NEWS2 shows low sensitivity and high specificity for delirium detection: a single site observational study of 13,908 patients

Authors: Emma RLC Vardy,[A] Schanhave Santhirasekaran,[B] Michael Cheng,[C] Atul Anand[D] and Alasdair M MacLullich[E]

ABSTRACT

Delirium affects 25% of hospital admissions of older people and is a serious medical condition with poor outcomes. 'New confusion' as a delirium indicator was incorporated into the 'alert, verbal, pain and unresponsive' (AVPU) level of consciousness scale in the National Early Warning Score 2 (NEWS2) in 2017. We measured sensitivity of non-alert NEWS2 (new confusion and/or V, P or U ratings) for delirium through comparison with the four 'A's test (4AT) delirium tool in 13,908 consecutive non-elective hospital admissions. We included NEWS2 scores 4 hours before or after 4AT. There were 2,802 (20%) admissions with positive 4AT and 594 (4.3%) with non-alert NEWS2 status. Sensitivity of NEWS2 for 4AT $\geq$4 was 17.8% (95% confidence interval (CI) 16.4–19.2), and specificity was 99.1% (95% CI 98.9–99.3). These findings suggest that NEWS2 in current practice has low sensitivity but high specificity for delirium. Further research is needed to improve routine inpatient monitoring for delirium.

KEYWORDS: NEWS2, delirium, 4AT, detection, alertness

Introduction

Delirium is a serious medical condition that, if not accurately diagnosed and managed, leads to adverse outcomes in hospital. Early detection of delirium is important and is known to influence clinical outcomes.[1] The Royal College of Emergency Medicine recommends cognitive assessment on admission to hospital and delirium detection specifically.[2] The National Early Warning Score 2 (NEWS2) is an aggregate scoring system to identify acutely unwell patients, calculated from routinely collected physiological measurements. The aim of the system is to improve early identification of significant acute illness in patients and to standardise the assessment and response. Developed in 2012 by the Royal College of Physicians, NEWS was updated to NEWS2 to include new confusion in in the consciousness section in 2017.[3] It was recognised early on that there was ambiguity on how to assess new confusion and, for this reason, additional supplementary guidance was published.[3] Anecdotally, NEWS2 remains variably utilised to assess for new confusion and, hence, delirium. Through analysis of large-scale routine data from a single hospital site, this study aimed to answer the following questions.

> How often are NEWS2 and/or the four 'A's test (4AT) completed in non-elective hospitalised patients aged over 65 years old?
> Is non-alert status on NEWS2 sensitive for delirium detection in a population screened using 4AT?
> How does the level of consciousness scoring on the NEWS2 compare to 4AT scoring?
> Whether patient factors (such as age or sex) influence the likelihood of new confusion scoring on NEWS2.

In answering these questions, we aimed to ascertain the validity of using NEWS2 to identify new confusion and, hence, possible delirium.

Methods

Study population and setting

This study was conducted in Salford, UK. Part of the Northern Care Alliance NHS Foundation Trust, Salford Royal is a tertiary referral hospital in the north of England covering a population of approximately 240,000 people. This study included consecutive non-elective admissions aged 65 years or older. All non-elective admissions admitted through the emergency department between 01 March 2020 and 30 March 2022 were included, where at least one 4AT was completed within 24 hours of first attendance and at least one NEWS2 assessment occurred within 4 hours either side of this delirium screen. The period of 4 hours

Authors: [A]consultant geriatrician, Northern Care Alliance NHS Foundation Trust, Salford, UK and School of Health Sciences and NIHR Applied Research Collaboration Greater Manchester, Manchester, UK; [B]junior data scientist, Northern Care Alliance NHS Foundation Trust, Salford, UK; [C]data scientist, Mersey Care NHS Foundation Trust, Prescot, UK; [D]senior clinical research fellow, Centre for Cardiovascular Science, Edinburgh, UK and honorary consultant in geriatric medicine, Royal Infirmary of Edinburgh, Edinburgh, UK; [E]professor of geriatric medicine, Usher Institute, Edinburgh, UK and Royal Infirmary of Edinburgh, Edinburgh, UK

was chosen as a pragmatic window based on advised frequency of assessment using NEWS2. This comes from the recommended clinical response to NEWS trigger thresholds and ranges from continuous monitoring for NEWS2 of 7 or above, to 12-hourly for a score of zero.[4] This was designed to allow fair comparison between assessments for new confusion using NEWS2, with possible delirium using the 4AT, carried out at similar points in time for an individual patient. The start date was chosen to mirror the update of NEWS2 to include assessment for new confusion in the level of consciousness category.

