaining that after the search identified cases with possible delirium, “the whole text was read and interpreted by two authors to ensure that it was truly a delirium diagnosis.” Then go on to say the result of the chart review was compared to the DEMO score. 3. Who were the authors doing the chart review? Makes a big difference if they were research assistants, pharamcists, geriatricians, psychiatrists, etc. Please specify. 4. I would also specify that the date of delirium onset was determined by chart review. 5. Regarding delirium at admission – this is still confusing. In the flow diagram, it says 450 patients were enrolled. Then 21 were excluded for delirium at admission. Then 429 underwent chart review. How were the 21 found to have delirium at admission prior to chart review, if chart review is how delirium was diagnosed? I presume simply by looking at the reason for admission in the EPR, prior to chart review? But, then it is strange that no further patients were excluded for delirium at admission after chart review. I would think that looking at the reason for admission would not identify 100% of patients who were delirious at admission. 21 of 450 patients (5%) is a VERY LOW number for delirium at admission – delirium is a common reason for hospital admission and most studies cite a higher percentage of delirium upon admission (usually >10%). 7. The question of what constituted an unclear diagnosis is still left unresolved. Why, if the chart said “patient seems confused” was the patient not considered delirious? If the chart said “patient seems confused but attention is normal and level of alertness is normal, without fluctuation, so this is not clearly delirium” I would agree this was not delirium, but if the chart says “patient seems confused and this is a clear change from his baseline very clear mental status” that is highly likely to be delirium. The authors need to explain more about the unclear cases that were excluded. 8. The “SETs” still don’t make sense. If your delirium diagnosis was made by chart review, why are you saying that the SETs were “delirium” or “delirious” as diagnosis? That suggests your outcome was diagnosed purely by searching for these terms, and not a thorough chart review. Just call it delirium or no delirium. 9. Finally - and this is the most critical point of this review - SET 4 also does not make sense. If you did a broader search for these delirium terms, and then updated your chart review, why not include the results of the updated chart review in the first three SETs? The key here is the FINAL delirium diagnosis by chart review, not the search strategy you used in the EPR to find potential cases. I would use the search strategy from SET 4 for SETs 1, 2, and 3, and eliminate SET 4 from the analysis. In fact, to expound on this further, the following sentence in the discussion is a red flag: “For this reason we performed a wider search using other words that might suggest delirium, however, this method resulted in more FN.”

	You can NOT have two different methods of outcome ascertainment and then just pick the one that resulted in better test characteristics for prediction model. Furthermore, as I stated above, it doesn't make sense how you can have different rates of sensitivity/specificity/FN/FP for SET 4 and SET 3. If your delirium diagnosis is based on chart review, and the two search strategies are identifying different numbers of delirious patients, then you have a major problem with your outcome ascertainment and you need to go back and review all the charts fully. In fact I think this is really critical for this study to have validity and future impact, and because it hasn't been addressed properly I think you need to go back and do this prior to publication. Minor points: A) Abstract: Change the sentences "DEMO is a satisfactory prediction model. The next step will be to validate the DEMO in a cohort where the outcome of delirium is assessed prospectively in person by the physician, and the DEMO model is used for retrospective measurements." To "DEMO is a satisfactory prediction model, but needs further prospective validation with in-person delirium ascertainment." B) Introduction: How is polypharmacy defined in DEMO? This has not been adequately addressed. Table 2, which is helpful, is not interpretable. The abbreviations need to be defined. Specifically for polypharmacy, what is "ATC-5th"? If V2 is simply the number of drugs a patient is prescribed, please explain that in Table 2. Also, it would be helpful to know what medicines were included in each category. C) Discussion: 1. From the first review: "Third paragraph - needs to be greatly expanded given the limitation of outcome ascertainment. The text has been adjusted Physicians would diagnose delirium and to validate the DEMO we performed a search on different terms; we aim at detecting delirium, not at making a diagnostic delirium" My point here is that you still need a valid outcome assessment to validate DEMO as a prediction rule. I understand that it is not a diagnostic tool – but you need to measure the outcome DEMO is predicting in a valid way. 2. Even though I said that the DEMO uses a clever way of identifying cognitive impairment by capturing the use of dementia drugs, you shouldn't say is was "clever" in your manuscript. Please revise this sentence to be a bit more humble. D) Overall: Manuscript still needs additional English-language copyediting, but it is much improved.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Jin Han
Vanderbilt University Medical Center, USA
Please state any competing interests or state 'None declared': None

