



## Brief Report

# Optimizing the month of the year backwards test for delirium screening of older patients in the emergency department



Wolfgang Hasemann, PhD<sup>a,\*</sup>, Florian F. Grossmann, MScN<sup>b</sup>, Roland Bingisser, MD<sup>b</sup>, Martina Hafner, MD<sup>c</sup>, Dieter Breil, MD<sup>d</sup>, Reto W. Kressig, MD<sup>d</sup>, Christian H. Nickel, MD<sup>b</sup>

<sup>a</sup> Department of Nursing and Allied Health Professions, University Hospital Basel, Hebelstr 2, CH - 4031 Basel, Switzerland

<sup>b</sup> Emergency Department, University Hospital Basel, Petersgraben 2, CH-4031 Basel, Switzerland

<sup>c</sup> University Clinic of Medicine, Kantonsspital Baselland, Rheinstrasse 26, CH-4410 Liestal, Switzerland

<sup>d</sup> University Center for Medicine of Aging, Felix Platter Hospital Basel, Burgfelderstrasse 101, CH - 4012 Basel, Switzerland

## ARTICLE INFO

## Article history:

Received 18 November 2018

Received in revised form 19 June 2019

Accepted 20 June 2019

## Keywords:

Delirium

Symptom assessment

Neuropsychological tests

Emergency medicine

Emergency nursing

Diagnostic testing

Psychometrics

## ABSTRACT

**Introduction:** Different scoring methods exist for the Month of the Year Backward Test (MBT), which is designed to detect inattention, the core feature of delirium. When used as a part of the modified Confusion Assessment Method for the Emergency Department (mCAM-ED), each error in the MBT scores one point. Because this scoring procedure is complex, we aimed to simplify the scoring method of the MBT.

**Methods:** This is a secondary analysis of a single center prospective validation study of the mCAM-ED comprising a sample of Emergency Department (ED) patients aged 65 or older presenting to our ED.

**Data collection:** Research assistants (RAs) who were trained nurses conducted the MBT. Geriatricians conducted the reference standard delirium assessment within 1 h of the RA.

**Results:** For the scoring method “number of errors”, optimal performance according the Youden index was achieved when 8 or more errors were reached resulting in an overall sensitivity of 0.95 and overall specificity of 0.94. The scoring method “number of errors in combination with time needed” resulted in a comparable result with minimally lower positive likelihood ratios. For the scoring method “last month in correct order”, optimal performance according the Youden index was achieved with the month of September resulting in an overall sensitivity of 0.90 and an overall specificity of 0.89.

**Discussion:** We suggest omitting the factor time and using a more practical scoring method with good performance: “last month in correct order” with the requirement to reach September to rule out delirium.

© 2019 Elsevier Inc. All rights reserved.

## 1. Introduction

Delirium in the Emergency Department (ED) is associated with increased hospital length of stay and 6-month-mortality [1,2]. As the main feature of delirium is inattention, measuring inattention with the Month of the Year Backwards Test (MBT) is a fast method to rule out delirium [3]. The MBT was previously suggested to be the best single delirium screening item, particularly due to its relative brevity and simplicity compared to other approaches [4,5]. However, difficulties in completing the MBT may also indicate the presence of significant cognitive problems such as dementia [3,6].

We previously developed and validated the modified Confusion Assessment Method for the Emergency Department (mCAM-ED), a two-step tool for delirium diagnosis in the emergency department (ED)

[7,8]. In the first step, the Month of the Year Backwards Test (MBT) is used to rule out delirium. Different scoring methods exist for the MBT such as reciting the month backwards errorlessly to July or January [3,9]. In our study, in which the MBT was one component of the mCAM-ED, we used the scoring methods of counting the amount of errors according to Stillman et al. [9]: Each error in the MBT, defined as every pair of months that is not recited in the correct order, scores one point, and if the patient needs longer than 30 s for this task, an additional point is added. Because this scoring procedure is complex, we aimed to simplify the scoring method of the MBT and to determine the best cut-off score for delirium screening in the ED.

## 2. Methods

### 2.1. Setting

This is a secondary analysis of a single center prospective validation study of the mCAM-ED [8] comprising a sample of ED patients aged 65 or older presenting to our ED. The 24/7 consecutive data collection

Abbreviations: mCAM-ED, modified Confusion Assessment Method for the Emergency Department; MBT, month of the year backwards test.

\* Corresponding author.

