**RESEARCH ARTICLE**

# Translation and cultural adaptation of the Greek version of the confusion assessment method diagnostic algorithm and the nursing delirium screening scale and their inter-rater reliability: A prospective cohort study

Ntalouka MP[1], Bareka M[1], Brotis AG[2], Chalkias A[1], Stamoulis K[1], Flossos A[1], Tzimas P[3], Arnaoutoglou E[1]

[1]Department of Anesthesiology, Faculty of Medicine, School of Health Sciences, University of Thessaly, Larissa
[2]Department of Neurosurgery, University Hospital of Larissa
Thessaly
[3]Department of Anesthesiology and Postoperative Intensive Care, Faculty of Medicine, School of Health Sciences, University of Ioannina, Epirus
Greece

**Abstract**

**Aim:** The lack of standardized tools limits the diagnosis of postoperative delirium (POD) in the Greek population. Our aim was the translation and the cultural adaptation of the confusion assessment method (CAM) diagnostic algorithm and the nursing delirium screening scale (nu-DESC) in the Greek surgical population, and the determination of their inter-rater reliability.

**Methods:** After Ethical approval and registration as a clinical trial (NCT04154176), a prospective cohort study was conducted in the Department of Anesthesiology, University Hospital of Larissa, Greece. Patients at least 60 years old, undergoing elective non-cardiac surgery, under general anesthesia were included.

**Results:** Data from 60 patients, 180 records in total, were analyzed. There was an "almost perfect agreement" between the raters with the use of CAM (Cohen's Kappa estimate: 0.960; 95 % CI: 0.905-1.000) and nu-DESC (Cohen's Kappa estimate: 0.981; 95 % CI: 0.944-1.000). The agreement on each specific question of CAM and nu-DESC ranged from "substantial" to "almost perfect agreement". Based on the CAM, the sensitivity and specificity of nu-DESC were 0.97 (95 % CI: 0.82-1.00) and 0.99 (95 % CI: 0.96-1.00), respectively. The Greek versions of CAM and nu-DESC showed a high inter-rater agreement.

**Conclusion:** With the translation, the cultural adaptation, and the determination of their inter-rater agreement, the CAM diagnostic algorithm and the nu-DESC may serve as reliable instruments for the detection of POD in the Greek population. HIPPOKRATIA 2020, 24(1): 8-14.

**Keywords:** Delirium, diagnosis, neuropsychological tests, confusion assessment method, nursing delirium screening scale, anesthesia, general, reproducibility of results, humans, Greece

**Corresponding author:** Maria P. Ntalouka, MD, PhD, MScC, Anesthesiologist, Department of Anesthesiology, University Hospital of Larissa, C' Wing, 2nd Floor, Mezourlo, PO Box 1425, 41110, Larisa, Thessaly, Greece, tel: +306973688099, fax: +302431043858, e-mail: maria.ntalouka@icloud.com

## Introduction

Postoperative delirium (POD) is a prevalent complication associated with increased morbidity and mortality in patients undergoing surgery and/or anesthesia[1-4]. POD is defined as a state of acute and fluctuating cerebral dysfunction, hallmarked by disturbances in attention and cognition, not attributable to pre-existing cognitive dysfunction[1-3,5]. The prevalence of POD varies based on the type of surgery and the procedural risk. For lower-risk procedures, such as general surgery, the prevalence is <13 %, whereas, for higher-risk procedures, including major abdominal surgery, it rises up to 50 %[3,6].

As stated by the 2017 guidelines of the European Society of Anaesthesiology (ESA), all patients undergoing surgery should be screened for POD with standardized tools[1]. Over time, several instruments were developed to diagnose POD[7,8], including the Confusion Assessment Method (CAM) diagnostic algorithm[9] and the Nursing Delirium Screening Scale (nu-DESC)[10]. Inouye et al[9] initially described CAM in the early 1990s[7,9,11]. The CAM diagnostic algorithm is based on the four cardinal elements of postoperative delirium: i) acute onset and fluctuating course, ii) inattention, iii) disorganized thinking, and iv) altered level of consciousness. It can be administrated in less than five minutes by non-psychiatrist physicians[9,11]. For the diagnosis of POD, the first two and either of the latter two features are required[9,11]. Nu-DESC was developed in 2005 by Gaudreau et al[10]. It is a five items screen-

ing scale, and it can be administrated in approximately one minute[11]. It assesses disorientation, inappropriate behavior, inappropriate communication, hallucination, and psychomotor retardation[11]. Each item is rated on a three-point scale (0-2), and the total score varies from zero to ten. The cutoff value for POD is reported to be two[11]. Although it was initially developed to be administrated by nurses, based on a recent study, experts suggest that there is no significant difference between the evaluation results from the nursing staff and the physician, which makes the tool equally usable by both groups[12].

As far as the Greek population is concerned, neither CAM nor nu-DESC has been translated and adapted, and thus their diagnostic accuracy is unknown. Our aim was the translation and the cultural adaptation of the CAM diagnostic algorithm and the nu-DESC in the Greek population and the determination of their inter-rater reliability under the above-mentioned setting. Our secondary aim was to estimate the diagnostic accuracy of nu-DESC and the incidence of POD in our study sample.