Please leave your comments for the authors below

I am still not fully convinced that using the EPR is an accurate enough method to identify patients who are delirious. I fully understand that DEMO tool was not meant to be a diagnostic tool, but "an aid to detect delirium." But even with this goal in mind, your reference standard should be reasonably accurate especially when the title of your manuscript includes the word "validation". Inouye et al. validated a chart based method of delirium identification and reported a sensitivity of 74% and specificity of 83% [1]. But this chart based method used the medical record from the entire hospitalization, and was never meant to identify delirium within a specified time window as you performed. They also used search for terms of an acute confusional state (e.g., delirium, mental status change, inattention, disorientation), hallucinations, agitation, inappropriate behavior, etc.). As a result, I suspect that your chart based delirium ascertainment method has a higher proportion of misclassification especially for your day 1 and 3 analyses. To address this issue, my suggestions would be to:

Comment 1) In the methods, mention if your emergency department and inpatient clinicians routinely screened for delirium. If so, what delirium assessment did they use? This would provide some reassurance that your method of delirium ascertainment is reasonable.

Response: It has been adjusted in the text. Both in the ER and in the wards patients are routinely screened for delirium. The first screening is performed by a validated checklist (Dutch healthcare inspectorate and Safety Management System) [35]. The results from this checklist give an indication for the risk to develop delirium. When the risk is high, the DOS method [20] is used to evaluate whether a patient has a delirium.

Comment 2) If they do not routinely screen for delirium, I would consider just presenting Set 4 where you search for all the terms related to an acute confusional state for the first 5 days of the DEMO analysis.

Response: Not applicable as patients are routinely screened for delirium

Comment 3) In the methods, please specify where you searched for your delirium-related terms. Did you also look at the nurses, consultants, physical therapy, occupational therapy, or speech therapy notes?

Response: It has been specified in the text. The search was performed in the patients' charts where different healthcare professionals including nurses, physiotherapists, speech therapists etc. note their findings regarding a patient.

Comment 4) I think it is important for the readers that you are more transparent about how well your chart based method reflects the patient's delirium status in the methods and the limitations. I would cite Inouye's study [1].

Response: Thank you for the tip! The methods and discussion/limitation sections have been adjusted mentioning the Inouye's study. Inouye et al. make clear that chart based method has certain limitations. In our study the chart method was used to establish the prediction value of the DEMO score.

Below are some additional comments:

Page 3, lines 37 to 49, Intro: You mention DEMO for the first time, but you don't explain what DEMO is until much later. You may want to consider shortening this section and simply state: (1) There is no effective treatment for delirium; (2) Preventing delirium is by far a more effective strategy to improve patient outcomes, (3) Risk models have been used to identify patients at higher risk for delirium development as these patients would most benefit from delirium prevention, (4) These models are based on manual evaluation of individual risk factors and may be difficult to implement, and (5) automated models are preferable and more feasible.
The section has been adjusted according comment.
"There is no effective treatment for delirium [24,25]. Preventing delirium is by far a more effective strategy to improve patient outcomes [1, 4,26-29]. Risk models have been used to identify patients at higher risk for delirium development as these patients would most benefit from delirium prevention; These models are based on manual evaluation of individual risk factors and may be difficult to implement, and automated models are preferable and more feasible [30-34]."

Page 4, lines 38 to 39, Methods: Patients with delirium were excluded. How were they excluded (e.g., chart review?). If they were excluded by chart review, I would state that it is possible that patients with delirium may have been included in your non-delirious cohort in the limitations.

Response: The text has been adjusted mentioning that it was performed by chart review. This issue has also been mentioned in the limitations section.

Table 1: I would consider changing the headings to "Ever Delirious" and "Never Delirious" to indicate that you classified these patients based upon the first 5 day of the hospitalization. If not,

Response: The table has been adjusted mentioning delirium or no-delirium during admission to clarify the concept.
We hope this is what was meant given the fact the last sentence (if not, ….) was not complete.

References
1. Inouye SK, Leo-Summers L, Zhang Y, Bogardus ST, Jr., Leslie DL, Agostini JV: A chart-based method for identification of delirium: validation compared with interviewer ratings using the confusion assessment method. Journal of the American Geriatrics Society 2005, 53(2):312-318.