4AT testing

The 4AT is a validated brief delirium assessment tool designed for clinical use.[5] It has four items: alertness, a four-point abbreviated mental test, a test of attention, and recognition of acute change or fluctuating course. The 4AT yields a 0–12 score range. We used the standard cut-offs of 4AT 0 (normal test), 4AT 1–3 (possible cognitive impairment, no delirium) and 4AT ≥4 (possible delirium ± cognitive impairment). The 4AT is embedded in the electronic health record (EHR) in Salford (Allscripts, Sunrise). Completion of the assessment is recommended for all non-elective admissions aged ≥65 years and recorded rates for Salford Royal hospital have been 49%.[1] If more than one eligible score was available for a patient, the first completed assessment was used.

NEWS2 assessment

NEWS2 includes six physiological parameters: respiration rate, oxygen saturation, systolic blood pressure, pulse rate, level of consciousness (including new-onset confusion) and temperature.[3] Each of these parameters is allocated a score of 0, 1, 2 or 3. The electronic form capture of NEWS2 level of consciousness is shown in supplementary material S1; Fig S1. For this study, NEWS2 assessments that were completed within 4 hours of the 4AT test were analysed. Sensitivity analyses were completed to assess shorter periods between 4AT and NEWS2 assessment at ≤3, ≤2 and ≤1 hour. Where more than one eligible NEWS2 score was available, the first completed assessment was used. New confusion is captured within a question on consciousness where a patient could also be rated as 'alert', or responsive to voice, or responsive to pain, or unresponsive (the AVPU scale). Note that altered level of consciousness is alone a highly specific indicator of delirium and complements assessment for new confusion, which typically includes assessment of cognition including attention.[6] Importantly, for any given NEWS2 assessment, only one state is possible. For example, a person cannot be scored as both alert and have new confusion. We, therefore, considered all non-alert NEWS2 states as possible delirium, reflecting the potential identification of both hyper- and hypoactive states. For ease, we refer to any of new confusion, responsive to voice, responsive to pain or unresponsive (ie CVPU) as a combined 'non-alert' NEWS2 category.

Outcomes

The primary outcome was the sensitivity of non-alert NEWS2 assessment scores (considered collectively as positive or negative) for 4AT possible delirium (score ≥4). Secondary outcomes included other measures of diagnostic performance: specificity,

positive predictive values (PPVs) and negative predictive values (NPVs). Given one 4AT component specifically assesses alertness, we further compared responses on this single question to NEWS2 recorded levels of consciousness. In patients with 4AT ≥4, age and sex differences between those classified as alert or non-alert by NEWS2 were also explored.

Statistical analysis

All analyses were conducted using R, including *tidyverse*, *lubridate*, *caret*, *epiR* and *RColorBrewer* packages.[7] Due to the categorical nature of the 4AT and NEWS2 assessments, the data are predominantly presented as absolute numbers (percentages) and continuous data are presented as mean ± standard deviation (SD). Sensitivity, specificity, PPVs and NPVs of the NEWS2 non-alert state for a referent 4AT measure ≥4 were determined using standard confusion matrices. NEWS2 consciousness responses were analysed at each 4AT score (between 0–12) and visualised using proportional stacked bar charts. Age and sex differences between groups were assessed by Wilcoxon rank and chi-squared tests, respectively.

Ethics

Access to data was provided in accordance with service evaluation of an existing intervention (NEWS2) recommended in local and national guidelines.

Results

The study period included 29,231 eligible hospital admissions, of which, 16,326 (56%) had at least one 4AT completed within 24 hours. The analysis population included 13,908 consecutive admissions where at least one NEWS2 score was entered within 4 hours of the admission 4AT. Most NEWS2 assessments occurred within 1 hour of the 4AT (65%; Fig 1).

There were 2,802 (20%) admissions with a 4AT ≥4 consistent with probable delirium. A total of 594 (4.3%) admissions had a non-alert status recorded on NEWS2. The sensitivity of a NEWS2 non-alert assessment for a 4AT ≥4 was 17.8% (95% confidence interval (CI) 16.4–19.2), with a specificity of 99.1% (95% CI 98.9–99.3). The full diagnostic performance metrics are reported in supplementary material S1; Table S1.

