

BMJ Open Diagnostic accuracy of the Ottawa 3DY and Short Blessed Test to detect cognitive dysfunction in geriatric patients presenting to the emergency department

David Barbic,¹ Brian Kim,² Qadeem Salehmohamed,² Kate Kemplin,³ Christopher R Carpenter,⁴ Skye Pamela Barbic^{5,6}

To cite: Barbic D, Kim B, Salehmohamed Q, *et al*. Diagnostic accuracy of the Ottawa 3DY and Short Blessed Test to detect cognitive dysfunction in geriatric patients presenting to the emergency department. *BMJ Open* 2018;**8**:e019652. doi:10.1136/bmjopen-2017-019652

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-019652>).

Received 18 September 2017
Revised 6 February 2018
Accepted 9 February 2018



¹Department of Emergency Medicine, University of British Columbia, Vancouver, British Columbia, Canada

²Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

³School of Nursing, University of Tennessee Chattanooga, Chattanooga, Tennessee, USA

⁴Division of Emergency Medicine, Washington University, St Louis, Missouri, USA

⁵Department of Occupational Therapy and Occupational Science, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

⁶Centre for Health Evaluation Outcome Sciences, University of British Columbia, Vancouver, British Columbia, Canada

Correspondence to
Dr David Barbic;
david.barbic@ubc.ca

ABSTRACT

Objectives Cognitive dysfunction (CD) is a common finding in geriatric patients presenting to the emergency department (ED). Our primary objective was to determine the diagnostic accuracy of the Ottawa 3DY (O3DY) and Short Blessed Test (SBT) as screening tools for the detection of CD in the ED. Our secondary objective was to estimate the inter-rater reliability of these instruments.

Methods We conducted a prospective cross-sectional comparative study at an inner-city academic medical centre (annual ED visit census 86 000). Patients aged 75 years or greater were evaluated for inclusion, 163 were screened, 150 were deemed eligible and 117 were enrolled. The research team completed the O3DY, SBT and Mini-Mental State Exam (MMSE) for each participant. Descriptive statistics were calculated. Sensitivity and specificity of the O3DY and SBT were calculated in STATA V.11.2 using the MMSE as our criterion standard.

Results We enrolled 117 patients from June to November 2016. The median ED length of stay at the time of completion of all tests was 1:40 (IQR 1:34–1:46). The sensitivity of the O3DY was 71.4% (95% CI 47.8 to 95.1), and specificity was 56.3% (46.7–65.9). Sensitivity of the SBT was 85.7% (67.4–99.9) and specificity was 58.3% (48.7–67.8). The receiver operating characteristic area under the curve was calculated for the O3DY (0.51; 95% CI 0.42 to 0.61) and SBT (0.52; 95% CI 0.43 to 0.61) relative to the MMSE. Inter-rater reliability for the O3DY ($k=0.64$) and SBT ($k=0.63$) were good.

Conclusion In a cohort of geriatric patients presenting to an inner-city academic ED, the O3DY and SBT tools demonstrate moderate sensitivity and specificity for the detection of CD. Inter-rater reliability for the O3DY and SBT were good. Future research on this topic should attempt to derive and validate ED-specific screening tools, which will hopefully result in more robust likelihood ratios for the screening of CD in ED geriatric patients.

INTRODUCTION

Background

Population ageing is happening throughout the world with more people living to advanced old age. By 2030, the geriatric population (65 years and older) is expected to reach 24%, 20.3% and 21.5% of the total population in

Strengths and limitations of this study

- This study used rigorous, prospective data collection to test the sensitivity and specificity of the Ottawa 3DY (O3DY) and Short Blessed Test (SBT) tools for screening for cognitive dysfunction in geriatric emergency department patients.
- This is the first study to examine the inter-rater reliability of the O3DY and SBT tools in geriatric emergency department patients.
- This study used research assistants to administer the O3DY, SBT and Mini-Mental State Exam, not geriatric nurses. In suburban, community and rural hospitals, which often do not have access to geriatric nurses, the results of our study may more accurately reflect the diagnostic accuracy of these screening tests for busy clinicians and those not as comfortable caring for the geriatric patient.
- The single-centre nature of this study may limit the generalisability of our results.

