Validation of an automated delirium prediction model (DElirium MOdel (DEMO)): an observational study

Carlota Mestres Gonzalvo,¹,² Hugo A J M de Wit,¹,³ Brigit P C van Oijen,¹ Debbie S Deben,¹ Kim P G M Hurkens,⁴ Wubbo J Mulder,⁵ Rob Janknegt,¹ Jos M G A Schols,⁶ Frans R Verhey,⁷ Bjorn Winkens,⁸ Paul-Hugo M van der Kuy¹,⁹

ABSTRACT

Objectives Delirium is an underdiagnosed, severe and costly disorder, and 30%–40% of cases can be prevented. A fully automated model to predict delirium (DEMO) in older people has been developed, and the objective of this study is to validate the model in a hospital setting.

Setting Secondary care, one hospital with two locations.

Design Observational study.

Participants The study included 450 randomly selected patients over 60 years of age admitted to Zuyderland Medical Centre. Patients who presented with delirium on admission were excluded.

Primary outcome measures Development of delirium through chart review.

Results A total of 383 patients were included in this study. The analysis was performed for delirium within 1, 3 and 5 days after a DEMO score was obtained. Sensitivity was 87.1% (95% CI 0.756 to 0.939), 84.2% (95% CI 0.732 to 0.915) and 82.7% (95% CI 0.734 to 0.893) for 1, 3 and 5 days, respectively, after obtaining the DEMO score. Specificity was 77.9% (95% CI 0.729 to 0.882), 81.5% (95% CI 0.766 to 0.856) and 84.5% (95% CI 0.797 to 0.884) for 1, 3 and 5 days, respectively, after obtaining the DEMO score.

Conclusion DEMO is a satisfactory prediction model but needs further prospective validation with in-person delirium confirmation. In the future, DEMO will be applied in clinical practice so that physicians will be aware of when a patient is at an increased risk of developing delirium, which will facilitate earlier recognition and diagnosis, and thus will allow the implementation of prevention measures.

INTRODUCTION

A delirium or acute confused state is a transient attention and cognition disorder that develops over a short period of time and occurs mainly in hospitalised patients and people aged 60 years and over. Delirium is an underdiagnosed, severe (increased mortality), costly and often preventable disorder.¹–³ Its severity and symptoms can vary considerably, but the main features are impaired cognitive and sensory functions, reduced consciousness and diminished attention. In addition, it is often accompanied by problems with psychomotor activity, the circadian rhythm and emotions.

The prevalence and incidence of delirium in the general population differ widely depending on the setting. The overall prevalence in the community is estimated to be 1%–2%. In a hospital setting, this prevalence increases to 10%–31% at the time of hospital admission and 3%–29% during hospitalisation. The incidence increases up to 87% when more specialised populations, such as the elderly and people in postoperative, intensive care and/or palliative care, are considered.⁴⁻¹¹ In 30%–40% of cases, delirium is preventable, which, along with its associated high costs (ranging from US$164 billion to US$182 billion per year), makes it a perfect target for interventions by healthcare professionals.¹ ¹¹⁻¹⁵⁻¹⁷⁻¹⁸ As a result, a great number of screening tools have been developed and are widely used to detect the early onset of delirium, which can in turn allow treatment measures to be introduced in a timely manner.¹⁶⁻²¹ These tools help healthcare professionals to establish and quantify symptoms associated with delirium.¹⁹⁻²⁵ Once the diagnosis has been established, the underlying medical condition can be targeted, and delirium can be managed appropriately.
There is no effective treatment for delirium. Preventing delirium is by far a more effective strategy to improve patient outcomes. Risk models have been used to identify patients at higher risk for delirium development because these patients would most likely benefit from delirium prevention. These models are based on manual evaluation of individual risk factors and may be difficult to implement, so automated models are preferable and more feasible.

**Screening instrument**

A fully automated model to predict delirium in older people (aged over 60 years) was developed at Zuyderland Medical Centre. This DEMOdel (DEMO) uses only electronically available data to predict the occurrence of delirium. The predictive variables include age, polypharmacy and the use of antidementia drugs, antidepressants, anti-Parkinson’s agents, antidiabetic drugs, analgesia and/or sleeping tablets (see box). This model can be applied hospital-wide and has an area under receiver operating characteristic (measure for model prediction quality) value of 0.770 (95% CI 0.736 to 0.804) with a sensitivity of 78.2% and a specificity of 63.7%, when 14.1% is used as a cut-off value for the predicted probability of developing delirium. DEMO was developed retrospectively but has not yet been validated.

Therefore, the objective of this study is to validate DEMO in a hospital setting. To do so, the system’s accuracy (main study parameter), that is, sensitivity (proportion of delirium patients who test positive) and specificity (proportion of non-delirium patients who test negative), will be calculated. In addition to these parameters, the positive predictive value (PPV), negative predictive value (NPV), positive and negative likelihood ratios (LR+, LR−) with their 95% CI will be computed.