Increasing 4AT scores were associated with a greater probability of a non-alert assessment on NEWS2 (Fig 2), but even at the highest possible 4AT score (12 points), most patients were assessed as alert on NEWS2 (67.1%). More than one in every four patients with an alert status on NEWS2 had an abnormal 4AT score >0 (Fig 3).

There was no difference in age between patients with both 4AT ≥4 and non-alert status on NEWS2 compared with those only identified by positive 4AT who were rated alert on NEWS2 (82.0 years ± 8.0 vs 82.4 years ± 7.7, respectively; p=0.24). There was also no significant difference in sex (57.8% female vs 56.8% female, respectively; p=0.70). Considering only those identified as non-alert by NEWS2, age did not differ by 4AT status (81.1 years ± 8.4 for 4AT <4 vs 82.0 years ± 8.0 for 4AT ≥4; p=0.32) and sex distributions were again similar (65.6% female vs 57.8% female, respectively; p=0.19).

Of those with 4AT ≥4, a total of 18% had a NEWS2 non-alert assessment within 4 hours, compared with 11% when only NEWS2
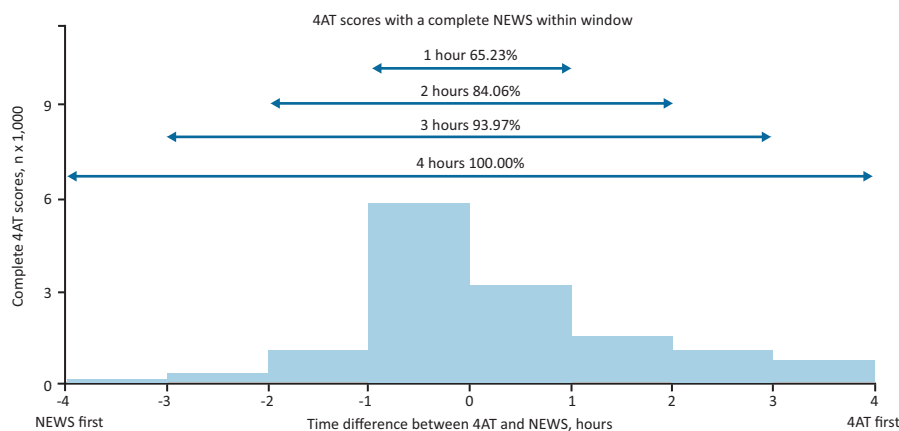
Fig 1. Time difference between the admission four 'A's test assessment and closest National Early Warning Score 2. 4AT = four 'A's test; NEWS = National Early Warning Score.

scores within 1 hour of 4AT were considered (supplementary material S1; Fig S2).

The specific alertness question on the 4AT scored 980 (7%) admissions as having 'clearly abnormal' alertness and 825 (6%) as having 'mild sleepiness'. Out of the patients classed as 'clearly abnormal' on 4AT, 70.8% were assessed as alert by NEWS2. Out of the patients classified with 'mild sleepiness' on 4AT, 88.5% were assessed as alert by NEWS2 (Fig 4).

Discussion

The novel findings in this study are that, in routine practice, NEWS2 has low sensitivity but high specificity for delirium based on 4AT scoring. Of those assessed as having clearly abnormal alertness on 4AT, more than three-quarters of these were assessed as being alert on NEWS2. Hence, despite updated guidance on how to assess for new confusion as part of NEWS2, the scoring as currently implemented cannot be relied upon for delirium detection.

There are several potential explanations for inaccurate completion of the consciousness domain on NEWS2. First, the level of consciousness assessment does not accommodate rating that a person may be both alert and confused, which may force clinicians into choosing one or other option. Second, the presentation on the EHR may also influence completion, in that new confusion is the last option available. Third, the term 'new confusion' can be interpreted in a number of ways in relation to duration of onset,

nature and severity of the perceived confusion. Fourth, scoring this item requires knowledge of the patient's baseline cognitive status.