Reviewer: 2
Vanja Douglas
University of California, San Francisco, USA
Please state any competing interests or state 'None declared': None declared
---
Please leave your comments for the authors below

Comment: The authors have addressed nearly all of the points from the first review, but there is a still a very major concern about the outcome assessment. The authors have clarified this in some regard, but confused the issue further by introducing a new search strategy in the EPR to identify delirium, which led to a different number of delirious patients and different values for the test characteristics of DEMO, which calls into question their entire outcome ascertainment methodology. I think this can all be rectified by going through each chart to identify delirium, rather than performing an electronic search. See further details below.

Response; Thank you!
Apart from the electronic search, the EPR during the admission period was studied to look for missing values. The text has been adjusted to clarify that it was not merely a search.

Major points:
Comment A) Introduction: Both reviewers asked you to change the part of the introduction mentioning pharmacological treatment of delirium; this was major comment #1 from my initial review. While some sentences have been added, the statement that "Treatment measures for delirium include both pharmacological and non-pharmacological symptom-targeted measures" remains misleading, and I am going to insist that it be removed. As I mentioned in my first review, I think it is misleading and potentially dangerous to say pharmacological measures have any role in the treatment of delirium.

Response: The section has been adjusted in the text.

Comment B) Methods:
1. In my original review I asked for a "flow diagram showing total number of patients over age 60 admitted during the study period, number excluded due to delirium at admission, number screened by DEMO, and then the number excluded due to unclear diagnosis, and the final number included." The total number of patients over age 60 admitted during the study period was not included in the flow diagram; please include.

Response: The diagram has been adjusted showing that 2534 patients were included. Also the point mention below has been addressed as patients with delirium at admission were excluded after performing the chart review.

2. Regarding the diagnosis of delirium, thank you for your clarification. This is helpful, but the manuscript text is still unclear. The paragraph that begins "A search in the EPR was performed per patient and date…" should explain the chart review aspect of the outcome assessment earlier. After "…professor of geriatric psychiatry," please move the sentence explaining that after the search identified cases with possible delirium, "the whole text was read and interpreted by two authors to ensure that it was truly a delirium diagnosis." Then go on to say the result of the chart review was compared to the DEMO score.

Response; We agree it makes more sense when described in this order. The text has been adjusted

3. Who were the authors doing the chart review? Makes a big difference if they were research assistants, pharmacists, geriatricians, psychiatrists, etc. Please specify.
KH (internist geriatrician) and CMG(hospital pharmacist) were the authors who performed the chart review.

Response: It has been mentioned in the text.

4. I would also specify that the date of delirium onset was determined by chart review.

Response: The text has been adjusted to “The date of delirium onset was determined by chart review.”

5. Regarding delirium at admission – this is still confusing. In the flow diagram, it says 450 patients were enrolled. Then 21 were excluded for delirium at admission. Then 429 underwent chart review. How were the 21 found to have delirium at admission prior to chart review, if chart review is how delirium was diagnosed? I presume simply by looking at the reason for admission in the EPR, prior to chart review? But, then it is strange that no further patients were excluded for delirium at admission after chart review. I would think that looking at the reason for admission would not identify 100% of patients who were delirious at admission. 21 of 450 patients (5%) is a VERY LOW number for delirium at admission – delirium is a common reason for hospital admission and most studies cite a higher percentage of delirium upon admission (usually >10%).

Response: Delirium at admission was identified by chart review. The flow diagram was confusing and it has been adjusted. As for the low incidence, we could imagine that some from the unclear data patients also presented with a delirium at admission. In that case the incidence would be higher, matching better with literature.

7. The question of what constituted an unclear diagnosis is still left unresolved. Why, if the chart said “patient seems confused” was the patient not considered delirious? If the chart said “patient seems confused but attention is normal and level of alertness is normal, without fluctuation, so this is not clearly delirium” I would agree this was not delirium, but if the chart says “patient seems confused and this is a clear change from his baseline very clear mental status” that is highly likely to be delirium. The authors need to explain more about the unclear cases that were excluded.

Response: I completely agree with the comment. As explained in the text the EPR (chart) was fully read by two authors to identify delirium patients. On the wither search this unclear cases were evaluated and classified as diagnosis positive or negative.

8. The “SETs” still don’t make sense. If your delirium diagnosis was made by chart review, why are you saying that the SETs were “delirium” or “delirious” as diagnosis? That suggests your outcome was diagnosed purely by searching for these terms, and not a thorough chart review. Just call it delirium or no delirium.