**METHODS**

This is an observational study of the ability of DEMO to predict delirium in an elderly hospital population. It was conducted in Zuyderland Medical Centre (locations Sittard and Heerlen) in the period from January 2016 to October 2016.

DEMO involves a daily analysis of all hospitalised patients aged ≥60 years at the different wards and predicts whether a patient is at risk of developing delirium in a 24-hour postanalysis period. The electronic patient record (EPR) was accessed at a later date to check for delirium diagnosis. In this study, DEMO was calculated prospectively, but the outcome was ascertained by chart review retrospectively.

Although delirium diagnosis was determined by chart review, delirium documentation in our hospital is robust. At admission, patients are routinely screened for delirium, both in the emergency department and in the ward. The first screening is performed by a checklist (Inspectie voor de Gezondheidszorg=Dutch Healthcare Inspectorate, Veiligheids management systeem=Safety Management System and Dutch guideline for delirium) This checklist consists of three questions: does the patient need help with self-care?, has the patient previously suffered a delirium?, does the patient suffer from memory disorders? When one of the questions is positively answered, the patient is at risk of developing delirium; in this case, the Delirium Observation Screening Scale (DOSS) method is used to evaluate whether a patient has delirium and it is subsequently noted in the chart.

Patients aged over 60 years who were admitted to Zuyderland were eligible for enrolment. From all patients admitted between 31 December 2015 and 31 October 2016, 450 patients were randomly selected (using [https://www.randomlists.com/team-generator](https://www.randomlists.com/team-generator)) and their charts extracted for review. Patients who, based on chart review, presented with delirium on admission were then excluded (see online supplementary figure 1).


The search was performed by first identifying where the different words appeared in the EPR, and then, if any of these words appeared, the whole EPR during the admission period was read and interpreted by two authors (KH (internist geriatrician) and CMG (hospital pharmacist)) to determine whether it was truly a delirium diagnosis. All notes were reviewed, including notes by physicians, nurses, physiotherapists and speech therapists. During the study, treating healthcare professionals (physicians, nurses, etc) were blinded to DEMO scores in order to avoid bias. If a diagnosis of delirium could not be established for a patient as a result of insufficient information
Delirium diagnosis based on chart review was then compared with the risk score from DEMO. The DEMO was dichotomised into two groups: high risk ≥14.1% and low risk <14.1% for this analysis. A two-by-two table was then constructed to calculate true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rates.

The predictive value of DEMO was determined for delirium developing within 1, 3 and 5 days after the DEMO score was calculated. It had been developed to predict delirium within the next 24 hours, but here we wished to also investigate whether its predictive value could be extended to 3 or 5 days.

In the study wherein the DEMO was developed, an incidence rate of 17.4% was used. Given the assumption of the same specificity of 0.75 (75%), we calculated that 33 delirium patients were needed based on the requirement that the lower limit of 95% CI would be at least 60% (width of 95% CI ≤0.30 (30%)). With regard to the specificity, the number of non-delirium patients would be much larger than the number of delirium patients, and hence, the width of the 95% CI for specificity would be <0.30.

It was assumed that at least 332 patients would be needed to identify 33 delirium patients. Taking into account the exclusion criteria and the possibility of a smaller percentage of patients who would develop delirium, a sufficient number of patients were screened to obtain 33 delirium patients (ie, 450 patients).

The sensitivity, specificity, PPV, NPV, LR+, LR− with corresponding 95% CIs were calculated with the use of an online calculator (http://vassarstats.net/ch1.html). The differences in PPV and NPV over time were tested using McNemar’s test. The differences in age and gender between delirium and non-delirium groups were tested by using independent-samples t-test and χ² test, respectively. IBM SPSS statistics for Windows (V.23.0) was used to perform these tests. A two-sided p≤0.05 was considered statistically significant.

RESULTS
The study lasted 8 months, for 450 patients chart review was undergone. Finally, a total of 383 patients were included, as 21 patients presented with delirium at admission, and for 46 patients there was insufficient information to determine delirium status (see online supplementary figure 1). The results of the diagnostic test (TP/FP/FN/TN) for 1, 3 and 5 days after DEMO analysis are shown in table 1. The analysis, including prevalence estimates, sensitivity, specificity, PPV, NPV and LRs, is presented in table 2. Although sensitivity decreased and specificity increased if the period increased from 1 day to 3 or 5 days after DEMO score was obtained, all values were rather high (sensitivity ≥0.827, specificity ≥0.779). PPV was statistically different p<0.001 for all three comparisons (1 vs 3 vs 5 days after DEMO analysis).
Patients who developed delirium within 5 days were significantly older (mean age 83.9 (SD 7.8)) compared with those who did not develop a delirium within 5 days (mean age 73.9 (SD 9.1); p<0.001). There was no significant difference in the percentage of males within the delirium and non-delirium groups (50.0% vs 50.1%, p=0.911).

