Comparison of verbal and computerised months backwards tests in a hospitalized older population

Martin Mulligan1 · Leona Lally1,2 · Dimitrios Adamis2 · David Meagher3 · Chris Exton4 · Colum Dunne5 · Geraldine McCarthy1,2

Received: 3 March 2022 / Accepted: 19 July 2022 / Published online: 5 August 2022
© The Author(s) 2022

Abstract

Background Delirium is extremely prevalent, yet underdiagnosed, in older patients and is associated with prolonged length of hospital stay and higher mortality rates. Impaired attention is the cardinal deficit in delirium and is a required feature in diagnostic criteria. The verbal months backwards test (MBT) is the most sensitive bedside test of attention, however, hospital staff occasionally have difficulty with its administration and interpretation. We hypothesise that the MBT on an electronic tablet may be easier and more consistent to use for both experienced and unexperienced medical professionals and, if the diagnostic efficacy was similar, aid delirium diagnosis.

Aim We aim to investigate the correlation of the verbal MBT with a computerised MBT application.

Methods Participants recruited (age > 65, n = 75) were allocated to different cohorts (Dementia and Delirium (DMDL), Dementia (DM), Delirium (DL), No Neurocognitive Disorder (NNCD)) and were administered both the verbal and electronic versions.

Results Correlation between measurements were: overall Spearman’s rho = 0.772 (p < 0.0001); DMDL rho = 0.666 (p < 0.0001); DL rho = 0.778 (p = 0.039); DM rho = 0.378 (p = 0.203); NNCD rho = 0.143 (p = 0.559).

Discussion Overall, and for the delirious subset, statistically significant agreement was present. Poor inter-test correlation existed in the groups without delirium (DM, NNCD).

Conclusions The MBTc correlates well with the MBTv in patients who are clinically suspected to have delirium but has poor correlation in patients without delirium. Visuospatial cognition and psychomotor deficits in a dementia cohort and mechanical factors (such as tremor, poor fingernail hygiene and visual impairment) in a group with no neurocognitive disorder may limit the utility of the MBTc in a hospitalised older population.

Keywords Delirium · Dementia · Attention · Months backwards test · Information technology

Introduction

Delirium is estimated to occur in 21% of all inpatients [1] and is associated with longer inpatient stays and higher mortality rates [2]. Delirium is more prevalent in older patients (> 65 years old), especially in those with dementia [1], and is misdiagnosed, detected late or missed in over 50% of cases [3]. Not only is cognitive impairment a key element of delirium, it’s also a well-recognised risk factor for the development of the condition [4]. Impaired attention is the cardinal cognitive deficit associated with delirium and is a required feature for diagnosis according to criteria such as those outlined in the DSM 5 and ICD-10 [5, 6]. The months backwards test (MBT) is a brief, simple bedside test of attention that is also thought to engage
short term memory, executive function and central processing [7, 8]. Bellelli et al. found that the MBT had a high sensitivity for delirium while Adamis et al. showed that the MBT was the single most sensitive test for identifying delirium in an elderly cohort (compared to digit span, vigilance A test, serial sevens) [9, 10]. Indeed, the MBT forms part of the test for attention in the Revised Delirium Rating Scale administration manual [11]. However, as evidenced by a systematic review of verbal MBT use, there are inconsistencies in its administration and interpretation [12]. The 22 articles in this review describe a variety of MBT metrics, the result of each decided and interpreted by a human assessor. The authors of this study propose that a human examiner may vary in test administration e.g. via error interpretation. A major challenge in detecting delirium is the lack of a convenient, universally applicable tool that assesses for impaired attention [13]. By employing computer technology, it is thought that inter-rater inconsistencies can be minimised [14]. We hypothesise that by having the test administered via a computer application, the test could be universally administered (by both experienced and inexperienced doctors, nursing staff and allied healthcare professionals) and interpreted with minimal need for prior training with reproducible results independent of the assessor. Given the evidence base already established (see above) for the verbal MBT, if a computerised version of the test correlated well with the verbal version that is administered by experienced and formally trained assessors, we hypothesise that the computerised MBT may be easier and more consistent to use for both experienced and unexperienced medical professionals alike.

