


ORIGINAL



Natural language processing diagnosed behavioral disturbance vs confusion assessment method for the intensive care unit: prevalence, patient characteristics, overlap, and association with treatment and outcome

Marcus Young^{1,5}, Natasha Holmes¹, Kartik Kishore¹, Nada Marhoon¹, Sobia Amjad^{1,2}, Ary Serpa-Neto^{1,3} and Rinaldo Bellomo^{1,3,4,5*} 

© 2022 The Author(s)

Abstract

Purpose: To compare the prevalence, characteristics, drug treatment for delirium, and outcomes of patients with Natural Language Processing (NLP) diagnosed behavioral disturbance (NLP-Dx-BD) vs Confusion Assessment Method for intensive care unit (CAM-ICU) positivity.

Methods: In three combined medical-surgical ICUs, we obtained data on demographics, treatment with antipsychotic medications, and outcomes. We applied NLP to caregiver progress notes to diagnose behavioral disturbance and analyzed simultaneous CAM-ICU.

Results: We assessed 2313 patients with a median lowest Richmond Agitation-Sedation Scale (RASS) score of -2 (-4.0 to -1.0) and median highest RASS score of 1 (0 to 1). Overall, 1246 (53.9%) patients were NLP-Dx-BD positive (NLP-Dx-BD^{pos}) and 578 (25%) were CAM-ICU positive (CAM-ICU^{pos}). Among NLP-Dx-BD^{pos} patients, 539 (43.3%) were also CAM-ICU^{pos}. In contrast, among CAM-ICU^{pos} patients, 539 (93.3%) were also NLP-Dx-BD^{pos}. The use of antipsychotic medications was highest in patients in the CAM-ICU^{pos} and NLP-Dx-BD^{pos} group (24.3%) followed by the CAM-ICU^{neg} and NLP-Dx-BD^{pos} group (10.5%). In NLP-Dx-BD^{neg} patients, antipsychotic medication use was lower at 5.1% for CAM-ICU^{pos} and NLP-Dx-BD^{neg} patients and 2.3% for CAM-ICU^{neg} and NLP-Dx-BD^{neg} patients (overall $P < 0.001$). Regardless of CAM-ICU status, after adjustment and on time-dependent Cox modelling, NLP-Dx-BD was associated with greater antipsychotic medication use. Finally, regardless of CAM-ICU status, NLP-Dx-BD^{pos} patients had longer duration of ICU and hospital stay and greater hospital mortality (all $P < 0.001$).

Conclusion: More patients were NLP-Dx-BD positive than CAM-ICU positive. NLP-Dx-BD and CAM-ICU assessment describe partly overlapping populations. However, NLP-Dx-BD identifies more patients likely to receive antipsychotic medications. In the absence of NLP-Dx-BD, treatment with antipsychotic medications is rare.

Keywords: Intensive care, Delirium, Critical illness, Quetiapine, Olanzapine

*Correspondence: Rinaldo.bellomo@austin.org.au

⁴ Department of Intensive Care, Austin Hospital, 145 Studley Rd, Heidelberg, VIC, Australia

Full author information is available at the end of the article

Introduction

Delirium is common in critically ill patients [1] and associated with poor outcomes [2, 3]. The fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM IV) defines delirium as a “disturbance in consciousness that is accompanied by a change in cognition that cannot be accounted for by a preexisting or evolving dementia” [4]. The manual describes three criteria that may be used to identify such a disturbance in consciousness; A, “reduced clarity of awareness of the environment; B, “accompanying change in cognition” or “development of a perceptual disturbance” and C, “develops over a short period of time and tends to fluctuate during the course of the day”.

The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM V) reconsiders delirium diagnosis listing five diagnostic criteria. Three describe disturbance in behavior (i.e., “disturbance in attention”, “develops over a short period” and “disturbance in cognition”), and two describe the absence of pre-existing disturbances or conditions (i.e., “preexisting, established or evolving neurocognitive disorder” and “a direct physiological consequence of another medical condition”) that, if present, would exclude a diagnosis of delirium [5]. Although, more restrictive than the DSM IV and with dependences on interpretation, the DSM V criteria have been found to identify a similar population of patients to the DSM IV criteria [6, 7].

Despite such constructs, the systematic identification of delirium in intensive care unit (ICU) patients through direct application of these criteria is considered challenging [8]. As a consequence, derivative methodologies have been developed to operationalize screening for delirium in the ICU [9]. Of these, the Confusion Assessment Method for ICU (CAM-ICU) [10, 11] and the Intensive Care Delirium Screening Checklist (ICDSC) [12], both derived from the DSM IV, have been recommended as the most reliable for use within the adult ICU [13]. In our ICUs, the CAM-ICU methodology is used by bedside nurses once per shift to assess patients for possible delirium. The results of this assessment are documented in the electronic progress notes. However, progress notes also contain nursing, medical, and allied health caregivers’ narrative observations, which may further describe a patient’s behavior and cognitive state.

We hypothesized that, in their clinical progress notes, caregivers might use words suggestive of disturbed behavior or cognition such as “agitated” and “confused”, which would increase the information available for the assessment of the patient’s cognitive state. In a recent study [14], we successfully used Natural Language Processing (NLP) techniques to detect these words in the

Take-home message

Natural language processing (NLP) of electronic caregiver notes is a novel and powerful epidemiological tool to identify behavioral disturbance in critically ill patients. NLP identifies more patients with abnormal behavior and disturbed cognitive state, who will receive antipsychotic medications, are more severely ill, likely to stay in intensive care unit (ICU) and hospital for longer, and more likely to die than CAM-ICU positive patients.

electronic clinical progress notes of a large cohort of critically ill patients. NLP is a software technology that is used to automatically analyze textual information in documentation such as progress notes. We referred to patients identified by such words as having “NLP-Diagnosed Behavioural Disturbance” (NLP-Dx-BD).

Accordingly, in this study, we compared the prevalence, clinical characteristics, outcomes, and antipsychotic medications-based treatment of patients with NLP-Dx-BD with those of patients with CAM-ICU-based delirium screening assessments.

Methods

Study design

This was a non-interventional, retrospective study, which used data derived from the electronic health records (EHR) of a university affiliated ICU system in Melbourne Australia. This study was approved by the Austin Hospital Human Research Ethics Committee (LNR/19/Austin/38), which waived the need for informed consent. We included all adult patients (≥ 18 years old) admitted to the ICUs of the Austin Hospital, Melbourne between 1 May 2019 and 31 December 2020. If a patient had multiple admissions only the first admission was considered for inclusion. No additional exclusion criteria were applied.

Data collection and manipulation

All baseline and outcome data were collected from the Australian and New Zealand Intensive Care Society Adult ICU Patient Database run by the Centre for Outcome and Resource Evaluation [15]. Using a proprietary intensive care clinical information system, we obtained electronic data from all progress notes entered into the ICU-specific electronic health record (EHR) by clinical staff including doctors, nurses, physiotherapists and other allied health professionals.