DISCUSSION AND CONCLUSION

In the current study, a previously developed model for predicting delirium has been validated. DEMO was calculated prospectively, and the outcome was ascertained by chart review retrospectively. Based on the current data and the high sensitivity and specificity, it can be concluded that DEMO is a satisfactory prediction model.

Another strength of DEMO is that it predicts delirium within 5 days post-analysis on a daily basis. This is a novel concept, as most delirium prediction rules apply at admission but not daily. Even though it is not clear whether there is a definite advantage to predicting delirium on a daily basis, as this could lead to information overload, it could eventually be something that is tracked along with vital signs and intake/output.

We found sensitivity and specificity rates that were higher than reported in the study by de Wit et al, which may be because his study only checked the patients' medical history for delirium and not the entire EPR. Moreover, de Wit et al had performed the search merely on the diagnosis of delirium. In the current study, the full EPR during the admission period was taken into account, and a wider set of terms was considered for delirium diagnosis. Furthermore, in the current study, in those cases in which delirium was not clear, these patients were excluded, whereas such patients had been included in the development of the delirium model.4

The present study does present some limitations. First, the validation of the DEMO depends on how and when a healthcare professional reports that a patient has developed delirium. It is well known that documentation of delirium is poor since the majority of delirium remains unrecognised by clinical teams.37 We therefore performed a wider search considering other words that might suggest delirium as delirium diagnosis and read through the whole EPR during the admission period. The number of delirium patients is noticeably higher than originally found, which can be explained by the search we performed. The DEMO is merely an aid to detect delirium, not a diagnostic tool by itself. Furthermore, for 46 patients there was insufficient information in the chart to determine delirium status, which could influence the generalisability of the present study.

In addition, as mentioned in the study by Inouye et al,38 using a chart review method has some limitations as it has a 30% false-positive rate and thus it is possible that patients with delirium at admission may have been included in the non-delirious cohort due to poor documentation in the chart.

Furthermore, the checklist used to screen the patients is a non-validated tool. Nevertheless, after that first check, the DOSS is used. The DOSS method is a validated method used by nurses to screen for delirium. Its sensitivity ranges from 89% to 100% and its specificity ranges from 88% to 96.6%.39 40 The DOSS scores and its conclusion (delirium/non-delirium) are recorded in the chart. In that way, and taking into account that the chart is a complete document in which different healthcare professionals note their findings, makes the outcome more reliable and strengthens the validity of the present study.

Another limitation of the present study is that it is a single-centre study (two hospital locations) located in the Netherlands and may not be generalisable in other settings.

The DEMO uses only electronically available data. Other important factors that could predict a delirium (previous delirium, cognitive impairment, severity of disease, visual impairment, etc) are not included in this model because they were not electronically available. If these data were also made electronically available, the predictive quality of DEMO could be improved.22 23 27 30

Taking into account that the registration of such factors is becoming increasingly important and mandatory, it is only a matter of time until these important factors can be used in the DEMO.23 In addition, DEMO already uses an alternative way of identifying cognitive impairment by including medications used for dementia.

The DEMO is a fully automated satisfactory prediction model that predicts delirium up to 5 days after analysis. The next step is to validate the DEMO in a cohort in which the outcome of delirium would be prospectively assessed in person and to use DEMO for retrospective measurements. In the future, DEMO will be applied to clinical practice so that physicians are alerted when a patient is at increased risk of developing delirium. This will facilitate earlier recognition and diagnosis and, thus, the implementation of prevention measures.
Contributors All authors have made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data. They all have been involved in drafting the manuscript and revising it critically for important intellectual content. They all have given final approval of the version to be published; and they all agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Conception or design of the work: CMG, HAJMvD, BPCvO, DSD, KPGM, WMJ, RJ, JMGAS, FRV, BW, PHMvK. Data collection: CMG, DSD, KPGM. Data analysis and interpretation: CMG, KPGM, FRV, BW, PHMvK. Drafting the article: CMG. Critical revision of the article: HAJMvD, BPCvO, JMGAS, FRV, BW, PHMvK. Final approval of the version to be published: CMG, HAJMvD, BPCvO, DSD, KPGM, WMJ, RJ, JMGAS, FRV, BW, PHMvK.

Funding This work was supported by ZonMw (the Netherlands Organisation for Health Research and Development) grant number (113101001).

Competing interests None declared.

Ethics approval The medical ethics committee METC Z (Medisch Ethische Toetsings Commissie van Zuyderland en Zuyd Hogescholen, Zuyderland Medical Centre, Heerlen) approved this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data are anonymised and will be confidentially handled. Only the investigators have access to the data. All patient data will be kept for as long as the project is being conducted.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, for as long as the project is being conducted.

REFERENCES