E-mail address: [wolfgang.hasemann@usb.ch](mailto:wolfgang.hasemann@usb.ch) (W. Hasemann).

covered an 11-day period in November 2015. Patients were excluded if they were treated in the resuscitation bay, were unable to communicate verbally (e.g. aphasia, coma) or had other communication problems (e.g. severe hearing impairment, language barrier).

## 2.2. Measurements

### 2.2.1. Month of the year backwards test (MBT)

Patients were asked to recite the months of the year in reverse order beginning with December. In this study, we compared different scoring methods:

1. “Number of errors”: the number of errors is counted. Errors are defined as every pair of months that is not recited in the correct order. (Range: 0–12 points)
2. “Number of errors in combination with time needed”: If a patient needed more than 30 s to carry out the task, this was defined as an additional error according to the approach of Stillman et al. [9]. (Range: 0–13 points)
3. “Last month in correct order”: The month until which a person was able to recite in reverse order before an error occurs.

Patients who were not able to meaningfully engage with the task at all (i.e. no month was mentioned at all or months were recited in an unusual order) were scored with the worst result or 12 error points. One additional point was given if patients needed more than 30 s for the task.

## 2.3. Definitions

### 2.3.1. Delirium

A delirium diagnosis was the primary endpoint. Reference diagnosis of delirium was established by geriatricians based on criteria of the Text Revision of the fourth version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR). Accordingly, delirium was defined as present when there was an acute change in consciousness and cognition with fluctuating symptoms over the course of the day which were caused by the direct physiological consequences of a general medical condition [10]. The geriatricians used the first 10 questions of the Swiss Minimal State Examination [11] as standardized approach and let the patient recite the days of the week backwards to avoid learning on the part of the patient. Acute changes in cognition were evaluated using information by proxies and chart review.

### 2.3.2. Dementia

We defined patients as having dementia when a dementia diagnosis was documented in the electronic health record.

## 2.4. Data collection

Research assistants (RAs) who were trained nurses conducted the MBT. On a 12-point scale, (depicting 12 months) patients' responses were drawn as curves by the RA. Board certified geriatricians working as attending physicians conducted the reference standard delirium assessment within 1 h of the RA. Additionally, they conducted an in depth delirium assessment. Geriatricians were blinded to the results of the mCAM-ED.

## 2.5. Statistical analysis

Demographic data were calculated as frequencies and percentages. Continuous variables were compared using a Mann-Whitney *U* test and a chi-square-test was used for binary variables. For each result of an MBT scoring method (“number of errors”, “number of errors in combination with time needed” and “last month in correct order”) true positive (TP), false positive (FP), true negative (TN), false negative (FN) results and a receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC) were calculated with their two-sided 95% confidence intervals (CI). Sensitivity and specificity as well as positive and negative predictive values (PPV and NPV) were calculated and presented together with their two-sided 95% CI. These were computed according to the method of Agresti & Coull [12]. In addition, the positive and negative likelihood ratios (LR+ and LR–) were calculated together with their two-sided 95% CIs and computed according to the method of Zhou et al. [13]. The optimal cut-off, that is the best achievable combination of sensitivity and specificity, was estimated with the Youden index. Data were calculated separately for patients with and without dementia.

## 2.6. Ethical considerations

The study was approved by the local ethics committee (identifier EKNZ-2015-123) and registered with [ClinicalTrials.gov](https://www.clinicaltrials.gov) (identifier NCT02782143) [8].

## 3. Results

For this subanalysis 286 patients with a median age of 80.02 (Q1 = 72.15; Q3 = 86.76), and of which 168 (58.7%) were female, were included. Dementia was documented in 41 (14.3%) out of 286 patients in the electronic medical record. According to the reference standard, 20 patients (7.0%) had delirium. Of those, 11 had delirium superimposed on dementia. Reciting the months backwards without any error was possible for 168 out of 236 (71.2%) patients without cognitive impairment, 8 out of 30 (26.7%) patients with dementia, 0 out of 9

**Table 1**

Highest vs. optimal cut-off for number of errors in the MBT.