During the study period, by ICU policy, patients received general care aimed at decreasing the risk of delirium, including frequent family visits, dimmed lights at night, minimal interaction to facilitate nighttime sleep cycling, and ensuring use of spectacles and hearing aids as necessary. CAM-ICU and Richmond

Agitation-Sedation Scale (RASS) assessments were recorded in the clinical progress notes by nursing staff during every 8-h shift. NLP techniques were then used to extract the detail of these assessments from the notes. Further, the clinical progress notes of all caregivers were analyzed using NLP tokenizing techniques (natural language toolkit; NLTK 3.5) [16]. As previously described, [14] progress notes were converted to sentence vectors, each vector was then searched for the presence of words, terms, or expressions, suggestive of behavioral disturbance in accordance with terms selected in a previously published survey of clinical staff [17] (etable 1a and 1b in online supplement). In this survey, clinical staff were asked to identify words, terms or expressions that they would use to describe a situation where they thought a patient was exhibiting disturbed behavior.

Finally, we obtained data on antipsychotic medications used in our ICUs (haloperidol, olanzapine, quetiapine, and risperidone) from the hospital EHR and correlated them with other study data. We hypothesized that treatment with antipsychotic medications may reflect caregivers discomfort with patients cognitive dysfunction and disturbed behavior. Accordingly, we chose this outcome as the clinically relevant primary outcome for our study.

Data definition

We searched nursing progress notes for CAM-ICU and RASS assessments using NLP. Simultaneously, we analyzed medical, nursing and allied health progress notes for words describing NLP-Dx-BD.

We classified patients as CAM-ICU positive, negative, or unable to be assessed (e.g. the patient does not speak English or the patient refuses to cooperate) or missing.

We then classified patients as CAM-ICU positive when, in at least one progress note, a CAM-ICU positive assessment was reported.

We classified patients as NLP-Dx-BD positive when, in at least one progress note, we found a word indicative of behavioral disturbance (eTable 1a and eTable 1b).

Exposure

The primary exposure of the present study was the combination of CAM-ICU and NLP-Dx-BD diagnoses. Patients were, therefore, classified in four groups:

Group 1: Negative for both assessments.

Group 2: Only CAM-ICU positive.

Group 3: Only NLP-Dx-BD positive.

Group 4: Positive for both assessments.

Outcomes

The primary outcome was use of antipsychotic medications. Such drugs included haloperidol, quetiapine,

olanzapine and risperidone. Despite controversy [18], these medications represent the most frequent medical intervention applied to agitation, perceived delirium and/or behavioral disturbance in ICU.

Secondary outcomes included the duration of mechanical ventilation (in ventilated patients), ICU and hospital length of stay, and ICU, hospital, and 28-day in-hospital mortality.

Statistical analysis

All continuous data are reported as medians (quartile 25%–quartile 75%) and categorical data as numbers and percentages. Baseline, clinical characteristics, and outcomes of the patients were compared among the groups using Fisher exact test and the Kruskal–Wallis test. The proportion of patients receiving medications over time is presented in Kaplan–Meier curves and compared with the log-rank test. Multivariable logistic regression models were used to assess the association of the exposure groups with hospital mortality, with Group 1 (CAM-ICU negative and NLP-Dx-BD negative) used as reference. To account for immortal time bias, we additionally conducted a time-dependent Cox proportional hazard model for the primary outcome and hospital mortality that considered all measurements available in each note. For the primary outcome only exposures happening before the first outcome (first time the patient received an antipsychotic) were included in the model to avoid exposures measured after the medication had already been given. The models were adjusted by age, type of admission, and by the Australian and New Zealand Risk of Death (ANZROD) after log transformation [19]. ANZROD is a validated and accurate predictor of mortality in ICUs in Australia and New Zealand [20]. Effect estimates were reported as hazard ratio (OR) with its 95% confidence interval (CI).

Sensitivity analyses were performed stratifying the cohort according to the use of mechanical ventilation. Overall rate of missing data was low, and is reported in the eTable 2 in the Online Supplement. All analyses were case complete analysis and were conducted in R v.4.0.3 (R Foundation) [21] and a P value < 0.05 was considered significant.

Results

The baseline characteristics of the study patients according to the four groups are shown in Table 1. Their median age was 63 years, most were male, most were admitted due to medical conditions, and the largest group of patients were admitted due to cardiovascular disease. The most prevalent pre-existing disorder was diabetes, followed by chronic lung disease. Overall, just above half received mechanical ventilation and just below half