In a preliminary study of a newly developed computerised months backwards test (application on a handheld android tablet; MBTc) comparing medically unwell elderly patients and cognitively intact young people, the MBTc had excellent sensitivity for detecting cognitive impairment [15]. This study found that the computerised MBT allows for accurate and efficient testing of general cognition in clinical populations. However, a limitation of this study was that dementia was only diagnosed in one patient; our study aims to explore MBTc use in this cohort. The aim of this pilot study is to observe if there is significant correlation between patient performance in the MBTc and the MBTv in a cohort with mixed neuropsychiatric conditions. We hypothesise that employing a standardised test on a handheld electronic tablet in an elderly cohort will allow ease of interpretation and reduce scoring inconsistencies for assessors of varying experience levels.

Methods

Participants were recruited from the Psychiatry Of Old Age Consultation Liaison Services at Galway University Hospital (GUH) and Sligo University Hospital (SUH) between April 2015 and July 2017. All participants were over the age of 65.

The battery of tests applied included the revised Delirium Rating Scale (DRS-R98), Mini-Mental State Examination, Delirium Motor Subtype Scale, Informant Questionnaire of Cognitive Decline in the Elderly as well as standard clinical assessment, chart review and collateral history [11, 16–18]. Assessments were conducted by raters (MM, LL) specifically trained in the use of the tests included and, to further enhance inter-rater reliability, ratings associated with any uncertainty were discussed and agreed. Patients were assessed at the bedside during the late morning/early afternoon when the anchors of the day are thought to be optimally active. The traditional verbal assessment was completed first followed by the computerised version. Both tests were completed within approximately 10 min of each other. Delirium was diagnosed if one scored ≥ 15 in the DRS-R98 and/or met DSM IV criteria. Dementia was defined as a clear history of meeting DSM-IV criteria (based on all available information including clinical case notes and a collateral history) or Informant Questionnaire on Cognitive Decline in the elderly (IQCODE) score of ≥ 3.5. Comorbid delirium-dementia was defined as the presence of both disorders. Patients with neither delirium nor dementia were designated ‘No Neurocognitive disorder’ (NNCD).

Patients included were those over 65 years of age referred to the inpatient Psychiatry Liaison Service at GUH and SUH and were divided into the following groups: meeting criteria for delirium in the context of background dementia (DMDL), meeting criteria for delirium only (DL), meeting criteria for dementia (DM) and no neurocognitive diagnosis (NNCD). Patients excluded were unconscious/sedated patients (with a reduced Glasgow Coma Scale score), patients who were unable to attempt both bedside tests of attention and patients receiving care in the intensive care unit, high dependency unit and emergency department.

A computerised version of the MBT and a verbal version of the MBT were administered. The application interface was designed to be user-friendly (see Fig. 1) presenting the twelve months on a 21 cm × 14 cm touch sensitive screen and asking the patient to touch the months in reverse order. The application month button design was modelled on the common abbreviated calendar appearance familiar to all age groups.

Due to the exponential number of possible answers (months skipped, months in an incorrect order, premature termination of attempt etc.), interpretation can be complex.
and requires a clear scoring system that removes inter-user and examiner variability (see varying examples in Fig. 2).

We employed a scoring algorithm to allow immediate and convenient interpretation of performance while reducing inter-user grading variability. A recent study identified that patients with delirium appeared to stop earlier in the reverse months progression than patients with dementia and this scoring system aims to take advantage of this differentiating characteristic [19]. The patient’s performance was assessed using a ‘Correct Progression’ scoring system: 1 point was allocated for correctly picking the final month first. Every correct progression after that was allocated 1 point. The patients were allocated a score out of 12 based on the above criteria. A cut-off of < 8 progressions was trialled in both (MBTv and MBTc) to separate inattention from attention. For example: Picking Dec. = 1 point; Dec. to Nov. = 1 point; Nov. to Oct. = 1 point; Sept. to Aug. = 1 point.