## Materials and Methods

*Setting*

A prospectively designed reliability study was performed. Ethical approval was provided by the Scientific Board of University Hospital of Larissa, Greece (42627/04/10/2019). The study was conducted in accordance with the Ethical Principles for Medical Research Involving Human Subjects (Declaration of Helsinki) and was registered (NCT04154176)[13]. All patients were fully informed about the protocol and provided written informed consent before participation. The study conforms with the STROBE Statement reporting standards for cohort studies, the ISPOR guidelines for translation and cultural adaptation, and the GRRAS Statement reporting standards for reliability/agreement studies[14-16].

*Participants*

We included consecutive patients who underwent elective non-cardiac surgery at the General University Hospital of Larissa during November 2019. Our inclusion criteria were: i) adult patients over 60 years of age, ii) American society of anesthesiologists (ASA) physical status I to III[3,7], iii) undergoing elective non-cardiac surgery under general anesthesia, iv) native speakers of the Greek language, v) eligible to leave the post-anesthesia care unit, and vi) an expected in-hospital stay of at least 24 hours following surgery. We exclude patients from our study if they had any of the following: i) refused to participate or sign the informed consent form[13], ii) underwent surgery or anesthesia within the last 30 days[3,17], iii) prior or current history involving an affliction of the central nervous system[3,5,8,18], iv) severe hearing or visual impairment[3], v) psychiatric disorders[3,7], vi) a score less than five according to the Geriatric Depression Scale (GDS-15)[3,7,19], vii) a score less than four for females and less than two for males according to the Lawton-Brody Instrumental Activities of Daily Living Scale (I. A. D. L. )[7,17,20], viii) alcohol consumption more than 35 units/week[3,7], ix) drug dependence[3,7], x) previous neuropsychological testing[3,7], xi) hemodynamical instability[3,7], and xii) peri-procedural desaturation (one or more events of $SpO_2$ <80 % for more than two minutes)[3,7].

*Translation and Cultural Adaptation*

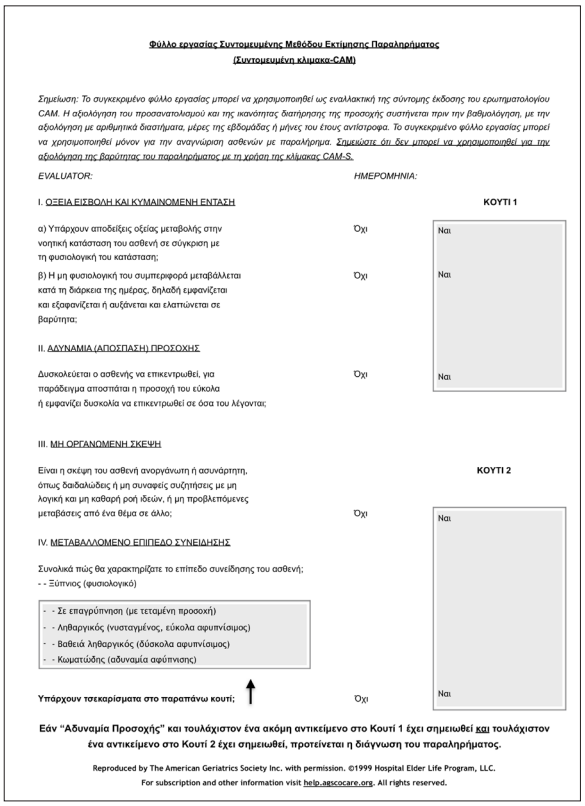
The translation process from English to Greek begun after gaining permission and consent from the authors/developers of both the original instruments[9,10]. Three authors performed three separate forward-translations, for each of the two instruments, from English to Greek. Subsequently, the translations were merged in one preliminary version[14]. Our next step was reconciliation. The Greek versions were adapted to the clinical setting without changing the meaning through a pilot study. Each phrase was extensively discussed between the translators. Different wording was tested throughout the process, and the final form was established after reaching a consensus between the two authors/developers[14]. A bilingual expert interpreter who was blinded to the original English version performed the back-translation into English[14]. The back translation review followed, and the back translation of the preliminary Greek version of both the instruments, was thoroughly compared with the original texts. After closed comparison and assessment, there were no substantial deviations from the original instruments[14]. The instrument developers reviewed the English back-translated version regarding the conformity of content and language and the agreement with the original versions. The appropriate modifications were made, and the latest step of the translation was approved (harmonization)[14]. The final Greek versions of CAM and nu-DESC underwent a structured evaluation by eight doctors from two different specialties (cognitive debriefing)[14]. The cognitive debriefing results were reviewed, and the translation was finalized[14]. The final translations of CAM (Figure 1) and nu-DESC (Figure 2) were proofread by three translators[14]. Four of the authors reviewed the process of the transaction (final report)[14].