Table 1 Baseline characteristics of included patients

	Overall population (n = 2313)	CAM-ICU positive (n = 578)		CAM-ICU negative (n = 1735)		P value
		NLP-Dx-BD positive (n = 539)	NLP-Dx-BD negative (n = 39)	NLP-Dx-BD positive (n = 707)	NLP-Dx-BD negative (n = 1028)	
Age, years	63.2 (50.5–73.7)	66.1 (52.3–76.1)	70.2 (62.2–77.9)	63 (51.7–74)	62 (48.7–72.2)	<0.001
Male gender—no. (%)	1390 (60.2)	335 (62.2)	21 (53.8)	433 (61.3)	601 (58.6)	0.384
Body mass index, kg/m ²	27.6 (24.3–32.7)	27.5 (24.8–34)	26.5 (21.6–30.5)	27.1 (22.3–32.1)	28.2 (24.6–33.2)	0.180
APACHE III	52 (38–67)	63 (49–78)	56 (46.5–67)	54 (40–69)	44 (33.0–58)	<0.001
ANZROD	0 (0–0.1)	0.08 (0.03–0.21)	0.04 (0.02–0.08)	0.04 (0.01–0.12)	0.02 (0.01–0.06)	<0.001
Type of admission—no. (%)						<0.001
Medical	1334 (57.7)	361 (67)	18 (46.2)	426 (60.3)	529 (51.5)	
Elective surgery	487 (21.1)	59 (10.9)	7 (17.9)	138 (19.5)	283 (27.5)	
Urgency surgery	492 (21.3)	119 (22.1)	14 (35.9)	143 (20.2)	216 (21)	
Planned admission—no. (%)	574 (24.8)	88 (16.3)	10 (25.6)	160 (22.6)	316 (30.7)	<0.001
MET call admission—no. (%)	437 (18.9)	115 (21.3)	4 (10.3)	135 (19.1)	183 (17.8)	0.197
Cardiac arrest—no. (%)	40 (1.7)	17 (3.2)	2 (5.1)	10 (1.4)	11 (1.1)	0.008
Acute respiratory failure—no. (%)	55 (2.4)	32 (6)	0 (0)	12 (1.7)	11 (1.1)	<0.001
Admission diagnosis—no. (%)						<0.001
Cardiovascular	669 (28.9)	118 (21.9)	13 (33.3)	186 (26.3)	352 (34.2)	
Gastrointestinal	379 (16.4)	115 (21.3)	6 (15.4)	104 (14.7)	154 (15)	
Respiratory	358 (15.5)	67 (12.4)	3 (7.7)	125 (17.7)	163 (15.9)	
Sepsis	278 (12)	67 (12.4)	5 (12.8)	93 (13.2)	113 (11)	
Neurological	199 (8.6)	73 (13.5)	3 (7.7)	78 (11)	45 (4.4)	
Metabolic	152 (6.6)	37 (6.9)	3 (7.7)	50 (7.1)	62 (6)	
Trauma	80 (3.5)	29 (5.4)	4 (10.3)	25 (3.5)	22 (2.1)	
Renal and genitourinary	116 (5)	21 (3.9)	2 (5.1)	32 (4.5)	61 (5.9)	
Musculoskeletal	45 (1.9)	9 (1.7)	0 (0)	7 (1)	29 (2.8)	
Hematological	29 (1.3)	3 (0.6)	0 (0)	5 (0.7)	21 (2)	
Gynecological	8 (0.3)	0 (0)	0 (0)	2 (0.3)	6 (0.6)	
ICU source of admission—no. (%)						<0.001
Operating room	969 (41.9)	176 (32.7)	21 (53.8)	280 (39.6)	492 (47.9)	
Emergency department	663 (28.7)	160 (29.7)	12 (30.8)	218 (30.8)	273 (26.6)	
Ward	438 (18.9)	114 (21.2)	3 (7.7)	141 (19.9)	180 (17.5)	
Other hospital (not ICU)	189 (8.2)	58 (10.8)	3 (7.7)	53 (7.5)	75 (7.3)	
Other hospital ICU	47 (2)	29 (5.4)	0 (0)	12 (1.7)	6 (0.6)	
ICU from the same hospital	5 (0.2)	1 (0.2)	0 (0)	2 (0.3)	2 (0.2)	
Other	2 (0.1)	1 (0.2)	0 (0)	1 (0.1)	0 (0)	
Co-existing disorders—no. (%)						
Diabetes	485 (21)	123 (22.8)	9 (23.1)	136 (19.2)	217 (21.1)	0.457
Chronic lung disease	300 (13)	60 (11.1)	5 (12.8)	106 (15)	129 (12.5)	0.229
Chronic kidney disease	287 (12.4)	73 (13.5)	3 (7.7)	100 (14.1)	111 (10.8)	0.126
Immunosuppression	229 (9.9)	51 (9.5)	3 (7.7)	73 (10.3)	102 (9.9)	0.952
Cirrhosis	207 (8.9)	80 (14.8)	1 (2.6)	61 (8.6)	65 (6.3)	<0.001
Chronic cardiovascular disease	115 (5)	21 (3.9)	4 (10.3)	42 (5.9)	48 (4.7)	0.133
Metastatic cancer	98 (4.2)	12 (2.2)	1 (2.6)	33 (4.7)	52 (5.1)	0.039
Leukemia	62 (2.7)	11 (2)	1 (2.6)	19 (2.7)	31 (3)	0.705
Lymphoma	25 (1.1)	5 (0.9)	0 (0)	10 (1.4)	10 (1)	0.779
Chronic immune disease	22 (1)	7 (1.3)	0 (0)	8 (1.1)	7 (0.7)	0.591
Liver failure	24 (1)	8 (1.5)	0 (0)	8 (1.1)	8 (0.8)	0.565

Table 1 (continued)

	Overall population (n = 2313)	CAM-ICU positive (n = 578)		CAM-ICU negative (n = 1735)		P value
		NLP-Dx-BD positive (n = 539)	NLP-Dx-BD negative (n = 39)	NLP-Dx-BD positive (n = 707)	NLP-Dx-BD negative (n = 1028)	
Organ support—no. (%)						
ECMO	7 (0.3)	3 (0.6)	0 (0)	4 (0.6)	0 (0)	0.058
Vasopressor or inotropes	1175 (51.2)	372 (69.5)	22 (56.4)	365 (52)	416 (40.8)	< 0.001
Invasive ventilation	1123 (48.9)	371 (69.2)	22 (56.4)	356 (50.7)	374 (36.7)	< 0.001
Non-invasive ventilation	173 (7.5)	46 (8.6)	4 (10.3)	61 (8.7)	62 (6.1)	0.095
Renal replacement therapy	170 (7.4)	101 (18.9)	2 (5.1)	41 (5.8)	26 (2.6)	< 0.001
Laboratory tests						
pH	7.39 (7.33–7.43)	7.39 (7.33–7.43)	7.38 (7.33–7.42)	7.39 (7.35–7.44)	7.39 (7.36–7.44)	0.064
PaO ₂ /FiO ₂	304 (212–376)	281 (172–360)	329 (241–397)	286 (198–367)	319 (243–395)	< 0.001
PaCO ₂ , mmHg	39 (35–44)	40 (35–45)	40.5 (35.2–46.5)	39 (35–44)	39 (35–44)	0.267
Lactate, mmol/L	2.1 (1.5–3.2)	2.5 (1.8–3.9)	2.2 (1.6–3.4)	2.2 (1.6–3.3)	2.0 (1.4–2.9)	< 0.001
Highest creatinine, μmol/L	92 (69–145)	108 (71–175)	95 (70–160)	95 (69–149)	87 (68–129)	< 0.001
Lowest platelet, × 10 ⁹ /L	175 (122–245)	171 (100–230)	161 (115–305)	176 (122–252)	179 (128–249)	0.007
Vital signs						
Lowest MAP, mmHg	65 (59–71)	65 (59–70)	63 (59–70)	63 (58–71)	66 (60–73)	< 0.001
Highest RR, breaths/min	21 (18–25)	20 (18–25)	20 (17–25)	22 (18–26.2)	21 (18–25)	0.394
Highest temperature, °C	37.2 (36.8–37.7)	37.3 (36.8–37.8)	37 (36.5–37.5)	37.3 (36.8–37.7)	37.2 (36.8–37.6)	0.090
Urine output, mL	1525 (1075–2130)	1415 (955–2045)	1585 (1120–2333)	1550 (1122–2205)	1550 (1100–2122)	0.001

Data are median (IQR) or N (%)

APACHE Acute Physiology and Chronic Health Evaluation, MET medical emergency team, ICU intensive care unit, ECMO extracorporeal membrane oxygenation, MAP mean arterial pressure, RR respiratory rate, NLP-Dx-BD natural language processing diagnosed behavioural disturbance, CAM-ICU Confusion Assessment Method for Intensive Care Unit

received vasopressor/inotropic drugs. However, irrespective of CAM-ICU status, the presence of NLP-Dx-BD positivity identified patients with greater illness severity. The mean RASS during ICU admission was -0.5 (-1 to -0.2), and was similar between the groups ($P=0.238$) (eFigure 1 in Online Supplement).