Fig. 1 User-friendly app interface

Fig. 2 Sample answers demonstrating the potential variability in responses that require scoring

Statistical analysis was conducted using SPSS. The correlation between the MBTv and MBTc was analysed using Spearman’s rank correlation coefficient with the $p < 0.05$ used as the cut-off for statistical significance. Also, a Bland–Altman plot (scatter plot) was conducted to compare the two measures by plotting the score differences of MBTv and MBTc against their means. Because the two measures were not normally distributed, their natural logarithms were used (ln). In addition, two logistic regression analyses were used with vMBT and cMBT alternatively as predictors of delirium. In these regression analyses, the dependent variable was delirium vs. no delirium and the independent variables were the MBTv or MBTc, age, gender, MMSE score, and previous history of dementia. The Backward Stepwise Wald method was used.

| Patient 1          | Patient 2          | Patient 3          |
|--------------------|--------------------|--------------------|
| December, October  | November, October  |
| September, June    | September, August  |
| July               | September, August  |
| May                | August, May        |
| April              | February, March    |
| January            | April, March       |
|                    | January            |
Results

Demographic data

A total of 75 patients were recruited across both hospital sites. 30 patients were recruited from GUH and 45 patients from SUH. The gender split was as follows: 28 (37.3%) males, 47 (62.7%) females. The mean age of the cohort was 80.3 (SD = 7.61) years old. All patients assessed were able to attempt both bedside tests of attention.

Cognitive subgroups based on clinical criteria

Patients were allocated a clinical diagnosis based on the aforementioned battery of clinical tests; Delirium Rating Scale (DRS-R98), Informant Questionnaire of Cognitive Decline in the Elderly (IQCODE) and DSM IV criteria for dementia and delirium (see also Methods section) and the final subgroup allocations were as follows: Thirty-six participants (n = 36 (48%)) were diagnosed with delirium in the context of an underlying dementia (DMDL), thirteen (n = 13 (17%)) were diagnosed with dementia only (DM), seven (n = 7 (9%)) were diagnosed with delirium only (DL), and nineteen (n = 19 (25%)) participants were considered to be cognitively intact (not meeting criteria for a diagnosis of delirium or dementia) (NCCD).

Scales (MMSE, MBTv, MBTc, DMSS)

The mean MMSE of the total group was 20.4 (SD = 6.6). The mean MMSE scores together with mean age in each subgroup is shown in Table 1. Using a cut-off of < 8 progressions to separate inattention from attention (described in detail in the Methods section), the results comparing MBTv results with the MBTc results per subgroup are described.

Of the total sample, the mean MBTv was 8.1 (SD = 4.6) and the mean MBTc was 6.3 (SD = 4.4).

Table 1  Summary table of mean age, mean MMSE, results from MBTv and MBTc per subgroup

| Sub group | Mean age (SD) | Mean MMSE (SD) | MBTv | MBTc |
|-----------|--------------|---------------|------|------|
| DMDL      | 82 (8)       | 15 (6)        | 10   | 26   | 6   | 30  |
| DM        | 79 (7)       | 21 (5)        | 12   | 1    | 10  | 3   |
| DL        | 81 (9)       | 24 (5)        | 5    | 2    | 3   | 4   |
| CI        | 79 (6)       | 28 (2)        | 19   | 0    | 17  | 2   |

Inattention (< 8 Progressions)

*Attention (= / > 8 Progressions)

Table 2  Spearman correlation of MBTv with MBTc

|          | rho  | p-value |
|----------|------|---------|
| Overall  | 0.772| <0.0001 |
| DMDL     | 0.666| <0.0001 |
| DL only  | 0.778| 0.039   |
| DM only  | 0.378| 0.203   |
| Cog intact| 0.143| 0.559   |