*Study procedures*

Two experienced anesthesiologists, who have been trained according to the original training manuals of the CAM and nu-DESC instruments, evaluated the patients for the presence of POD[10,21]. The two raters were blinded and had no access to each other's evaluation. They independently administered the CAM and nu-DESC questionnaires during the first postoperative day after the discharge from the post-anesthesia care unit. The assessments of the two different rates were completed within 60 minutes. CAM was administrated within five minutes and nu-DESC within one in accordance with the international standards[11]. The patients who were diagnosed with POD received the appropriate treatment according to the hospital's protocol.

*Statistical analysis*

Descriptive statistics were used to summarize gen-



**Φύλλο εργασίας Συντομευμένης Μεθόδου Εκτίμησης Παραληρήματος**
**(Συντομευμένη κλίμακα-CAM)**

*Σημείωση: Το συγκεκριμένο φύλλο εργασίας μπορεί να χρησιμοποιηθεί ως εναλλακτική της σύντομης έκδοσης του ερωτηματολογίου CAM. Η αξιολόγηση του προσανατολισμού και της ικανότητας διατήρησης της προσοχής συστήνεται πριν την βαθμολόγηση, με την αξιολόγηση με αριθμητικά διαστήματα, μέρες της εβδομάδας ή μήνες του έτους αντίστροφα. Το συγκεκριμένο φύλλο εργασίας μπορεί να χρησιμοποιηθεί μόνον για την αναγνώριση ασθενών με παραλήρημα. Σημειώστε ότι δεν μπορεί να χρησιμοποιηθεί για την αξιολόγηση της βαρύτητας του παραληρήματος με τη χρήση της κλίμακας CAM-S.*

*EVALUATOR:* *ΗΜΕΡΟΜΗΝΙΑ:*

I. ΟΞΕΙΑ ΕΙΣΒΟΛΗ ΚΑΙ ΚΥΜΑΙΝΟΜΕΝΗ ΕΝΤΑΣΗ — **ΚΟΥΤΙ 1**

α) Υπάρχουν αποδείξεις οξείας μεταβολής στην νοητική κατάσταση του ασθενή σε σύγκριση με τη φυσιολογική του κατάσταση; — Όχι / Ναι

β) Η μη φυσιολογική του συμπεριφορά μεταβάλλεται κατά τη διάρκεια της ημέρας, δηλαδή εμφανίζεται και εξαφανίζεται ή αυξάνεται και ελαττώνεται σε βαρύτητα; — Όχι / Ναι

II. ΑΔΥΝΑΜΙΑ (ΑΠΟΣΠΑΣΗ) ΠΡΟΣΟΧΗΣ

Δυσκολεύεται ο ασθενής να επικεντρωθεί, για παράδειγμα αποσπάται η προσοχή του εύκολα ή εμφανίζει δυσκολία να επικεντρωθεί σε όσα του λέγονται; — Όχι / Ναι

III. ΜΗ ΟΡΓΑΝΩΜΕΝΗ ΣΚΕΨΗ — **ΚΟΥΤΙ 2**

Είναι η σκέψη του ασθενή ανοργάνωτη ή ασυνάρτητη, όπως διαδιλωδιες ή μη συναφείς συζητήσεις με μη λογική και μη καθαρή ροή ιδεών, ή μη προβλεπόμενες μεταβάσεις από ένα θέμα σε άλλο; — Όχι / Ναι

IV. ΜΕΤΑΒΑΛΛΟΜΕΝΟ ΕΠΙΠΕΔΟ ΣΥΝΕΙΔΗΣΗΣ

Συνολικά πώς θα χαρακτηρίζατε το επίπεδο συνείδησης του ασθενή;
- - Ξύπνιος (φυσιολογικό)
- - Σε επαγρύπνηση (με τεταμένη προσοχή)
- - Ληθαργικός (νυσταγμένος, εύκολα αφυπνίσιμος)
- - Βαθειά ληθαργικός (δύσκολα αφυπνίσιμος)
- - Κωματώδης (αδυναμία αφύπνισης)

**Υπάρχουν τσεκαρίσματα στο παραπάνω κουτί;** — Όχι / Ναι

**Εάν "Αδυναμία Προσοχής" και τουλάχιστον ένα ακόμη αντικείμενο στο Κουτί 1 έχει σημειωθεί και τουλάχιστον ένα αντικείμενο στο Κουτί 2 έχει σημειωθεί, προτείνεται η διάγνωση του παραληρήματος.**