Characteristics of CAM-ICU delirium and NLP behavioral disturbance diagnosis

The time to first CAM-ICU positivity or to first NLP-Dx-BD assessment is shown in eFigure 2 in the Online Supplement. The frequency of words used to describe NLP-Dx-BD is shown in eTable 3 and, as word cloud, in eFigure 3 in the Online Supplement. Overall, 32% of the patients with NLP-Dx-BD were identified on day 0 (day of admission), but only 11% of CAM-ICU positive patients were identified on this day. In addition, the majority of patients positive for NLP-Dx-BD (37%) and of CAM-ICU (29%) were identified on day 1.

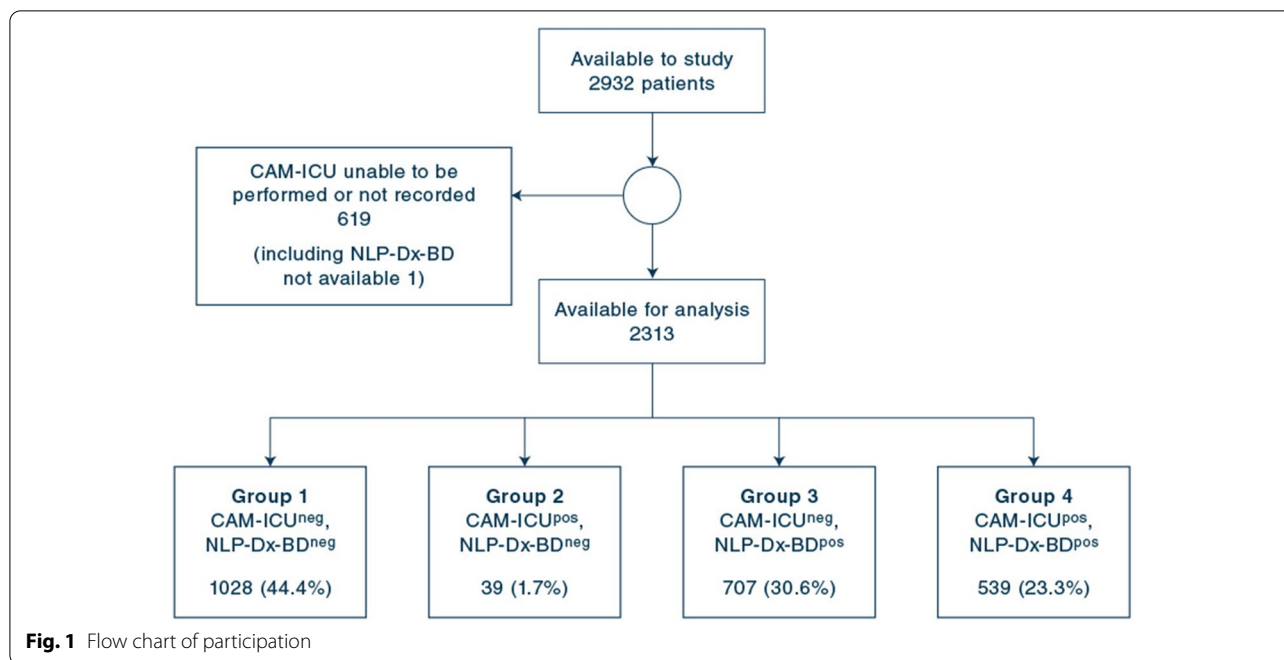
Among NLP-Dx-BD-positive patients, the median number of notes with BD positive words was 3 (1–8) and the median cumulative number of BD positive words was 4 (2–13) per patient. Among CAM-ICU positive patients, the median number of notes that were positive

for NLP-Dx-BD was 7 (3–18) and the median cumulative number of words was 11 (4–30). The median number of ICU-shift notes per patient where CAM-ICU was reported was 3 (2–7). Among CAM-ICU positive patients a median of 2 (1–4) notes were positive. The number of notes per patient with CAM-ICU reported as “unable to be done” was 0 (0–2).

Classification according to NLP-Dx-BD and CAM-ICU

We studied 2932 patients. Among these, NLP-Dx-BD or CAM-ICU assessments were not available and reported as “unable to be performed” or missing in 1 (0%) and in 619 (21.1%) patients, respectively, leaving 2313 patients with complete data for analysis. Of these, 1246 (53.9%) were NLP-Dx-BD positive and 578 (25%) were CAM-ICU positive. Among NLP-Dx-BD positive patients 539 (43.3%) were CAM-ICU positive, while among CAM-ICU positive patients 539 (93.3%) were NLP-Dx-BD positive.

When assessing the four key groups, 1028 (44.4%) were categorized in Group 1; 39 (1.7%) in Group 2; 707 (30.6%) in Group 3; and 539 (23.3%) in Group 4. The distribution of patients into these groups shows that NLP-Dx-BD identified more patients than CAM-ICU (Fig. 1). It also



demonstrates that, overall, only 44% of patients were negative for both assessments during their ICU stay.

We additionally assessed categorization according to different levels of conscious state from coma to agitation using the lowest and highest RASS score (eFigure 4 and eFigure 5, Online Supplement) to illustrate the impact of such states during each shift on the classification of patients according to NLP-Dx-BD and CAM-ICU status. NLP identified more patients than CAM-ICU when the shift RASS score was in the awake range and when the highest shift RASS score was in the coma or agitated range.

Primary outcome: use of antipsychotic medications

The use of antipsychotic medications was highest in patients in Group 4 (24.3%) (Table 2 and Fig. 2). The group with

the second greatest use of these medications was Group 3 (10.5%). For all antipsychotic medications, the rate of use in Group 2 patients was 5.1%, similar to that of Group 1 at 2.3%. (Table 2). The impact of NLP-Dx-BD on the use of antipsychotic medications was confirmed by univariate and multivariate Cox models that treated groups as time-dependent variables (eTable 4 and eTable 5).

By day 3, most of such treatment had been given, with the exception of Group 4 patients who continued to accrue further treatment up to one week after ICU admission (Fig. 3).