Cut-off	Highest			Optimal		
	Healthy if <1 error			Healthy if <8 errors		
Population	Overall	Without dementia	With dementia	Overall	Without dementia	With dementia
Sensitivity	1.00 [0.84; 1.00]	1.00 [0.70; 1.00]	1.00 [0.74; 1.00]	0.95 [0.76; 0.99]	0.89 [0.57; 0.98]	1.00 [0.74; 1.00]
Specificity	0.66 [0.60; 0.72]	0.71 [0.65; 0.77]	0.27 [0.14; 0.44]	0.94 [0.91; 0.97]	0.96 [0.92; 0.98]	0.83 [0.66; 0.93]
PPV	0.18 [0.12; 0.26]	0.12 [0.06; 0.21]	0.33 [0.20; 0.50]	0.56 [0.39; 0.71]	0.44 [0.25; 0.66]	0.69 [0.44; 0.86]
NPV	1.00 [0.98; 1.00]	1.00 [0.98; 1.00]	1.00 [0.68; 1.00]	1.00 [0.98; 1.00]	1.00 [0.98; 1.00]	1.00 [0.87; 1.00]
LR+	2.96 [2.50; 3.50]	3.47 [2.84; 4.24]	1.36 [1.10; 1.69]	16.85 [10.20; 27.82]	20.98 [10.96; 40.14]	6.00 [2.70; 13.36]
LR-	0.00 [0.00; NaN]	0.00 [0.00; NaN]	0.00 [0.00; NaN]	0.05 [0.01; 0.36]	0.12 [0.02; 0.74]	0.00 [0.00; NaN]
TP	20	9	11	19	8	11
FP	90	68	22	15	10	5
TN	176	168	8	251	226	25
FN	0	0	0	1	1	0

Highest sensitivity was achieved with a cut-off <1 error and optimal ratio between sensitivity and specificity was achieved with <8 errors.

PPV = positive predictive value; NPV = negative predictive value; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; TP = true positive; FP = false positive; TN = true negative; FN = false negative; NaN = not a number.

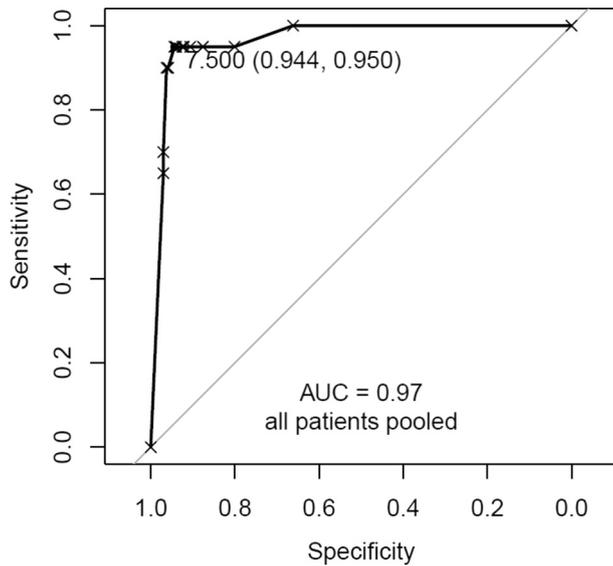


Fig. 1. ROC for number of errors in the MBT.

(0%) patients with delirium and 0 out of 11 (0%) patients with delirium superimposed on dementia. Two out of 236 patients (0.8%) without cognitive impairment, 3 out of 30 patients (10.0%) with dementia, 2 out of 9 patients (22.2%) with delirium and 7 out of 11 patients (63.6%) with delirium superimposed on dementia were not able to meaningfully engage in the MBT.

For the scoring method “number of errors”, optimal performance according the Youden index was achieved when 8 or more errors were reached resulting in an overall sensitivity of 0.95 (95% CI = 0.76; 0.99) and overall specificity of 0.94 (95% CI = 0.91; 0.97) (Table 1). We found a sensitivity of 1.00 (95% CI = 0.74; 1.00) and 0.89 (95% CI = 0.57; 0.98) and a specificity of 0.83 (95% CI = 0.66; 0.93) and 0.96 (95% CI = 0.92; 0.98) in patients with and without dementia, respectively. The overall AUC was 0.97 (Fig. 1).

The scoring method “number of errors in combination with time needed” resulted in a slight decrease in positive likelihood ratios.

For the scoring method “last month in correct order”, optimal performance according the Youden index was achieved with the month of September resulting in an overall sensitivity of 0.90 (95% CI = 0.70; 0.97) and an overall specificity of 0.89 (95% CI = 0.85; 0.92). We found a sensitivity of 0.91 (95% CI = 0.62; 0.98) and 0.89 (95% CI = 0.57; 0.98) and a specificity of 0.73 (95% CI = 0.56; 0.86) and 0.91 (95% CI = 0.87; 0.94) in patients with and without dementia, respectively (Table 2). The AUC in the overall ROC analysis was 0.96.

Table 2  
Highest vs. optimal cut-off for last month in correct order in the MBT.