**Figure 1:** The translated and adapted Greek version of the confusion assessment method (CAM) diagnostic algorithm.

**Nursing delirium screening scale (Nu-DESC)**

**Ημερομηνία:**

| Χαρακτηριστικά και περιγραφή | Βαθμολογία συμπτωμάτων 0-2 | | |
|---|---|---|---|
| **Συμπτώματα/χρονική διάρκεια** | **24-8.00** | **8.00-16.00** | **16.00-24** |
| Αποπροσανατολισμός<br>Λεκτικές ή συμπεριφορικές εκδηλώσεις αποπροσανατολισμού στον χρόνο ή τον τόπο ή λανθασμένης αντίληψης των ατόμων στο περιβάλλον | | | |
| Ανάρμοστη συμπεριφορά<br>Ανάρμοστη συμπεριφορά στον τόπο ή τα άτομα | | | |
| Ψευδαισθήσεις, παραισθήσεις<br>Όραση ή ακοή πραγμάτων που δεν υπάρχουν | | | |
| Ψυχοκινητική υστέρηση<br>Καθυστερημένη αντίδραση, λίγες ή καθόλου αυτόματες πράξεις/λέξεις, πχ όταν ο ασθενής ενοχλείται η αντίδραση καθυστερεί ή/και ο ασθενήςδεν αφυπνίζεται | | | |
| Συνολική βαθμολογία (από το σύνολο των 10) | | | |
| Τα συμπτώματα βαθμολογούνται από το 0 έως το 10 ανάλογα με την παρουσία και την ένταση των συμπτωμάτων, και η κάθε βαθμολογία αθροίζεται για να οριστεί η τελική συνολική βαθμολογία. | | | |
| Βαθμολογία >2 στο nu-Desc δηλώνει την ύπαρξη παραληρήματος στο 86% των ασθενών. | | | |

**Figure 2:** The translated and adapted Greek version of the nursing delirium screening scale (nu-DESC).

eral study parameters, including the demographic data and the questionnaire scores. Count data were summarized in counts and percentages, whereas continuous parameters by their central tendency and range. Moreover, we used repeated-measures analysis of variance to test for statistically important differences among repeated measurements. The index of dissimilarity, percentage of agreement, and Cohen's kappa estimate were used to estimate the inter-rater agreement analysis. The former is a statistic used to measure the overall difference between two percent distributions. It ranges from 0 to 100 and indicates the proportion of items required to be reallocated to make the two distributions the same[22.] The percentage of agreement is a basic measure for inter-rater reliability calculated by the ratio of the number of ratings in agreement over the total number of ratings[23]. Additionally, the kappa statistic represents a quantitative measure of the magnitude of agreement between observers beyond what would be expected by chance. Kappa estimates range from "0.01 to 0.20", "0.21 to 0.40", "0.41 to 0.60", "0.61 to 0.80", and "0.81 to 1.00", and correspond to "slight agreement", "fair agreement", "moderate agreement", "substantial agreement", and "almost perfect or perfect agreement", respectively[24.] We used Cronbach's alpha statistic to evaluate the internal consistency for both scales. Values of alpha ≥0.9, 0.9>alpha≥0.8, 0.8>alpha≥0.7, 0.7>alpha≥0.6, 0.6>alpha≥0.5, alpha <0.5 represented excellent, good, acceptable, questionable, and poor scale internal consistency, respectively. The diagnostics accuracy of nu-DESC was assessed in terms of sensitivity, specificity, positive predictive value, negative predictive value, and positive and negative likelihood ratios, using CAM recordings as the reference tool. All data were analyzed using the statistical environment R[25]. Statistical significance for all analyses was set at $p < 0.05$.

*Study size*

The required sample size (questionnaires/records) for the inter-observer agreement analysis using Cohen's Kappa statistic (delirium prevalence: 0.25; the true Cohen's Kappa statistic: 0.96; kappa under the null hypothesis: 0.7; power analysis: 0.90; type I error: 0.05; two-sided) was estimated to be 65 questionnaires/records. With an anticipated dropout rate of 20 %, our study required a minimum of 78 records.

## Results

A total of 79 patients were eligible and included initially. Of them, 19 patients refused re-evaluation (Figure 3). Three pairs of records were available for each patient; at 8, 16, and 22 hours post-operatively. Our final sample size consisted of 180 records/questionnaires in total. There were 33 (55 %) females and 27 (45 %) males with a mean age of 71 (standard deviation: 6.23) years. Table 1 and Table 2 summarize the main characteristics of our study sample. There were not any missing data for each variable of interest.

*Inter-observer reliability and incidence of POD*

According to the first observer, eight (13.3 %), eleven (18.3 %), and twelve (20 %) patients were diagnosed with POD using CAM at the initial examination, 16-hour, and 22-hour, respectively. For the second observer, there were eight (13.3 %), eleven (18.3 %), and ten (16.7 %) patients with POD at the 8-hour, 16-hour, and 22-hour examination, respectively. Based on the nu-DESC, the first observer identified nine (15 %) cases with POD at the 8-hour evaluation, and eleven (18 %) and thirteen (22 %) cases at the 16-hour and 22-hour re-examination, respectively. At the same time points, the second observer recorded eight (13 %), eleven (18 %), and thirteen (22 %) patients with POD. According to the repeated measurement analysis, there was no statistically important difference in POD occurrence (Figure 4). The two independent raters completed 180 paired evaluations (Table 3 and Table 4). There was an "almost perfect agreement" between the two raters with the use of CAM [Cohen's Kappa estimate: 0.960; 95 % Confidence Interval (CI): 0.905-1.000] and nu-DESC (Cohen's Kappa estimate: 0.981; 95 % CI: 0.944-1.000). The agreement on each specific question of CAM and nu-DESC ranged from "substantial agreement" to "almost perfect agreement", as depicted in Table 5 and Table 6. The internal consistency of both questionnaires ranged between "excellent" and "good" (Table 7). Considering the Cronbach's alpha values, the measurements at 16 hours provide significant information if a measurement is dropped.