Verbal and computerised test correlation

Given the evidence base for the verbal MBT, if a computerised version of the test correlated well with the verbal version administered by experienced and formally trained assessors, we hypothesise that the computerised MBT may be easier and more consistent to use for both experienced and unexperienced medical professionals alike. To assess this correlation, a Spearman’s correlation test (MBTv and MBTc were normally distributed) was used to calculate the inter-test correlation coefficient (rho). Overall, a statistically significant positive inter-test correlation between the MBTv and MBTc (p < 0.0001) was found. Similarly, statistically significant positive correlations were found for the groups of DMDL and DL group (p < 0.0001). However, poor correlation was found in the two groups without delirium DM and NNCD. (See also Table 2).

Bland–Altman plot

The Bland–Altman plot is depicted in Fig. 3

Figure 3 shows that the two methods are comparable (with the mean of difference being close to 0). It also shows systematic variation with the mean of the two measurements towards to the right which reflects more patients who were cognitively intact than those with previous history of dementia (as described above). The standardised Beta coefficient showed no significant differences between the two measurements (B = –0.219, t = –1.770, p = 0.082).
Logistic regression

The results of the two logistic regression analysis performed for the MBTv and MBTc (cut-off point for both < 8) are illustrated in Tables 3, 4.

From the tables above seems that both MBTv and MBTc are significant predictors of delirium together with MMSE when controlling for other variables.

Discussion

These results imply very strong correlation between the MBTv (done by formally trained experienced assessors) with the MBTc both overall and in the subgroups meeting a clinical diagnosis of delirium (DMDL, DL). There is poor correlation in the dementia only group and in the group with no neurocognitive disorder (DM, NNCD). Given that poor attention is a cardinal feature of delirium and is required in both the verbal and computerised version, it is unsurprising that participants who met the clinical criteria for delirium did poorly in both. When reviewing the poor correlation between the verbal and computerised versions in the dementia group, more participants had less than 8 correct progressions in the computerised version than in the verbal version (3 vs 1; DM subgroup – see Table 1). Visuospatial cognition such as spatial orientation can be impaired in dementia and may partly account for the difference in completion of the computerised version in the dementia subgroup. Psychomotor deficits in group due to underlying neurodegeneration

![Bland–Altman plot](image)

**Table 3** Regression analysis including MBTv

|          | B    | S.E  | Wald | df  | Sig  | Exp (B) |
|----------|------|------|------|-----|------|---------|
| MBTv (inattention) | 2.589 | 1.183 | 4.791 | 1   | 0.029 | 13.313  |
| MMSE score    | −0.174 | 0.076 | 5.316 | 1   | 0.021 | 0.840   |
| Constant      | 3.411  | 1.807 | 3.565 | 1   | 0.059 | 30.309  |

**Table 4** Regression analysis including MBTc

|          | B    | S.E  | Wald | df  | Sig  | Exp (B) |
|----------|------|------|------|-----|------|---------|
| MBTc (inattention) | 1.715 | 0.727 | 5.560 | 1   | 0.018 | 5.556   |
| MMSE score    | −0.195 | 0.074 | 7.053 | 1   | 0.008 | 0.823   |
| Constant      | 3.651  | 1.804 | 4.096 | 1   | 0.043 | 38.499  |
can also result in poor computerised performance [20]. Poor correlation between the verbal and computerised version also exists in the cohort with no neurocognitive disorder (NNCD). In the elderly hospitalised group with NNCD, practical limitations of the computerised tablet noted by the authors were the presence of a tremor (influencing month selection), visual difficulties (cataracts, diabetic retinopathy) and, surprisingly, length of fingernails (obstructing touchscreen utility). The explanation for the poor correlation between the MBTv and MBTc score is likely to be due to the reduced performance in the computerised test due to the aforementioned mechanisms. In addition, the distribution of subjects, high homogeneity and small sample sizes of the groups (in this pilot study) may be another reason for the low correlation of the two tests in the groups without delirium. Given that the correlation between tests was high when numbers were large in the entire sample, a larger follow-on study to this initial pilot study with larger sample sizes in each subgroup is needed to fully assess the utility of this tool.