Response: It has been adjusted in the text (also in combination with comment 9). It makes it more clear when not using the word SET and just talking about delirium diagnosis 1,3 or 5 days after the DEMO analysis.

9. Finally - and this is the most critical point of this review - SET 4 also does not make sense. If you did a broader search for these delirium terms, and then updated your chart review, why not include the results of the updated chart review in the first three SETs? The key here is the FINAL delirium diagnosis by chart review, not the search strategy you used in the EPR to find potential cases. I would use the search strategy from SET 4 for SETs 1, 2, and 3, and eliminate SET 4 from the analysis.

Response: Thank you for this critic comment. We do realize it had to be adjusted.
As suggested we have used the search strategy 1, 3 and 5 days after the demo considering all search terms as a delirium diagnosis as the method used was a full chart review.
The text and results have been adjusted accordingly.

In fact, to expound on this further, the following sentence in the discussion is a red flag: “For this reason we performed a wider search using other words that might suggest delirium, however, this method resulted in more FN.” You can NOT have two different methods of outcome ascertainment and then just pick the one that resulted in better test characteristics for prediction model. Furthermore, as I stated above, it doesn’t make sense how you can have different rates of sensitivity/specificity/FN/FP for SET 4 and SET 3. If your delirium diagnosis is based on chart review, and the two search strategies are identifying different numbers of delirious patients, then you have a major problem with your outcome ascertainment and you need to go back and review all the charts fully. In fact I think this is really critical for this study to have validity and future impact, and because it hasn’t been addressed properly I think you need to go back and do this prior to publication.
As mentioned above this issue has been solved in the article. The chart review was documented for all search terms and this analysis has now been used in the current version of the article.

Minor points:
A) Abstract: Change the sentences “DEMO is a satisfactory prediction model. The next step will be to validate the DEMO in a cohort where the outcome of delirium is assessed prospectively in person by the physician, and the DEMO model is used for retrospective measurements.” To “DEMO is a satisfactory prediction model, but needs further prospective validation with in-person delirium ascertainment.”

Response: It has been adjusted as suggested.

B) Introduction:
How is polypharmacy defined in DEMO? This has not been adequately addressed. Table 2, which is helpful, is not interpretable. The abbreviations need to be defined. Specifically for polypharmacy, what is “ATC-5th”? If V2 is simply the number of drugs a patient is prescribed, please explain that in Table 2. Also, it would be helpful to know what medicines were included in each category.

Response: The table has been adjusted. The abbreviations have been defined. As for polypharmacy, it refers to the number of drugs a patients was prescribed. The drugs in the other categories are also mentioned, specifying that all drugs within the ATC (anatomical therapeutic chemical classification system) mentioned were included.

C) Discussion:
1. From the first review: “Third paragraph - needs to be greatly expanded given the limitation of outcome ascertainment.
The text has been adjusted
Physicians would diagnose delirium and to validate the DEMO we performed a search on different terms; we aim at detecting delirium, not at making a diagnostic delirium”
My point here is that you still need a valid outcome assessment to validate DEMO as a prediction rule. I understand that it is not a diagnostic tool – but you need to measure the outcome DEMO is predicting in a valid way.

Response: It has been adjusted in the text adding that the first screening for delirium is performed using a validated checklist, and the fact that healthcare professionals other than physicians also write in the chart.

2. Even though I said that the DEMO uses a clever way of identifying cognitive impairment by capturing the use of dementia drugs, you shouldn't say is was "clever" in your manuscript. Please revise this sentence to be a bit more humble.

Response: It has been adjusted in the text by saying an alternative way of identifying cognitive impairment.

D) Overall:
Manuscript still needs additional English-language copyediting, but it is much improved.

Response: The revised manuscript has been copy-edited by another editing service.

VERSION 3 – REVIEW

REVIEWER	Jin Han Vanderbilt University Medical Center USA
REVIEW RETURNED	17-Jul-2017

GENERAL COMMENTS	Thank you so much for your revisions. The description of how delirium was ascertained still needs additional detail. Page 4, lines 22 to 30, Methods: I am unfamiliar with the Dutch Healthcare Inspectorate Safety Management System checklist. The reference you provided is a website that links to a page in a foreign language. If it has been validated, what is its sensitivity and specificity? Can you provide some details of this checklist as a supplemental? You stated that the DOSS was performed in patients at high risk. Who performed the DOSS (nurses vs physicians) and what is its sensitivity and specificity? Please provide a citation for this assessment. Page 5, lines 8 to 14, Methods: Was a positive DOSS or the presence of these terms indicative of delirium? Page 7, lines 22 to 32, Limitations: The reason to cite the Inouye reference is to state that the chart review method has a 30% false positive rate, and that misclassification may have occurred. I had difficulty understanding what you wrote for this citation.