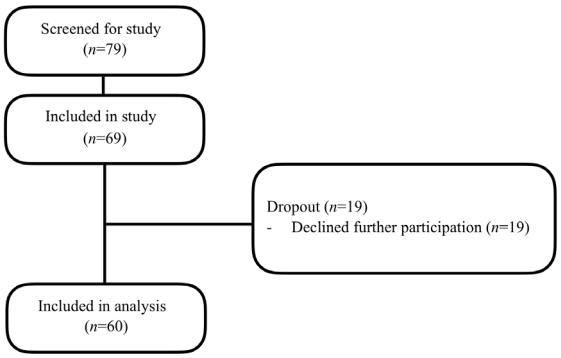



**Figure 3:** Flowchart depicting the patient sample selection.

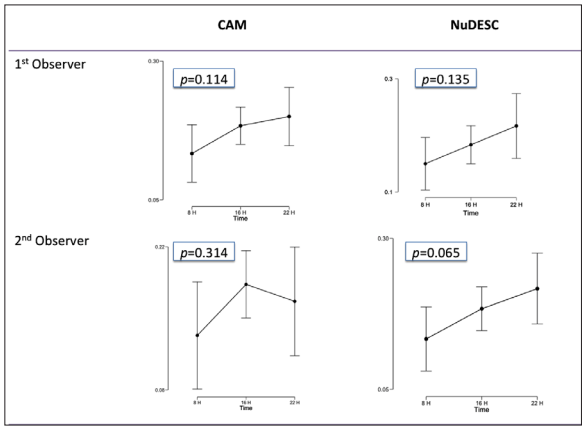



**Figure 4:** A two-by-two graph visualizing the interaction plots between the two observers scores (mean values and 95 % confidence interval of the occurrence of postoperative delirium) and time for the confusion assessment method (CAM) diagnostic algorithm and the nursing delirium screening scale (nu-DESC). The scores were measured at 8, 16, and 22 hours after surgery; p values indicate the statistical significance among the repeated measurements. Accordingly, there was no statistically important change in the scores with time in our study sample.

**Table 1:** Summary of the count parameters of the 60 consecutive patients who underwent elective non-cardiac surgery and were included in the study.

| | Subgroup | Counts | % |
|---|---|---|---|
| Gender | M | 33 | 55 |
| | F | 27 | 45 |
| ASA physical status | 2 | 35 | 58.3 |
| | 3 | 25 | 41.7 |
| Type of Surgery | General | 21 | 35 |
| | Vascular | 11 | 18.3 |
| | Gynaecological | 12 | 20 |
| | Urogical | 8 | 13.3 |
| | ENT | 3 | 5 |
| | Orthopaedic | 2 | 3.3 |
| | Neurosurgical | 2 | 3.3 |
| | Orthognathic | 1 | 1.7 |

M: male, F: female, ASA: American society of anesthesiologists, ENT: Ear, nose, throat.

**Table 2:** Summary of the continuous parameters of the 60 consecutive patients who underwent elective non-cardiac surgery and were included in the study.

| | Age (years) | Anesthesia Duration | Surgery Duration |
|---|---|---|---|
| Mean | 71.1 | 138 | 123 |
| Median | 71.07 | 130 | 110 |
| Standard deviation | 6.23 | 66.35 | 62.1 |
| Minimum | 60 | 30 | 25 |
| Maximum | 85 | 420 | 370 |

Duration of anesthesia and surgery is reported in minutes.

**Table 3:** Two by two table between the two observers for the confusion assessment method (CAM) diagnostic algorithm.

| | | 2nd Observer | |
|---|---|---|---|
| | | 0 | 1 |
| 1st Observer | 0 | 146 | 2 |
| | 1 | 4 | 28 |

**Table 4:** Two by two table between the two observers for the nursing delirium screening scale (nu-DESC).

| | | 2nd Observer | |
|---|---|---|---|
| | | 0 | 1 |
| 1st Observer | 0 | 147 | 0 |
| | 1 | 1 | 32 |

**Table 5:** Inter-rater agreement analysis of the confusion assessment method (CAM) diagnostic algorithm.

| | Index of dissimilarity | Percentage of agreement | Cohen's kappa estimate * | |
|---|---|---|---|---|
| Acute onset / fluctuating course | 0.097 | 98.3 | 0.930 (0.870-1.00) | Almost perfect agreement |
| Inattention | 0.033 | 99.4 | 0.979 (0.939-1.00) | Almost perfect agreement |
| Disorganized thinking | 0.413 | 92.7 | 0.683 (0.518-0.849) | Substantial agreement |
| Altered level of consciousness | 0.448 | 92.1 | 0.654 (0.479-0.828) | Substantial agreement |
| Total | 0.0645 | 98.9 | 0.960 (0.905-1.000) | Almost perfect agreement |

Values in brackets for Cohen's kappa represent 95 % confidence interval.