Secondary clinical outcomes

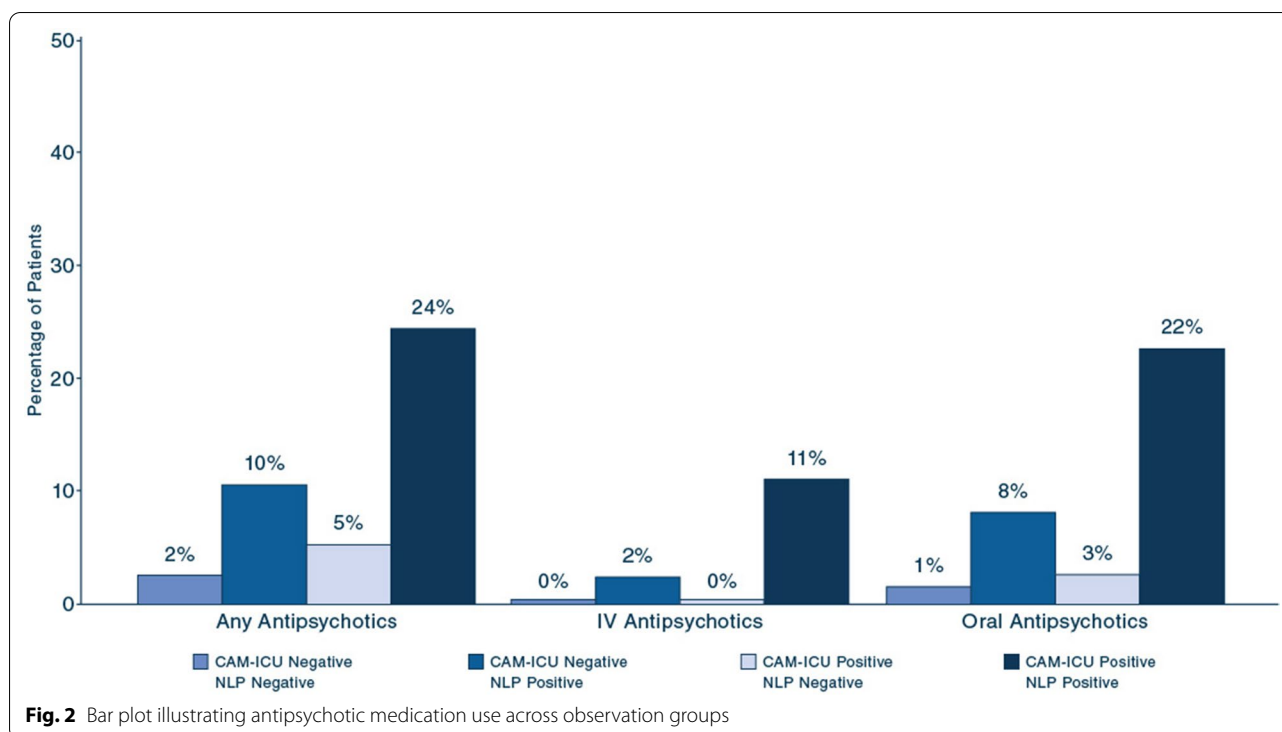
Clinical outcomes of study patients are shown in Table 3. Overall hospital mortality was 6.3%. Mortality was

Table 2 Use of antipsychotic medications (APM) in the study population

	Overall population (n = 2313)	CAM-ICU positive (n = 578)		CAM-ICU negative (n = 1735)		P value
		NLP-Dx-BD positive (n = 539)	NLP-Dx-BD negative (n = 39)	NLP-Dx-BD positive (n = 707)	NLP-Dx-BD negative (n = 1028)	
Any APM	231 (10)	131 (24.3)	2 (5.1)	74 (10.5)	24 (2.3)	<0.001
Haloperidol	76 (3.3)	59 (10.9)	0 (0)	16 (2.3)	1 (0.1)	<0.001
Olanzapine	40 (1.7)	27 (5)	0 (0)	10 (1.4)	3 (0.3)	<0.001
Quetiapine	170 (7.3)	111 (20.6)	1 (2.6)	48 (6.8)	10 (1)	<0.001
Risperidone	4 (0.2)	0 (0)	0 (0)	3 (0.4)	1 (0.1)	0.262

Data are median (IQR) or N (%)

APM antipsychotic medication, NLP natural language processing, CAM-ICU Confusion Assessment Method for ICU, RASS Richmond agitation sedation scale, ICU intensive care unit; CAM-ICU Confusion Assessment Method for Intensive Care Unit



highest in Group 4 (10.8%), lower but similar in Group 2 and Group 3 (7.7% and 7.5%, respectively), and lowest in Group 1. Duration of ICU and hospital stay was significantly longer in the presence of NLP-Dx-BD positivity, irrespective of CAM-ICU status.

On univariable analysis, both Group 4 (OR, 3.77 [95% CI 2.43–5.95]; $P < 0.001$) and Group 3 (OR, 2.53 [95% CI, 1.62–4.00]; $P < 0.001$) were associated with increased risk for hospital mortality (eTable 6 in Online Supplement). On multivariable analysis, only Group 3 (OR 1.69 [95% CI 1.05–2.76]; $P = 0.03$) was independently associated with increased risk for hospital mortality (eTable 7). On univariate modelling that treated groups as time-dependent variables both NLP-Dx-BD and CAM-ICU status were associated with mortality but, on multivariable modelling, no overall effect was found (eTable 8 and eTable 9).

Discussion

Key findings

In a cohort of more than two thousand ICU patients, we compared the prevalence, characteristics, treatment with antipsychotic medications, and outcomes of patients with NLP-diagnosed behavioral disturbance (NLP-Dx-BD) and patients with Confusion Assessment Method for ICU (CAM-ICU) positivity. We found that more than half were NLP-Dx-BD positive and a quarter were CAM-ICU positive. Among NLP-Dx-BD positive patients, four out of ten were CAM-ICU positive. In contrast, among

CAM-ICU positive patients, nine out of ten were NLP-Dx-BD positive. NLP-Dx-BD identified significantly more patients likely to receive antipsychotic medications and, in the absence of NLP-Dx-BD, treatment with antipsychotic medications was uncommon. Finally, regardless of CAM-ICU status, NLP-Dx-BD was associated with significantly longer duration of ICU and hospital stay and greater hospital mortality (Table 4).

Relationship to previous studies

Dependent on risk factors and delirium phenotype, the prevalence of CAM-ICU positive patients has been reported between 10 and 89% [22–25]. Our prevalence of 25% falls within this range. CAM-ICU positive patients appear to have a hospital mortality rate of between 10.7 and 27% [26–28]. Our hospital mortality rate of 10.6% falls just below this range. Further, our finding that NLP-Dx-BD identified more patients with behavioral and/or cognitive disturbance earlier in the critical care episode than CAM-ICU is also consistent with previous findings that, in routine practice, the use of CAM-ICU may delay the detection of delirium [29]. Our observations that the prevalence of behavioral disturbance as detected by CAM-ICU was lower than with NLP-Dx-BD are also similar to findings from previous studies where CAM-ICU was found to detect lower rates of delirium than unstructured assessments [29, 30].