There are important implications for patient escalation. Mohammed *et al* completed a two-centre study to estimate the impact of adding delirium to NEWS on the number of medium- or high-level alerts that trigger assessment by a clinician with expertise in acute care or a clinician with expertise in critical care, respectively.[8] The study found that medium-level alerts would increase by 25.7% and 20.7% at each site, and high-level alerts would increase by 26% and 26.1%. A positive score on the 4AT is a strong prognostic marker, with a recent study finding that 23% of patients with an admission 4AT score of ≥4 died in that admission.[1] Yet 4AT scores of ≥4 do not trigger escalation alerts in the same standardised way as higher NEWS2 scores. Perceived workload concerns about escalating patients for 'confusion' alone may, in part, account for why healthcare staff are more likely to identify delirium using a 4AT but are not recording this via NEWS2. Further work should explore this issue.

Limitations

There are some limitations of this study that should be acknowledged. We included only data on hospital admission. It is possible that there is variable completion of NEWS2 dependent on the completing group of staff; for example, medical admissions unit staff as opposed to geriatric medical ward staff involved
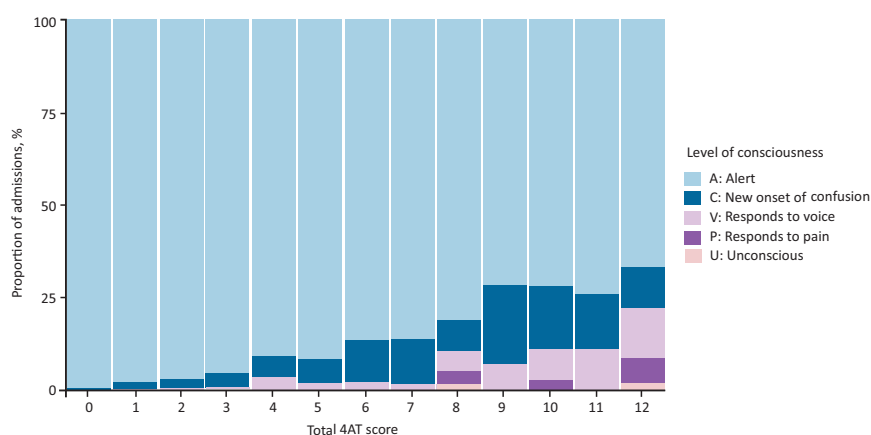

Fig 2. Breakdown of level of consciousness assessment by National Early Warning Score 2 at each scoring level of four 'A's test (4AT) assessment. The 4AT thresholds for clinical interpretation are 4AT 0 (normal test), 4AT 1–3 (possible cognitive impairment) and 4AT ≥4 (possible delirium ± cognitive impairment). 4AT = four 'A's test.
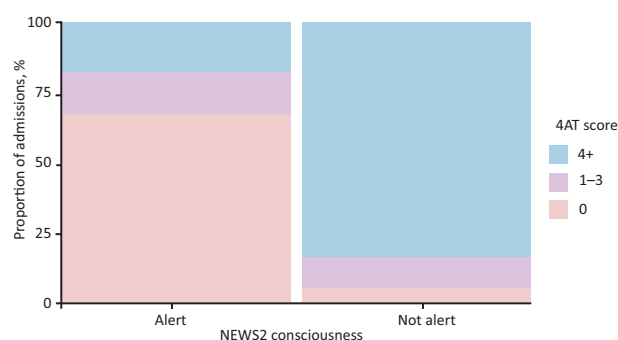
Fig 3. Distribution of four 'A's test scores among patients with alert and non-alert status on National Early Warning Score 2. 4AT = four 'A's test; NEWS = National Early Warning Score.

later in an admission. Of the 29,231 admissions, we were only able to analyse data for 13,908 admissions for this study. This was because of incomplete 4AT, a lack of 4AT completion within 24 hours and lack of NEWS2 completion within 4 hours of available 4AT scores. In this centre, there are ongoing efforts to encourage use of the 4AT for delirium screening where completion on admission is not mandatory; some centres show higher rates of completion, suggesting that higher rates of 4AT completion are possible with implementation work, including education and mandatory completion.[1] In a prior study, patients without a 4AT measured had outcomes that were broadly intermediate between 4AT scores of 0 and scores of 1–3, and 4 or more, suggesting that non-completion is unlikely to signal more severe cases in the sample as a whole.[1] Additionally, the data presented are from a single centre and there may be variation in completion accuracy across other organisations. Finally, only people aged 65 years and over were included in the analysis. However, it is unlikely that age influenced NEWS2 scoring given that in those with 4AT ≥4 there was no difference in age between those scored as alert or non-alert on the NEWS2.