**Table 6:** Inter-rater agreement analysis of the nursing delirium screening scale (nu-DESC).

| | Index of dissimilarity | Percentage of agreement | Cohen's kappa estimate * | |
|---|---|---|---|---|
| Disorientation | 0.037 | 99.4 | 0.9731 (0.920-1.000) | Almost perfect agreement |
| Inappropriate behavior | 0.210 | 97.1 | 0.815 (0.656-0.975) | Almost perfect agreement |
| Inappropriate communication | 0.157 | 98.9 | 0.928 (0.821-1.000) | Almost perfect agreement |
| Illusions / Hallucination | 0.307 | 97.7 | 0.806 (0.619-0.993) | Almost perfect agreement |
| Psychomotor retardation | 0.272 | 97.1 | 0.709 (0.458-0.960) | Substantial agreement |
| Total | 0.033 | 99.4 | 0.981 (0.944-1.000) | Almost perfect agreement |

Values in brackets for Cohen's kappa represent 95 % confidence interval.

**Table 7:** Scale reliability statistics. Considering the Cronbach's alpha values and Cronbach's alpha values if a measurement is dropped.

| Scale | 1st Observer | | | 2nd Observer | | |
|---|---|---|---|---|---|---|
| | Cronbach's alpha | Cronbach's alpha (if item dropped) | | Cronbach's alpha | Cronbach's alpha (if item dropped) | |
| **CAM** | 0.912 | 8 hours | 0.913 | 0.904 | 8 hours | 0.906 |
| | | 16 hours | 0.813 | | 16 hours | 0.759 |
| | | 22 hours | 0.940 | | 22 hours | 0.902 |
| **nu-DESC** | 0.918 | 8 hours | 0.887 | 0.897 | 8 hours | 0.886 |
| | | 16 hours | 0.813 | | 16 hours | 0.762 |
| | | 22 hours | 0.940 | | 22 hours | 0.902 |

CAM: Confusion assessment method, nu-DESC: Nursing delirium screening scale. Sixty of the observations were used, 0 were excluded list wise, and 60 were provided. Values of alpha ≥0.9, 0.9>alpha≥0.8, 0.8>alpha≥0.7, 0.7>alpha≥0.6, 0.6>alpha≥0.5, alpha <0.5 represented excellent, good, acceptable, questionable, and poor scale internal consistency, respectively.

*Diagnostic accuracy of nu-DESC*

For the first observer, CAM and nu-DESC were at an "almost perfect agreement" (Cohen's Kappa estimate: 0.960; 95 % CI: 0.91-1.000). Based on CAM, nu-DESC was characterized by a sensitivity and specificity as high as 0.94 (95 % CI: 0.80-0.99) and 1.00 (95 % CI: 0.98-1.00), respectively (Figure 5). Likewise, for the second reviewer, there was an "almost perfect agreement" between the two scales (Cohen's Kappa estimate: 0.90; 95 % CI: 0.82-0.98). With the later values, based on CAM, the sensitivity and specificity of nu-DESC were 0.97 (95 % CI: 0.82-1.00) and 0.99 (95 % CI: 0.96-1.00), respectively (Table 8).

## Discussion

POD can prove to be a devastating complication for the surgical population[3,5]. Experts strongly recommend screening all patients for delirium, in each shift, up to the 5th postoperative day[1]. CAM and nu-DESC are among the recommended tools based on the guidelines released by ESA in 2017[1]. Of note, both instruments are not time-consuming; they do not require extensive training and can be administrated by non-psychiatrists[1,3,5,11].

This study provides the first official Greek translation and cultural adaptation of CAM and nu-DESC, according to the ISPOR guidelines, and the first evaluation of their inter-rater agreement. The Greek version of CAM and nu-DESC have a high inter-rater agreement, and ultimately, nu-DESC was characterized by significant diagnostic accuracy based on CAM. The incidence of POD in our cohort was less than 20 %, using CAM and nu-DESC.

Our findings are comparable with the originally de-

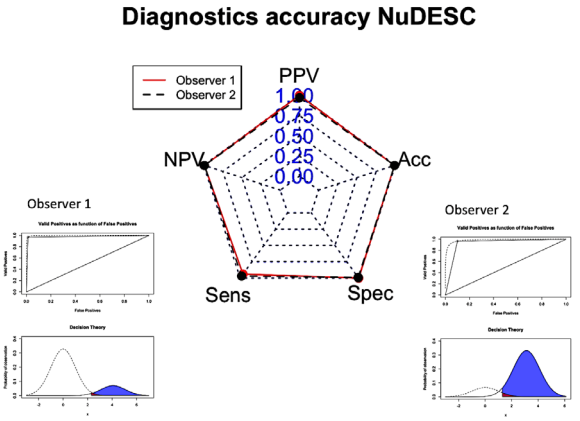



**Figure 5:** Nursing delirium screening scale (nu-DESC) is an excellent tool in screening for postoperative delirium. The spider plot visualizes the accuracy of nu-DESC as a screening tool for delirium in postoperative patients based on the confusion assessment method (CAM) diagnostic algorithm as a gold standard. The inserts highlight the diagnostic accuracy analysis's inherent details, including the area under the curve (AUC).