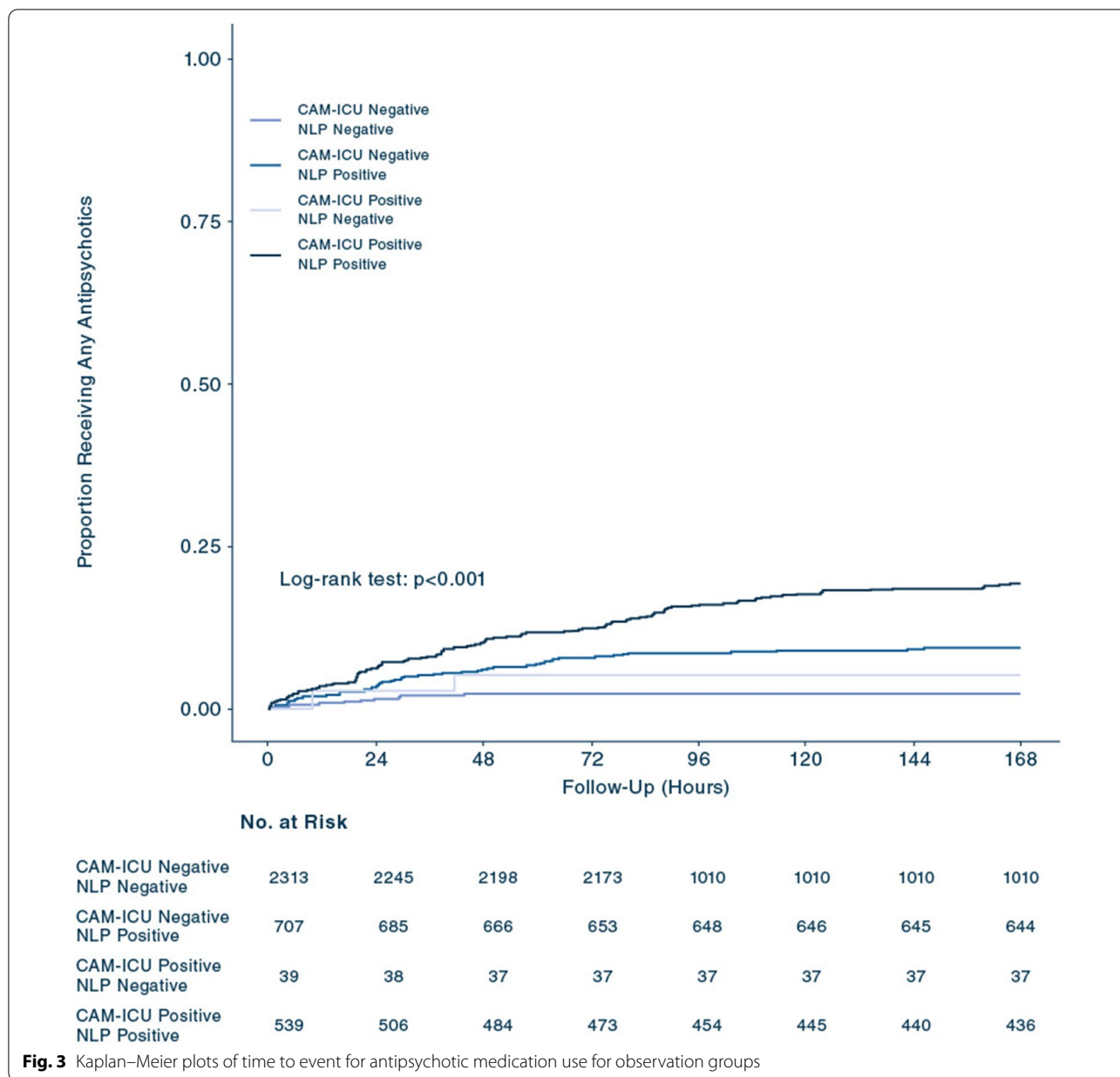


Fig. 3 Kaplan–Meier plots of time to event for antipsychotic medication use for observation groups

CAM-ICU was not obtained in one in five patients. This is aligned with work by Terry et al. who found a missed documentation rate of more than half of available opportunities in a pair of American medical and surgical ICUs [31] and with data from Vanderbilt University Medical Centre showing a 16% non-compliance for CAM-ICU assessment in an institution highly dedicated to CAM-ICU assessment [32]. Moreover, Kanova et al. reported in a prospective study that 14% of patients could not be assessed for CAM-ICU due to prolonged coma [33]. Our observations indicate that, during a given shift, and in the presence of RASS defined coma, CAM-ICU

assessment was missing in most patients. However, this was also true to a similar extent for NLP-Dx-BD words and coma was an uncommon state in our cohort.

Finally, our observations of the prevalence, characteristics, and outcome of NLP-Dx-BD are aligned with those found in a much larger population of more than 12,000 patients recently investigated by our group [14]. Moreover, the words used to define this condition are supported by our survey of the terms that clinicians would use to describe the presence of an acute behavioral disturbance [17].

Table 3 Clinical Outcomes of Included Patients

	Overall population (<i>n</i> = 2313)	CAM-ICU positive (<i>n</i> = 578)		CAM-ICU negative (<i>n</i> = 1735)		<i>P</i> value
		NLP positive (<i>n</i> = 539)	NLP negative (<i>n</i> = 39)	NLP positive (<i>n</i> = 707)	NLP negative (<i>n</i> = 1028)	
Duration of ventilation, days ^a	0 (0–7.4)	0 (0–41.3)	0 (0–14.2)	0 (0–8.4)	0 (0–2.4)	< 0.001
ICU length of stay, days	2.1 (1.2–4.3)	5.6 (2.7–10.7)	1.7 (0.8–2.7)	2.7 (1.6–4.6)	1.4 (0.8–2.3)	< 0.001
Hospital length of stay, days	10.4 (6.1–19.7)	17.5 (10.2–29.9)	8.8 (4.9–15)	10.4 (6.8–18.6)	8.2 (5.1–14.1)	< 0.001
ICU mortality—no. (%)	90 (3.9)	34 (6.3)	3 (7.7)	35 (5)	18 (1.8)	< 0.001
Hospital mortality—no. (%)	146 (6.3)	58 (10.8)	3 (7.7)	53 (7.5)	32 (3.1)	< 0.001
28-day mortality—no. (%)	119 (5.2)	42 (7.9)	3 (7.7)	45 (6.4)	29 (2.8)	< 0.001

Data are median (IQR) or *N* (%)

ICU intensive care unit, NLP natural language processing, CAM-ICU Confusion Assessment Method for Intensive Care Unit

^a Among patients who received invasive ventilation

Table 4 Univariable and multivariable models with hospital mortality as outcome

Group	Univariable Model		Multivariable Model	
	Odds ratio (95% CI)	<i>P</i> value	Odds ratio (95% CI)	<i>P</i> value
CAM-ICU ^{neg} and NLP-Dx-BD ^{neg}	1 (Reference)		1 (Reference)	
CAM-ICU ^{neg} and NLP-Dx-BD ^{pos}	2.53 (1.62–4)	< 0.001	1.69 (1.05–2.76)	0.032
CAM-ICU ^{pos} and NLP-Dx-BD ^{neg}	2.59 (0.6–7.69)	0.129	1.22 (0.18–4.78)	0.798
CAM-ICU ^{pos} and NLP-Dx-BD ^{pos}	3.77 (2.43–5.95)	< 0.001	1.59 (0.98–2.6)	0.060

NLP natural language processing, CAM-ICU Confusion Assessment Method for Intensive Care Unit

Implications of study findings

Our findings imply that NLP-Dx-BD may identify more patients with abnormal behavioral and disturbed cognitive state than CAM-ICU. Moreover, they imply that NLP-Dx-BD positive patients are significantly more likely to receive antipsychotic medications than CAM-ICU positive patients. Furthermore, they suggest that NLP-Dx-BD positive patients are more severely ill, more likely to stay in ICU and hospital for longer, and more likely to die. This is consistent with a previous study that found that agitated behavior, even without diagnosed delirium, correlated with poorer outcomes [34]. It is also consistent with clinical observations that the level of sedation and behavioral symptoms often fluctuate over an 8-h or 12-h shift. Thus, NLP-Dx-BD will capture behavioral screening features over the whole period but CAM-ICU, which represents a 'snapshot' at a single point of time during the shift, may not. Finally, our findings suggest that NLP may be particularly more sensitive during shifts where a given RASS score has identified the presence of either deep sedation or agitation.