REVIEWER	Vanja Douglas, MD University of California, San Francisco, USA
REVIEW RETURNED	11-Jul-2017

GENERAL COMMENTS	Thank you for your revision; this manuscript is much improved. The outcome assessment is much more clear now. English writing is vastly improved. Introduction, Screening instrument - refers to Table 2 (supplementary) but in the submission, the DEMO model is Table 1. Methods: Please see the attached file with comments and tracked changes to the methods section. I had to rearrange the writing in the methods substantially to make it comprehensible. I believe my changes are all accurate with respect to the study methods. Please review and accept them if you agree. Results: 1. Tables need help. First, you keep saying "(supplementary tables)" - why? You only seem to have 4 tables and 1 figure - can't they all be in the text and not supplementary? Table 4 is unnecessary. Just say what the mean age and gender of your cohort was and delete Table 4. For example: Patients who developed delirium were older (mean age xx (SD)) compared to those who did not (mean age xx (SD)). X% of patients with delirium were male compared to X% of patients without delirium. 1a. The paragraph about differences in age between delirium and non-delirium patients is not necessary. Just state your statistics in the sentence above. The granularity of differences between day 1 and day 5 patients does not matter. 2. Table 2 is still a little confusing. Column headers say "delirium during admission" but you really mean "Delirium by day 1 after DEMO" and "Delirium by day 3 after DEMO". Right? I would eliminate the Day 1, Day 3, and Day 5 column headers, and change "Delirium during admission" to say "Delirium by day 1 after DEMO" | "No delirium by day 1" | "Delirium by day 3 after DEMO" | No delirium by day 3 | Delirium by day 5 after DEMO | No delirium by day 5. 3. You can't say that sensitivity decreases over time or specificity increases, since these differences are not statistically significant. However, you show discuss the PPV and NPV of the DEMO over time. PPV goes from 43% to 65% from day 1 to day 5. This is not presented in any tables, so would be nice to state in the text, and a statistical comparison should be made. 4. Incidentally, if DEMO is done on day 0, and then delirium outcome measured on days 1-5, why is the total number of DEMO positives different on Days 1, 3, and 5? There are 125 positive DEMOs on day1, 125 on day3, and 127 on day5. That should not be the case. Similarly, the totals for DEMO negatives don't add up (ie. day 5 is different). Figure: First box should be "Eligible patients" or "Patients admitted during the study period". N=46 unclear data/unclear diagnosis should be "N=46 insufficient information in chart to determine delirium status"

VERSION 3 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Jin Han
Institution and Country: Vanderbilt University Medical Center, USA Competing Interests: None declared

Thank you so much for your revisions. The description of how delirium was ascertained still needs additional detail.

Comment: Page 4, lines 22 to 30, Methods: I am unfamiliar with the Dutch Healthcare Inspectorate Safety Management System checklist. The reference you provided is a website that links to a page in a foreign language. If it has been validated, what is its sensitivity and specificity? Can you provide some details of this checklist as a supplemental? You stated that the DOSS was performed in patients at high risk. Who performed the DOSS (nurses vs physicians) and what is its sensitivity and specificity? Please provide a citation for this assessment.

Response: The questionnaire used to evaluate the initial risk for delirium is not validated. It is a questionnaire developed by an expert team in combination with the healthcare inspectorate. The questionnaire is mention as a quality indicator for delirium risk assessment and it is, according to the Dutch guidelines, the method to follow to firstly screen patients. I have added the guidelines in the reference list and could send the checklist as supplemental but they are in Dutch. Therefore I have added the questions in the checklist in the article text.
More information about the DOSS method has been added to the text including sensitivity and specificity and references.

Page 5, lines 8 to 14, Methods: Was a positive DOSS or the presence of these terms indicative of delirium?

Response: A positive DOSS score was not interpreted as delirium. The terms mentioned in the methods section were used to perform the initial search and then the complete chart review took place asserting, taken into account the given information, if a patient had truly developed a delirium.