Spec: specificity, Sens: sensitivity, Acc: accuracy, PPV: positive predictive value, NPV: negative predictive value.

veloped versions of CAM and nu-DESC. In particular, Inouye et al reported inter-rater reliability for CAM as high as 93 %[9]. Regarding the features of CAM, the agreement ranged from "substantial agreement" for the features of disorganized thinking and altered level of consciousness to "almost perfect agreement" for the acute onset/fluctuating course and inattention. Inouye et al described agreement at 56 % and 100 % for the altered level of consciousness and the disorganized thinking, respectively[9]. In another study, Martins et al validated the European Portuguese version of CAM and stated that the inter-rater reliability in their study was 1.0 for the acute onset, 0.78 for inattention, 0.65 for disorganized thinking, and 1.0 for altered level of consciousness[26]. These minor disparities in the agreement among the different translations could be attributed to the distinctive characteristics in the two raters' experience, the non-homogeneity of the study populations, and, more importantly, the patients' dissimilarities regarding their educational background.

Likewise, Gaudreau et al focused on the agreement between CAM and nu-DESC, and the reported sensitivity and specificity rates were as high as 85.7 % and 86.8 %, respectively[9,10]. In our study, the features disorientation, inappropriate behavior, inappropriate communication, and illusions/hallucinations were associated with "almost perfect agreement". However, the feature psychomotor retardation was characterized by "substantial agreement". One possible explanation could be the high proportion of the hyperactive type of POD (23/60) in our study sample. Moreover, our sample's homogeneity (patients over the age of sixty and without dementia, who underwent elective non-cardiac surgery under general anesthesia) could be another reasonable justification. The predictive value of nu-DESC can vary considerably between populations with low versus high prevalence of delirium, and the positive and negative predictive values of our results (Table 8) highlight the significance of the homogeneity of our sample[10,12].

Additionally, the fact that the postoperative delirium's assessment was performed by anesthesiologists, who were able to evaluate the fluctuating daily course of POD, may strengthen the high inter-rater reliability of our study. The fluctuating course of postoperative delirium makes its clinical routine detection and diagnosis quite challenging. Thus, the dedicated and experienced physicians play a central role in the diagnosis of postoperative delirium[3]. Although nu-DESC was initially developed to be administrated by nurses, Hägi-Pedersen et al performed the validation of the Danish version of nu-DESC successfully by estimating the inter-rater agreement between medical doctors and nurses[12]. Their study showed that the nu-DESC is comprehensible and that there is no significant difference between the nursing staff and medical doctors' evaluation results, which suggests that the tool may be equally usable by both groups[12].

A limitation in the present study could be the selection of the population. Besides the numerous advantages of our study sample's homogeneity, further studies are necessary for assessing the validity of CAM and nu-DESC to other surgical populations (non-elective procedures, cardiac surgery). Additionally, we did not choose to compare either CAM or nu-DESC to Diagnostic and Statistical Manual of Mental Disorders (DSM) V or IV because an isolated comparison did not seem meaningful[9-12,21,27,28]. Other CAM and nu-DESC translations were only compared to DSM IV, and the whole postoperative delirium focus has shifted significantly with the progression from DSM IV to V[1,10-12,26,29]. Moreover, the utility

**Table 8:** Accuracy parameters of the nursing delirium screening scale (nu-DESC) as a screening tool for postoperative delirium, based on the confusion assessment method (CAM) diagnostic algorithm.

| | 1st Observer | 2nd Observer |
|---|---|---|
| **Apparent prevalence** | 0.17 (0.12-0.24) | 0.16 (0.11-0.23) |
| **True prevalence** | 0.18 (0.13-0.25) | 0.16 (0.11-0.23) |
| **Sensitivity** | 0.94 (0.80-0.99) | 0.97 (0.82-1.00) |
| **Specificity** | 1.00 (0.98-1.00) | 0.99 (0.96-1.00) |
| **Positive predictive value** | 1.00 (0.89-1.00) | 0.97 (0.82-1.00) |
| **Negative predictive value** | 0.99 (0.95-1.00) | 0.99 (0.96-1.00) |
| **Positive likelihood ratio** | Inf (NaN, Inf) | 142.90 (20.24-1008.95) |
| **Negative likelihood ratio** | 0.06 (0.02-0.23) | 0.03 (0.01-0.24) |

Values in brackets represent 95 % confidence interval, Inf: infinity, NaN: not a number.

of DSM has been questioned during the past years and, based on current literature, there is only a 30 % overlap between the DSM-IV and DSM-V in delirium diagnosis under strict criteria[3,4,10-12,28,29]. Of note, both screening tools, CAM and nu-DESC, have been abundantly validated in comparison to DSM in numerous different languages and settings[1,10-12,21].

With the translation, the cultural adaptation, and the determination of the inter-rater agreement of the Greek versions of the CAM and nu-DESC, our study provides two reliable screening tools for the diagnosis of POD in our country. Moreover, as nu-DESC can detect all psychomotor types of delirium, its implementation in daily practice may increase awareness about POD in Greece. Both CAM and nu-DESC are internationally recognized scales recommended by the ESA. Thus, they may be used in multi-center national and international clinical research, improving the safety and the peri-operative care of patients undergoing surgery and/or anesthesia.