Strengths and limitations

Our study has several strengths. It studied a novel approach to the assessment of the epidemiology of behavioral disturbance in critically ill patients. It compared such assessment with delirium assessment performed by the CAM-ICU methodology. It also conducted such comparisons in a large and heterogeneous group of ICU patients, thus increasing the external validity of the observations. The incidence of CAM-ICU positive assessment was consistent with the literature and our study applied a primary outcome based on therapy, which is likely to be relevant to clinicians as well as patients. Finally, by showing that NLP-Dx-BD positivity captured almost all of patients with CAM-ICU positive status, it provided additional indirect evidence of the validity of its construct.

We acknowledge several limitations. Our study was undertaken in a large tertiary intensive care unit system in a university affiliated hospital of a resource-rich country. Therefore, its findings may not apply to other intensive care units in low or middle-income countries.

Multiple progress notes were missing a positive or negative CAM-ICU assessment. However, this may reflect the well-described operational challenges in applying CAM-ICU and had minimal material impact on our findings. We did not provide data on non-pharmacological interventions including family visits and environmental management. However, many were applied by unit policy and the efficacy of such interventions remains uncertain as demonstrated in a recent stepped wedge cluster controlled trial [35]. Patients were not assessed for the presence of delirium by an independent, psychiatrically trained clinician. However, we were studying the relationship between patient populations identified through the use of alternative screening methodologies. Thus, we were not attempting to confirm that the patients identified by such screening would then be diagnosed with delirium by such clinicians. Therefore, the changes in behavior may have reflected drug withdrawal, exacerbation of dementia or of pre-existing psychiatric disorders. Care givers generating the progress notes may have been aware of CAM-ICU results at the time of writing their notes; however, it is impossible to determine if an NLP-Dx-BD positive note preceded or followed a CAM-ICU assessment. Further, significantly more patients were NLP-Dx-BD positive than CAM-ICU positive, thus demonstrating that NLP-Dx-BD positivity was often logically determined independently of CAM-ICU assessment. Clinical progress notes may include errors in annotation that incorrectly characterize a patient's behavior [36, 37]. However, our study analyzed multiple progress notes (nursing, medical, allied health) that documented the same shift making systematic errors unlikely. There may also have been errors in the records of administration of anti-psychotic medication. However, our ICU uses an audited medication management system that accounts for the prescription, retrieval and administration of medications, which would minimize such errors. Finally, we did not assess long term cognitive outcomes. Thus we cannot make any statement about the long term sequelae of NLP-Dx-BD.

Conclusion

In conclusion, Natural Language Processing of electronic caregiver notes and the Confusion Assessment Method for ICU (CAM-ICU) describe partly overlapping populations. However, NLP-Dx-BD identified more patients with behavioral and/or cognitive disturbance than CAM-ICU, while identifying more than 90% of CAM-ICU positive patients as having such a disturbance. Moreover, NLP-Dx-BD identified significantly more patients who went on to receive treatment with antipsychotic medications. In contrast, in the absence of NLP-Dx-BD, very few CAM-ICU positive patients received such treatment.

Further, irrespective of CAM-ICU status, NLP-Dx-BD also identified patients with longer duration of ICU and hospital stay and greater hospital mortality. These observations suggest that NLP-Dx-BD may be a novel and useful epidemiologic screening tool in ICU patients.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s00134-022-06650-z>.

Author details

¹ Data Analytics Research and Evaluation (DARE) Centre, Austin Health and The University of Melbourne, Heidelberg, VIC, Australia. ² School of Computing and Information Systems, The University of Melbourne, Parkville, Melbourne, VIC, Australia. ³ Australian and New Zealand Intensive Care Research Centre, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia. ⁴ Department of Intensive Care, Austin Hospital, 145 Studley Rd, Heidelberg, VIC, Australia. ⁵ Department of Critical Care, School of Medicine, The University of Melbourne, Parkville, Melbourne, VIC, Australia.

Acknowledgements

The authors wish to acknowledge the contribution of Michael Reade, Raghaven Murugan, Antonio Paulo Nassar Junior, Yahya Shehabi, Lakhmir Chawla and Michele Gaca for their review of this manuscript and subsequent guidance.

Funding

Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Conflicts of interest

The authors declare they have no conflict of interest.

Open Access

This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 27 September 2021 Accepted: 15 February 2022

Published: 23 March 2022

References

- van den Boogaard M, Schoonhoven L, van der Hoeven JG et al (2012) Incidence and short-term consequences of delirium in critically ill patients: a prospective observational cohort study. *Int J Nurs Stud* 49:775–783. <https://doi.org/10.1016/j.ijnurstu.2011.11.016>
- Salluh JF, Wang H, Schneider EB et al (2015) Outcome of delirium in critically ill patients: systematic review and meta-analysis. *BMJ* 350:h2538–h2538. <https://doi.org/10.1136/bmj.h2538>
- Cole MG, Ciampi A, Belzile E, Zhong L (2008) Persistent delirium in older hospital patients: a systematic review of frequency and prognosis. *Age Ageing* 38:19–26. <https://doi.org/10.1093/ageing/afn253>