Page 7, lines 22 to 32, Limitations: The reason to cite the Inouye reference is to state that the chart review method has a 30% false positive rate, and that misclassification may have occurred. I had difficulty understanding what you wrote for this citation.

Response: Thank you for the remark. The text has been adjusted mentioning the 30% and stating more clearly that this is a limitation.

Reviewer: 2
Reviewer Name: Vanja Douglas, MD
Institution and Country: University of California, San Francisco, USA Competing Interests: None declared

Comment: Thank you for your revision; this manuscript is much improved. The outcome assessment is much more clear now.

Response: Thank you!

Comment: English writing is vastly improved.

Response: Thank you!

Comment: Introduction, Screening instrument - refers to Table 2 (supplementary) but in the submission, the DEMO model is Table 1.

Response: It has been adjusted.

Comment: Methods: Please see the attached file with comments and tracked changes to the methods section. I had to rearrange the writing in the methods substantially to make it comprehensible. I believe my changes are all accurate with respect to the study methods. Please review and accept them if you agree.

Response: Thank you! It has been rewritten as suggested.

Results:
1. Tables need help. First, you keep saying "(supplementary tables)" - why? You only seem to have 4 tables and 1 figure - can't they all be in the text and not supplementary? Table 4 is unnecessary. Just say what the mean age and gender of your cohort was and delete Table 4. For example: Patients who developed delirium were older (mean age xx (SD)) compared to those who did not (mean age xx (SD)). X% of patients with delirium were male compared to X% of patients without delirium.
1a. The paragraph about differences in age between delirium and non-delirium patients is not necessary. Just state your statistics in the sentence above. The granularity of differences between day 1 and day 5 patients does not matter.

Response: The tables have been integrated in the text.
Table 4 has been deleted and the findings have been described as text.

“Patients who developed delirium were older (mean age 83.9 (sd 7.8)) compared to those who did not (mean age 73.9 (sd 9.1)). 50.0% patients with delirium were male compared to 50.1% of patients without delirium. A statistically significant difference in mean age was found between delirium and non-delirium (equal variances assumed, $p<0.001$, mean difference=10.0, 95%CI 7.6 to 12.5). There was no significant difference in the percentage of delirium between men and women ($p=0.911$) ”

2. Table 2 is still a little confusing. Column headers say "delirium during admission" but you really mean "Delirium by day 1 after DEMO" and "Delirium by day 3 after DEMO". Right? I would eliminate the Day 1, Day 3, and Day 5 column headers, and change "Delirium during admission" to say "Delirium by day 1 after DEMO" | "No delirium by day 1" | "Delirium by day 3 after DEMO" | No delirium by day 3 | Delirium by day 5 after DEMO | No delirium by day 5.

Response: The headers have been adjusted as suggested.

3. You can't say that sensitivity decreases over time or specificity increases, since these differences are not statistically significant. However, you show discuss the PPV and NPV of the DEMO over time. PPV goes from 43% to 65% from day 1 to day 5. This is not presented in any tables, so would be nice to state in the text, and a statistical comparison should be made.

Response: The PPV and NPV have been calculated and added to table 2. Statistical comparison has been performed.

4. Incidentally, if DEMO is done on day 0, and then delirium outcome measured on days 1-5, why is the total number of DEMO positives different on Days 1, 3, and 5? There are 125 positive DEMOs on day1, 125 on day3, and 127 on day5. That should not be the case. Similarly, the totals for DEMO negatives don't add up (ie. day 5 is different).

Response: Thank you for the comment! The numbers were wrongly inserted. It has been adjusted.

Figure:
First box should be "Eligible patients" or "Patients admitted during the study period".
N=46 unclear data/unclear diagnosis should be "N=46 insufficient information in chart to determine delirium status"
The text in the figure has been adjusted.

VERSION 4 – REVIEW

REVIEWER	Jin Han Vanderbilt University Medical Center USA No Competing Interest
REVIEW RETURNED	09-Sep-2017

GENERAL COMMENTS	Thank you for your revisions. 1) I appreciate the additional details of the initial delirium screen which is a checklist of 3 questions. I still have concerns of how sensitive the checklist is; there does not seem to be any validation data. 2) With regards to the EPR keywords on page 4, can you clarify that both the DOSS and the key words were used to define the presence or absence of delirium? 3) On page 5, you state that " If a diagnosis of delirium could not be established for a patient as a result of insufficient information in the chart, this patient was excluded from the analysis". In the first paragraph of the Results, I would state that you could come up with a diagnosis in 46 patients. This is a significant proportion and should be mentioned in the limitations. This may limit the generalizability of your study.