Cut-off	Highest			Optimal		
	Healthy if = January			Healthy if ≥September		
Population	Overall	Without dementia	With dementia	Overall	Without dementia	With dementia
Sensitivity	1.00 [0.84; 1.00]	1.00 [0.70; 1.00]	1.00 [0.74; 1.00]	0.90 [0.70; 0.97]	0.89 [0.57; 0.98]	0.91 [0.62; 0.98]
Specificity	0.66 [0.60; 0.72]	0.71 [0.65; 0.77]	0.27 [0.14; 0.44]	0.89 [0.85; 0.92]	0.91 [0.87; 0.94]	0.73 [0.56; 0.86]
PPV	0.18 [0.12; 0.26]	0.12 [0.06; 0.21]	0.33 [0.20; 0.50]	0.38 [0.26; 0.53]	0.28 [0.15; 0.46]	0.56 [0.34; 0.75]
NPV	1.00 [0.98; 1.00]	1.00 [0.98; 1.00]	1.00 [0.68; 1.00]	0.99 [0.97; 1.00]	1.00 [0.97; 1.00]	0.96 [0.79; 0.99]
LR+	2.96 [2.50; 3.50]	3.47 [2.84; 4.24]	1.36 [1.10; 1.69]	8.26 [5.68; 11.99]	9.99 [6.25; 15.97]	3.41 [1.83; 6.35]
LR-	0.00 [0.00; NaN]	0.00 [0.00; NaN]	0.00 [0.00; NaN]	0.11 [0.03; 0.42]	0.12 [0.02; 0.77]	0.12 [0.02; 0.81]
TP	20	9	11	18	8	10
FP	90	68	22	29	21	8
TN	176	168	8	237	215	22
FN	0	0	0	2	1	1

Highest sensitivity was achieved with a cut-off of January and optimal ratio between sensitivity and specificity was achieved with September.

PPV = positive predictive value; NPV = negative predictive value; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; TP = true positive; FP = false positive; TN = true negative; FN = false negative; NaN = not a number.

## 4. Discussion

### 4.1. Main discussion

In comparison with the current scoring method “number of errors in combination with time needed”, in the scoring method “number of errors” an optimal performance of the MBT to indicate delirium in ED patients was achieved with a cut-off of 8 errors or more. Although slightly lower in performance, the scoring method “last month in correct order” using the September cutoff is easier. Using time as a factor did not provide additional value in the “number of errors in combination with time needed”.

Our approach of rating patients who are not able to meaningfully engage with the MBT with the worst score seems adequate. This is supported by another study which found that these patients are often delirious [14].

In our study, the MBT showed a better performance than in other studies [5,15,16]. There are 3 possible explanations for these findings. First, it may be the population studied (Switzerland versus US). Second, a different reference standard was used (psychiatrists versus geriatricians) [5,16]. Third, the case mix might be different between settings (ED versus non-ED) and countries (Switzerland versus US). In our study, geriatricians served as reference standard and assessed patients within only 1 h after the RAs in the busy ED environment. This narrow time frame limited the chance for acute changes between the assessments. In another ED study investigating performance criteria of different MBT cut-offs (July and January, respectively) performance was lower [16].

The positive likelihood ratio was lower in patients with dementia than in patients without dementia in both scoring methods; i.e. number of errors and last month in correct order. This was true when the scores were maximized for sensitivity and even when optimized for the Youden index. Thus, the MBT does not qualify to rule in delirium, but its sensitivity is sufficient to rule out delirium.

Although for the scoring method “number of errors” the Youden index provided the optimal cut-off of 8 errors or more to identify delirium, in clinical practice the transcription of the patients’ responses and its evaluation require additional effort and time. We therefore suggest omitting the factor time and using a more practical scoring method with good performance: “last month in correct order” with the requirement to reach September to rule out delirium.

### 4.2. Strengths and limitations

Due to the relatively low delirium prevalence, and a lower sample size as anticipated, some confidence intervals could not be calculated

with a high precision. However, our sample consisted of a consecutive sample of ED patients and consequently minimizes selection bias.

#### 4.3. Conclusion

In summary, the application of the mCAM-ED can be simplified by scoring the MBT with the method “last month in correct order” where to rule out delirium September has to be reached without an error.

#### Financial support

Scientific funds of the University Hospital Basel.

#### Declaration of Competing Interest

None.