4. American Psychiatric Association (1998) Diagnostic and statistical manual of mental disorders: DSM-IV; includes ICD-9-CM codes effective 1. Oct. 96, 4. ed., 7. print. Washington, DC
5. Association AP, Association AP (2013) Diagnostic and statistical manual of mental disorders: DSM-5, 5th edn. American Psychiatric Association, Washington, D.C.
6. Adamis D, Rooney S, Meagher D et al (2015) A comparison of delirium diagnosis in elderly medical inpatients using the CAM, DRS-R98, DSM-IV and DSM-5 criteria. *Int Psychogeriatr* 27:883–889. <https://doi.org/10.1017/S1041610214002853>
7. Meagher DJ, Morandi A, Inouye SK et al (2014) Concordance between DSM-IV and DSM-5 criteria for delirium diagnosis in a pooled database of 768 prospectively evaluated patients using the delirium rating scale-revised-98. *BMC Med* 12:164. <https://doi.org/10.1186/s12916-014-0164-8>
8. Kotfis K, Marra A, Ely EW (2018) ICU delirium—a diagnostic and therapeutic challenge in the intensive care unit. 50:13
9. Neufeld KJ, Nelliott A, Inouye SK et al (2014) Delirium diagnosis methodology used in research: a survey-based study. *Am J Geriatr Psychiatry* 22:1513–1521. <https://doi.org/10.1016/j.jagp.2014.03.003>
10. Ely EW, Inouye SK, Bernard GR et al (2001) Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *JAMA* 286:2703. <https://doi.org/10.1001/jama.286.21.2703>
11. Inouye SK (1990) Clarifying confusion: the confusion assessment method: a new method for detection of delirium. *Ann Intern Med* 113:941. <https://doi.org/10.7326/0003-4819-113-12-941>
12. Bergeron N, Dubois M-J, Dumont M et al (2001) Intensive care delirium screening checklist: evaluation of a new screening tool. *Intensive Care Med* 27:859–864. <https://doi.org/10.1007/s001340100909>
13. Barr J, Fraser GL, Puntillo K et al (2013) Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. *Crit Care Med* 41:263–306. <https://doi.org/10.1097/CCM.0b013e3182783b72>
14. Young M, Holmes N, Robbins R et al (2021) Natural language processing to assess the epidemiology of delirium-suggestive behavioural disturbances in critically ill patients. *Crit Care Resusc* 23:145–154
15. Stow PJ, Hart GK, Higlett T et al (2006) Development and implementation of a high-quality clinical database: the Australian and New Zealand Intensive Care Society Adult Patient Database. *J Crit Care* 21:133–141. <https://doi.org/10.1016/j.jcrrc.2005.11.010>
16. Bird S, Klein E, Loper E (2009) Natural language processing with Python. O'Reilly, Newton
17. Holmes NE, Amjad S, Young M et al (2019) Using language descriptors to recognise delirium: a survey of clinicians and medical coders to identify delirium-suggestive words. *Crit Care Resusc* 21:4
18. Nikoio R, Neufeld KJ, Oh ES et al (2019) Antipsychotics for treating delirium in hospitalized adults: a systematic review. *Ann Intern Med* 171:485. <https://doi.org/10.7326/M19-1860>
19. Paul E, Bailey M, Pilcher D (2013) Risk prediction of hospital mortality for adult patients admitted to Australian and New Zealand intensive care units: development and validation of the Australian and New Zealand Risk of Death model. *J Crit Care* 28:935–941. <https://doi.org/10.1016/j.jcrrc.2013.07.058>
20. Eldho P, Bailey M, Kasza J, Pilcher D (2016) The ANZROD model: better benchmarking of ICU outcomes and detection of outliers. *Crit Care Resusc* 18:12
21. R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
22. Devlin JW, Fraser GL, Joffe AM et al (2013) The accurate recognition of delirium in the ICU: the emperor's new clothes? *Intensive Care Med* 39:2196–2199. <https://doi.org/10.1007/s00134-013-3105-7>
23. Agarwal V, Oneill PJ, Cotton BA et al (2010) Prevalence and risk factors for development of delirium in burn intensive care unit patients. *J Burn Care Res* 31:706–715. <https://doi.org/10.1097/BCR.0b013e3181eebee9>
24. Guenther U, Popp J, Koecher L et al (2010) Validity and Reliability of the CAM-ICU Flowsheet to diagnose delirium in surgical ICU patients. *J Crit Care* 25:144–151. <https://doi.org/10.1016/j.jcrrc.2009.08.005>
25. Haenggi M, Blum S, Brechbuehl R et al (2013) Effect of sedation level on the prevalence of delirium when assessed with CAM-ICU and ICDSC. *Intensive Care Med* 39:2171–2179. <https://doi.org/10.1007/s00134-013-3034-5>
26. Tomasi CD, Grandi C, Salluh J et al (2012) Comparison of CAM-ICU and ICDSC for the detection of delirium in critically ill patients focusing on relevant clinical outcomes. *J Crit Care* 27:212–217. <https://doi.org/10.1016/j.jcrrc.2011.05.015>
27. Yamamoto T, Mizobata Y, Kawazoe Y et al (2020) Incidence, risk factors, and outcomes for sepsis-associated delirium in patients with mechanical ventilation: a sub-analysis of a multicenter randomized controlled trial. *J Crit Care* 56:140–144. <https://doi.org/10.1016/j.jcrrc.2019.12.018>
28. Pauley E, Lishmanov A, Schumann S et al (2015) Delirium is a robust predictor of morbidity and mortality among critically ill patients treated in the cardiac intensive care unit. *Am Heart J* 170:79–86.e1. <https://doi.org/10.1016/j.ahj.2015.04.013>
29. van Eijk MM, van den Boogaard M, van Marum RJ et al (2011) Routine use of the confusion assessment method for the intensive care unit: a multicenter study. *Am J Respir Crit Care Med* 184:340–344. <https://doi.org/10.1164/rccm.201101-0065OC>
30. Reade MC, Eastwood GM, Peck L (2011) Routine use of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) by bedside nurses may underdiagnose delirium. *Crit Care Resusc* 13:9
31. Terry KJ, Anger KE, Szumita PM (2015) Prospective evaluation of inappropriate unable-to-assess CAM-ICU documentations of critically ill adult patients. *J Intensive Care* 3:52. <https://doi.org/10.1186/s40560-015-0119-y>
32. Soja SL, Pandharipande PP, Fleming SB et al (2008) Implementation, reliability testing, and compliance monitoring of the Confusion Assessment Method for the Intensive Care Unit in trauma patients. *Intensive Care Med* 34:1263–1268. <https://doi.org/10.1007/s00134-008-1031-x>
33. Kanova M, Sklienka P, Roman K et al (2017) Incidence and risk factors for delirium development in ICU patients—a prospective observational study. *Biomed Pap Med Fac Univ Palacky Olomouc Czechoslov* 161:187–196. <https://doi.org/10.5507/bp.2017.004>
34. Marquis F, Ouimet S, Riker R et al (2007) Individual delirium symptoms: Do they matter?*. *Crit Care Med* 35:2533–2537. <https://doi.org/10.1097/01.CCM.0000284506.43390.F3>
35. Rood PJT, Zegers M, Ramnarain D et al (2021) The Impact Of Nursing Delirium Preventive Interventions in the ICU: a multicenter cluster-randomized controlled clinical trial. *Am J Respir Crit Care Med* 204:682–691. <https://doi.org/10.1164/rccm.202101-0082OC>
36. Ratnapalan S, Brown K, Cieslak P et al (2012) Charting errors in a teaching hospital. *Pediatr Emerg Care* 28:268–271. <https://doi.org/10.1097/PEC.0b013e3182494fb8>
37. Carroll AE, Tarczy-Hornoch P, O'Reilly E, Christakis DA (2003) Resident documentation discrepancies in a neonatal intensive care unit. *Pediatrics* 111:976–980. <https://doi.org/10.1542/peds.111.5.976>