**Conflict of Interest**

Authors declare no conflict of interest and not receiving any specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Acknowledgements**

Adapted from: Inouye SK, et al. Ann Intern Med. 1990; 113: 941-948. Confusion Assessment Method. Copyright 2003, Hospital Elder Life Program, LLC. Not to be reproduced without permission.

Permission granted by the American Geriatrics Society, 2019. For more information visit: help.agscocare.org. Disclaimer: “No responsibility is assumed by the AGS or the Hospital Elder Life Program, LLC for any injury and/or damage to persons or property arising out of the application of any of the content at: help.agscocare.org”.

Adapted from: Gaudreau JD, et al. J Pain Symptom Manage. 2005; 29: 368-375. Fast, systematic, and continuous delirium assessment in hospitalized patients: the nursing delirium screening scale.

**References**

1. Aldecoa C, Bettelli G, Bilotta F, Sanders RD, Audisio R, Borozdina A, et al. European Society of Anaesthesiology evidence-based and consensus-based guideline on postoperative delirium. Eur J Anaesthesiol. 2017; 34: 192-214.
2. Hood R, Budd A, Sorond FA, Hogue CW. Peri-operative neurological complications. Anaesthesia. 2018; 73 Suppl 1: 67-75.
3. Rengel KF, Pandharipande PP, Hughes CG. Postoperative delirium. Presse Med. 2018; 47: e53-e64.
4. Ntalouka MP, Arnaoutoglou E, Tzimas P. Peri-operative Cognitive Alterations and the Ageing Brain. Clin Surg. 2019; 4: 2419.
5. Strøm C, Rasmussen LS, Sieber FE. Should general anaesthesia be avoided in the elderly? Anaesthesia. 2014; 69 Suppl 1: 35-44.
6. Ntalouka MP, Arnaoutoglou E, Tzimas P. Postoperative cognitive disorders: an update. Hippokratia. 2018; 22: 147-154.
7. Noimark D. Predicting the onset of delirium in the post-operative patient. Age Ageing. 2009; 38: 368-373.
8. Krenk L, Rasmussen LS. Postoperative delirium and postoperative cognitive dysfunction in the elderly - what are the differences? Minerva Anestesiol. 2011; 77: 742-749.
9. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. Ann Intern Med. 1990; 113: 941-948.
10. Gaudreau JD, Gagnon P, Harel F, Tremblay A, Roy MA. Fast, systematic, and continuous delirium assessment in hospitalized patients: the nursing delirium screening scale. J Pain Symptom Manage. 2005; 29: 368-375.
11. Grover S, Kate N. Assessment scales for delirium: A review. World J Psychiatry. 2012; 2: 58-70.
12. Hägi-Pedersen D, Thybo KH, Holgersen TH, Jenses JJ, Gaudreau JD, Radtke FM. Nu-DESC DK: the Danish version of the nursing delirium screening scale (nu-DESC). BMC Nurs. 2017; 16: 75.
13. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013; 310: 2191-2194.
14. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005; 8: 94-104.
15. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011; 64: 96-106.
16. PLOS Medicine Editors. Observational studies: getting clear about transparency. PLoS Med. 2014; 11: e1001711.
17. Feinkohl I, Winterer G, Spies CD, Pischon T. Cognitive Reserve and the Risk of Postoperative Cognitive Dysfunction. Dtsch Arztebl Int. 2017; 114: 110-117.
18. Needham MJ, Webb CE, Bryden DC. Postoperative cognitive dysfunction and dementia: what we need to know and do. Br J Anaesth. 2017; 119: i115-i125.
19. de Craen AJ, Heeren TJ, Gussekloo J. Accuracy of the 15-item geriatric depression scale (GDS-15) in a community sample of the oldest old. Int J Geriatr Psychiatry. 2003; 18: 63-66.
20. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. 1969; 9: 179-186.
21. Inouye SK. The Confusion Assessment Method (CAM): Training Manual and Coding Guide. Yale University School of Medicine, New Haven, CT, 2003.
22. Allen R, Burgess S, Davidson R, Windmeijer F. More reliable inference for the dissimilarity index of segregation. Econom J. 2015; 18: 40-66.
23. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012; 22: 276-282.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33: 159-174.
25. The R Project for Statistical Computing. The R Foundation. Available at: https://www.R-project.org, date accessed: 12/12/19.
26. Martins S, Lourenço C, Pinto-de-Sousa J, Conceição F, Paiva JA, Simões MR, et al. Validation study of the European Portuguese version of the Confusion Assessment Method (CAM). Int Psychogeriatr. 2015; 27: 777-784.
27. Wei LA, Fearing MA, Sternberg EJ, Inouye SK. The Confusion Assessment Method: a systematic review of current usage. J Am Geriatr Soc. 2008; 56: 823-830.
28. Meagher DJ, Morandi A, Inouye SK, Ely W, Adamis D, Maclullich AJ, et al. 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. 2014; 12: 164.
29. Adamis D, Meagher D, Rooney S, Mulligan O, McCarthy G. A comparison of outcomes according to different diagnostic systems for delirium (DSM-5, DSM-IV, CAM, and DRS-R98). Int Psychogeriatr. 2018; 30: 591-596.