#### References

- [1] Emond M, Boucher V, Carmichael PH, Voyer P, Pelletier M, Gouin E, et al. Incidence of delirium in the Canadian emergency department and its consequences on hospital length of stay: a prospective observational multicentre cohort study. *BMJ Open* 2018;8, e018190. <https://doi.org/10.1136/bmjopen-2017-018190>.
- [2] Han JH, Shintani A, Eden S, Morandi A, Solberg LM, Schnelle J, et al. Delirium in the emergency department: an independent predictor of death within 6 months. *Ann Emerg Med* 2010;56:244–52 e1. <https://doi.org/10.1016/j.annemergmed.2010.03.003>.
- [3] Meagher J, Leonard M, Donoghue L, O'Regan N, Timmons S, Exton C, et al. Months backward test: a review of its use in clinical studies. *World J Psychiatry* 2015;5: 305–14. <https://doi.org/10.5498/wjp.v5.i3.305>.
- [4] Zhou W, Ngo LH, Inouye SK, Gallagher J, Husser EK, Fick DM, et al. Does a year have 6 months or 12? Implications for delirium detection among hospitalized older general medicine patients. *J Gen Intern Med* 2019;34:354–5. <https://doi.org/10.1007/s11606-018-4704-3>.
- [5] Fick DM, Inouye SK, Guess J, Ngo LH, Jones RN, Saczynski JS, et al. Preliminary development of an ultrabrief two-item bedside test for delirium. *J Hosp Med* 2015;10: 645–50. <https://doi.org/10.1002/jhm.2418>.
- [6] Cunningham EL, Passmore AP, McAuley DF, McGuinness B. Reciting the months of the year backwards: what is a 'normal' score? *Age Ageing* 2015;44:537–8. <https://doi.org/10.1093/ageing/afv056>.
- [7] Grossmann FF, Hasemann W, Graber A, Bingisser R, Kressig RW, Nickel CH. Screening, detection and management of delirium in the emergency department – a pilot study on the feasibility of a new algorithm for use in older emergency department patients: the modified Confusion Assessment Method for the Emergency Department (mCAM-ED). *Scand J Trauma Resusc Emerg Med* 2014;22:19. <https://doi.org/10.1186/1757-7241-22-19>.
- [8] Hasemann W, Grossmann FF, Stadler R, Bingisser R, Breil D, Hafner M, et al. Screening and detection of delirium in older ED patients: performance of the modified Confusion Assessment Method for the Emergency Department (mCAM-ED). A two-step tool. *Intern Emerg Med* 2017;1–8. <https://doi.org/10.1007/s11739-017-1781-y>.
- [9] Stillman MJ, Rybicki LA. The bedside confusion scale: development of a portable bedside test for confusion and its application to the palliative medicine population. *J Palliat Med* 2000;3:449–56. <https://doi.org/10.1089/jpm.2000.3.449>.
- [10] American Psychiatric Association. Delirium, dementia, and amnesic and other cognitive disorders. Diagnostic and statistical manual of mental disorders DSM-IV-TR. 4th ed. Washington, DC: American Psychiatric Association; 2000; 135–47.
- [11] Monsch AU, Foldi NS, Ermini-Funfschilling DE, Berres M, Taylor KI, Seifritz E, et al. Improving the diagnostic accuracy of the mini-mental state examination. *Acta Neurol Scand* 1995;92:145–50.
- [12] Agresti A, Coull BA. Approximate is better than “exact” for interval estimation of binomial proportions. *Am Stat* 1998;52:119–26.
- [13] Zhou X-H, McClish DK, Obuchowski NA. Statistical methods in diagnostic medicine. 2nd ed. Oxford: Wiley-Blackwell; 2011.
- [14] Hasemann W, Duncan N, Clarke C, Nouzova E, Rutter L-M, Keerie C, et al. Utility of the month of the year backwards test (MOTYB) in detecting delirium: analysis of error patterns and overall performance in 149 older hospitalised patients. Presented at the 11 Annual Meeting of the European Delirium Association; 2017 Oslo, Norway.
- [15] Hendry K, Quinn TJ, Evans J, Scortichini V, Miller H, Burns J, et al. Evaluation of delirium screening tools in geriatric medical inpatients: a diagnostic test accuracy study. *Age Ageing* 2016;45:832–7. <https://doi.org/10.1093/ageing/afw130>.
- [16] Marra A, Jackson JC, Ely EW, Graves AJ, Schnelle JF, Dittus RS, et al. Focusing on inattention: the diagnostic accuracy of brief measures of inattention for detecting delirium. *J Hosp Med* 2018. <https://doi.org/10.12788/jhm.2943>.