A possible limitation of this study is that the oral battery of tests (as described previously), including the MBTv, was conducted before the patient attempted the MBTc version; this may have affected the performance in the latter (improved due to the possible benefit of a ‘verbal practice-run’ or disimproved due to patient fatigue); randomization and reversal of the chronology may be of interest in future studies to mitigate the effects of the previously described practice effect and patient fatigue.

Advantages of the tablet was that employment of an algorithm allowed instantaneous interpretation (relative to verbal interpretation) with no inter-user variability. The application’s easy-to-use interface does not require interpretation training and can be used by all health care professionals. A financial review identifying the cost–benefit of the purchase of a moderately priced handheld tablet use vs. reduction in the length of inpatient stay through early identification of deterioration (for example, if used on routine nursing rounds) would also help evaluation of its practical implementation.

As part of the Integrated Care Programme for Older Persons there is an emphasis on the role of integrated computer technology (ICT) in older persons care [21]. Given the visuospatial limitations in the dementia subgroup and the physical limitations in the group with no neurocognitive disorder (tremor, long fingernails) that our study has identified, one could envisage future development of an application based on advancing speech recognition technology that allows a would allow a hybrid voice recognition computerised months backwards test that would circumvent visuomotor problems yet have all the previously mentioned benefits of standardisation.

Conclusions

The MBTc correlates well with the MBTv in patients who are clinically suspected to have delirium but has poor correlation in patients without delirium. Visuospatial cognition and psychomotor deficits in a dementia only cohort and mechanical factors (such as tremor, poor fingernail hygiene and visual impairment) in a group with no neurocognitive disorder may limit the utility of the MBTc in a hospitalised older population. A larger follow-on study to this pilot study is needed to fully evaluate its use.

Author contributions All the authors have contributed in the design of the study, the interpretation of the data and drafting of the manuscript. More specifically, MM and LL were involved in data collection, DM, CE and CD in the development of the software, DA in the analysis of the data and in the overall supervision and coordination of the project.

Funding Open Access funding provided by the IRel Consortium. This research was funded by the Wellcome Trust, 215 Euston Road, London, NW1 2BE, UK.

Declarations

Conflicts of interest The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Ethical approval All the procedures conducted in this study were in accordance with the 1964 Helsinki Declaration and its later amendments. Ethical approval was granted by Research Ethics Committees at Galway University (GUH) and Sligo University Hospitals (SUH).

Informed consent Informed consent was obtained and all participants were asked to partake in an initial battery of tests. The procedures and rationale for the study were explained to all patients and a patient information leaflet was provided to the patient (and/or the next of kin/responsible caregiver where appropriate). Due to level of cognitive impairment, some participants were unable to provide informed consent. In these cases, informed consent was augmented by communicating with their next of kin (where possible) or a responsible caregiver (in accordance with Helsinki Guidelines for Medical research involving Human Participants).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits 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, 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/4.0/.

Springer
References

1. Ryan DJ, O’Regan NA, Caoimh RÓ et al (2013) Delirium in an adult acute hospital population: Predictors, prevalence and detection. BMJ Open 3:1–10. https://doi.org/10.1136/bmjopen-2012-001772

2. Dani M, Owen LH, Jackson TA et al (2018) Delirium, frailty, and mortality: interactions in a prospective study of hospitalized older people. J Gerontol Ser A Biol Sci Med Sci 73:415–418. https://doi.org/10.1093/gerona/glx214