Conclusion

Our data bring into question the role of NEWS2 as currently implemented for detection of delirium. It appears that the current scoring system may not reliably be used to detect delirium. One option may be improved options for recording alertness and confusion in the next update to the NEWS2. Alternatively, the way that NEWS2 is completed in EHRs could be altered to directly link
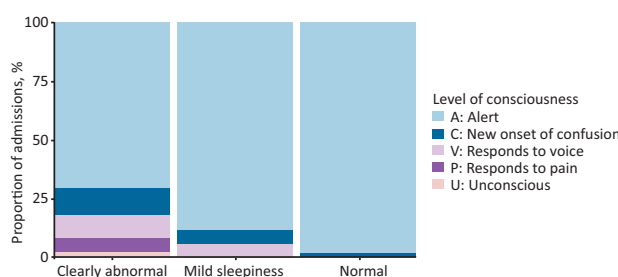

Fig 4. Relationship between alertness assessment by four 'A's test and level of consciousness on National Early Warning Score 2. 4AT = four 'A's test.

to the 4AT, at least on admission. The recommendation of the triple assessment using the 4AT, NEWS2 and Clinical Frailty Scale recommended by the Getting It Right First Time (GIRFT) geriatric medicine report[9] may be a meaningful way forward for older patients. ∎

Summary

What is known?

The 4AT is a well validated tool to assess for delirium. NEWS2 was updated in 2017, in recognition of the seriousness of delirium as an indicator of acute illness, to incorporate new confusion.

What is the question?

Is NEWS2 reliable in the detection of possible delirium?

What was found?

The NEWS2 is very specific in detection of possible delirium but is insensitive and many cases are missed.

What is the implication for practice now?

NEWS2 cannot be relied upon as a delirium detection method, and the completion of 4AT in parallel (for example, as part of the GIRFT recommended triple assessment for older people) might be considered in any future iterations.

Supplementary material

Additional supplementary material may be found in the online version of this article at www.rcpjournals.org/clinmedicine: S1 – Additional table and figures.

Conflicts of interest

Alasdair M MacLullich is the main author of the 4AT (see www. the4AT.com); the 4AT is free to download and use, and there are no current or future financial interests. Alasdair M MacLullich was co-chair of the committee that produced the 2019 Scottish Intercollegiate Guidelines Network (SIGN) guideline on delirium in which the 4AT is recommended.

References

1 Anand A, Cheng M, Ibitoye T *et al*. Positive scores on the 4AT delirium assessment tool at hospital admission are linked to mortality, length of stay and home time: two-centre study of 82,770 emergency admissions. *Age Ageing* 2022;51:afac051.

2 The Royal College of Emergency Medicine. *Assessing for cognitive impairment in older people*. RCEM, 2021. https://rcem.ac.uk/wp-content/uploads/2021/10/RCEM_Older_People_National_Report_2019_20.pdf [Accessed 11 July 2022].

3 Royal College of Physicians. *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS*. RCP, 2017. www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2 [Accessed 11 July 2022].

4 Royal College of Physicians. *Chart 2: thresholds and triggers*. RCP, 2017. www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2 [Accessed 11 July 2022].

5 MacLullich AMJ. *The 4 "A"s Test*. www.the4AT.com [Accessed 11 July 2022].

6 European Delirium Associaiton and American Delirium Society. The DSM-5 criteria, level of arousal and delirium diagnosis: inclusiveness is safer. *BMC Medicine* 2012;12:141.

7   R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, 2020. www.R-project.org [Accessed 11 July 2022].

8   Mohammed MA, Faisal M, Richardson D *et al*. The inclusion of delirium in version 2 of the National Early Warning Score will substantially increase the alerts for escalating levels of care: findings from a retrospective database study of emergency medical admissions in two hospitals. *Clin Med* 2019;19: 104–8.

9   Hopper A. *Getting it Right First Time Report: Geriatric Medicine*. NHS, 2021. www.gettingitrightfirsttime.co.uk/medical-specialties/geriatric-medicine [Accessed 11 July 2022].

**Address for correspondence: Professor Emma Vardy, Salford Care Organisation, Stott Lane, Salford M6 8HD, UK.
Email: emma.vardy@nca.nhs.uk
Twitter: @emmavardy2**