	4) In the second paragraph of the discussion, I am still concerned with the statement indicating that DEMO can predict delirium on a daily basis. What you showed is that DEMO can predict delirium for up to 5 days after hospitalization. A true daily prediction model should incorporate data from the day before the day of interest.

REVIEWER	Vanja Douglas University of California, San Francisco; USA
REVIEW RETURNED	25-Sep-2017

GENERAL COMMENTS	Thank you for addressing all of my comments and for your patience with me as your reviewer. This paper is an interesting and important contribution to the delirium literature. I think it is essentially ready for acceptance, although there are still a few minor stylistic issues. For example, in Table 3, there are too many numbers. The authors write Prevalence = "0.162 (16.2%)". This pattern is followed for every value - sensitivity, specificity, PPV and NPV. Why do you write a decimal and follow it with a %? Just write one or the other. I think it would make the most sense to just write the %. All you are currently doing is showing one number next to the same number multiplied by 100. Presumably this can be corrected during copy editing, but that decision is up to the editors.

VERSION 4 – AUTHOR RESPONSE

Reviewer: 1

Please leave your comments for the authors below Thank you for your revisions.

Comment 1) I appreciate the additional details of the initial delirium screen which is a checklist of 3 questions. I still have concerns of how sensitive the checklist is; there does not seem to be any validation data.

Response: We do agree that a validated tool would be of much more value , but unfortunately, this checklist is not a validated tool. As mentioned in the manuscript the checklist is provided by the Dutch healthcare inspectorate and included in the guidelines so every hospital is meant to use it.

Nevertheless, this checklist is used to evaluate whether a patients might be prone to develop a delirium and thus other methods to detect delirium should be applied. In that way we don't think the fact this checklist is not validated affects our study. In addition it has been added in the limitations section.

Comment 2) With regards to the EPR keywords on page 4, can you clarify that both the DOSS and the key words were used to define the presence or absence of delirium?

Response: We didn't use the DOSS as a key word or as a delirium diagnostic tool. The fact the DOSS is mentioned is purely to show that the delirium documentation is robust as whenever a DOSS score is high it is documented in the chart that the patient has a delirium (or one of the other keywords is used). The text has been adjusted making the link between the DOSS and the documentation in the chart. In addition it has been added in the discussion.

Comment 3) On page 5, you state that " If a diagnosis of delirium could not be established for a patient as a result of insufficient information in the chart, this patient was excluded from the analysis". In the first paragraph of the Results, I would state that you could come up with a diagnosis in 46 patients. This is a significant proportion and should be mentioned in the limitations. This may limit the generalizability of your study.

Response: We assume it is meant that for 46 patients we could NOT come up with a diagnosis. The text has been adjusted as suggested in the result section and it has been mentioned as a limitation.

Comment 4) In the second paragraph of the discussion, I am still concerned with the statement indicating that DEMO can predict delirium on a daily basis. What you showed is that DEMO can predict delirium for up to 5 days after hospitalization. A true daily prediction model should incorporate data from the day before the day of interest.

Response: What we mean with "on a daily basis" is that DEMO daily screens all patients and then predicts delirium for up to 5 days. The sentence has been adjusted to:" Another strength of DEMO is that it predicts delirium within 5 days post-analysis on a daily basis.". We hope this makes it more clear for the reader.

Reviewer: 2

Please leave your comments for the authors below

Comment: Thank you for addressing all of my comments and for your patience with me as your reviewer. This paper is an interesting and important contribution to the delirium literature. I think it is essentially ready for acceptance, although there are still a few minor stylistic issues. For example, in Table 3, there are too many numbers. The authors write Prevalence = "0.162 (16.2%)". This pattern is followed for every value - sensitivity, specificity, PPV and NPV. Why do you write a decimal and follow it with a %? Just write one or the other. I think it would make the most sense to just write the %. All you are currently doing is showing one number next to the same number multiplied by 100. Presumably this can be corrected during copy editing, but that decision is up to the editors.

Response: The table has been adjusted showing only the percentages. We agree it makes the table more clear.