3. Kean J, Ryan K (2008) Delirium detection in clinical practice and research: critique of current tools and suggestions for future development. J Psychosom Res 65:255–259. https://doi.org/10.1016/j.jpsychores.2008.05.024

4. Witlox J, Eurelings LSM, De Jonghe JFM et al (2010) Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. JAMA 304:443–451. https://doi.org/10.1001/jama.2010.1013

5. Association AP (2014) Diagnostic and Statistical Manual of Mental Disorders: DSM-5 (5th edition).

6. World Health Organization (WHO) (1993) The ICD-10 classification of mental and behavioural disorders. World Health Organization

7. Rudolph JL, Jones RN, Grande LJ et al (2006) Impaired executive function is associated with delirium after coronary artery bypass graft surgery. J Am Geriatr Soc 54:937–941. https://doi.org/10.1111/j.1532-5415.2006.00735.x

8. Grober E, Hall CB, Hahn SR et al (2011) Memory impairment and executive dysfunction are associated with inadequately controlled diabetes in older adults. J Prim Care Community Health 2:229–233. https://doi.org/10.11177/2150131911409945

9. Bellelli G, Morandi A, Davis DH et al (2014) Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. Age Ageing 43:496–502. https://doi.org/10.1093/aging/aft021

10. Adams D, Meagher D, Murray O et al (2016) Evaluating attention in delirium: a comparison of bedside tests of attention. Geriatr Gerontol Int 16:1028–1035. https://doi.org/10.1111/egi.12592

11. Trezepac PT, Mittal D, Torres R et al (2001) Validation of the delirium rating scale-revised-98: comparison with the delirium rating scale and the cognitive test for delirium. J Neuropsychiatry Clin Neurosci 13:229–242. https://doi.org/10.1176/jnp.13.2.229

12. Meagher J, Leonard M, Donoghue L et al (2015) Months backward test: a review of its use in clinical studies. World J Psychiatry 5:305. https://doi.org/10.5498/wjp.v5.i3.305

13. O’Hanlon S, O’Regan N, MacLullich AM et al (2014) Improving delirium care through early intervention: from bench to bedside to boardroom. J Neurol Neurosurg Psychiatry 85:207–213. https://doi.org/10.1136/jnnp-2012-304334

14. King D, Brughelli M, Hume P et al (2013) Concussions in amateur rugby union identified with the use of a rapid visual screening tool. J Neurol Sci 326:59–63. https://doi.org/10.1016/j.jns.2013.01.012

15. Donoghue L, Elias T, Leonard M et al (2016) Development of a novel computerised version of the month backwards test: a comparison of performance in hospitalised elderly patients and final year medical students. Comput Biol Med 68:1–8. https://doi.org/10.1016/j.compbiomed.2015.10.010

16. Folstein MF, Folstein SE, McHugh PR (1975) “Mini mental state. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. https://doi.org/10.3744/snajp.2003.40.2.021

17. Meagher D, Moran M, Raju B et al (2008) A new data-based motor subtype schema for delirium. J Neuropsychiatr Clin Neurosci 20:185–193. https://doi.org/10.1176/appi.neuropsych.20.2.185

18. Jorm AF (1996) Assessment of cognitive impairment and dementia using informant reports. Clin Psychol Rev 16:51–73. https://doi.org/10.1016/0272-7358(95)00056-9

19. Hasemann W, Duncan N, Clarke C et al (2021) Comparing performance on the months of the year backwards test in hospitalised patients with delirium, dementia, and no cognitive impairment: an exploratory study. Eur Geriatr Med 12:1257–1265. https://doi.org/10.1007/s41999-021-00521-4

20. Arvanitakis Z, Shah RC, Bennett DA (2019) Diagnosis and management of dementia: review. JAMA 322:1589–1599. https://doi.org/10.1001/jama.2019.4782

21. Health Service Executive (2016) Integrated Care Programme for Older Persons (ICP). https://www.hse.ie/eng/about/who/cspd/icp/older-persons/. Accessed 3 Jan 2022

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