Canada, the USA and Europe, respectively.^{1–3} In Canada, an estimated 16.9% of Canadians are aged 65 years or older. Geriatric patients are increasingly frequent visitors to the emergency department (ED) due to their complex and multiple comorbidities, limited same-day access to primary care or requisite diagnostic resources, caregiver stress and limited transportation to physician offices.^{4–8} Age-related changes manifest as geriatric syndromes like cognitive dysfunction (CD) which further complicate timely evaluation of acute complaints in the busy ED. Most elderly individuals presenting to the ED with CD live at home and have not been previously diagnosed with CD.⁹

CD is not a specific diagnosis, rather it is a group of symptoms which includes cognitive impairment, delirium and various stages of dementia.¹⁰ Changes in mental status can be attributed to electrolyte abnormalities,¹¹ cerebral hemisphere pathologies and

neurological abnormalities, hypoxic/anoxic states, polypharmacy and medication interactions¹² and cardiovascular pathologies.¹³

CD is a common diagnostic challenge faced by clinicians in the ED, with delirium and dementia present in 10%–40% of ED geriatric patients.^{14–16} The true prevalence of CD among these populations is unclear since CD includes an array of symptoms such as memory loss, disorientation and so on. Therefore, clinicians may face difficulty in collecting an accurate history.¹⁴ These delays likely will adversely impact ED performance measures such as length of stay (LOS) and adverse events experienced by individual patients.¹⁷ CD is also negatively associated with repeat visits to the ED,¹⁸ hospitalisation and death.^{17 19 20}

CD in this population can be difficult for ED clinicians to recognise and is infrequently assessed by ED providers.²¹ Consequently, geriatric competencies for emergency medicine residents have been developed that include core aspects of identifying and treating CD in the ED.²² In addition, geriatric ED guidelines which include the formal assessment of CD have been developed and endorsed by emergency medicine associations in the USA and Canada²³ as well as ED-specific, field-tested quality indicators for process and structural design in Australia.^{24 25}

The unpredictability of the ED setting presents notable challenges to the assessment of CD. Multiple barriers to accurately identifying CI in the ED include the absence of privacy, excessive noise, clinicians' time constraints and a lack of validated screening tools.²⁶ ED clinicians are poor detectors of CD in the elderly, failing to identify up to 80% of cases.^{7 21 27} Reducing error in this type of assessment is critical. Patients with CD and dementia experience frequent transitions of care,²⁸ and an objective, quantifiable screen for CD or dementia would allow ED providers to review current cognitive test performance with the patients' primary care provider via telephone or the electronic medical record. Further, when CD is undetected by ED clinicians, such diagnostic omissions may continue to admitting physicians who may also fail to detect CD.²⁹ There is no evidence that post-ED dementia screening interventions can improve patient outcomes, but there is ample data implying an opportunity to improve outcomes. Seventy to eighty per cent of ED patients with dementia had no history of dementia, so ED clinicians, hospitalists and geriatricians cannot rely on patient history or reporting.³⁰ Inpatient physicians under-recognise dementia³¹ and lack confidence in the management of persons with dementia.³² As a result, the failure to detect CD, combined with relationships between CD and poor patient outcomes, highlights the need to improve methods to screen for CD in geriatric populations presenting to the ED.^{15 27 33}

Goals of this investigation

Over 40 screening tools exist to screen for CD.^{22 30 33} Many of these instruments have been critiqued for not

being conducive for use in the ED due to their length, complexity of scoring or lack of integration into electronic medical records or have never been evaluated in the ED setting.³⁵ Two recently studied short assessments are the Ottawa 3DY (O3DY) and Short Blessed Test (SBT). Little evidence exist to estimate the extent to which these measures are fit for purpose to measure CD in a Canadian ED setting. As a result, the primary objective of this study was to determine the diagnostic accuracy of the O3DY and SBT as screening tools for the detection of CD in an ED setting. Additionally, our secondary objective was to estimate the inter-rater reliability of the instruments to understand how robust they stand as unambiguous measurement tools for ED clinicians.

METHODS

Study setting and design

This prospective, cross-sectional, convenience sampling study was performed over 5 months in 2016 at an urban, inner-city academic ED with an annual census of 86 600 visits, of which an estimated 16.7% of patients are older than 65 years. Ethical approval for this study was granted by the Research Ethics Board of the Providence Healthcare Research Institute. This study adhered to the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guidelines for the conduct of diagnostic test accuracy studies (see online supplementary appendix 1).³⁶

Participant characteristics and sampling

Over a 5-month span (June to November 2016) research assistants systematically screened patients for participation in this study if they met the inclusion criteria of being aged 75 or older and presented to the ED on Monday to Friday between 9 am and 4 pm. Our exclusion criteria were similar to the study by Wilding *et al* in an effort to identify and enrol a similar patient population.³⁷ We excluded patients who were: triaged as Canadian Emergency Department Triage and Acuity Scale level 1 (resuscitation), if their condition was deemed too critical for evaluation, patients requiring emergent ED administration of medications which might negatively affect their neurological and/or executive function (eg, opioids, benzodiazepines), patients with significant communication barriers affecting evaluation (eg, visual, verbal or auditory impairments), patients with overt hallucinations, agitation or confusion, patients who did not speak English, patients from nursing homes or long-term care facilities, patients with a previous diagnosis of cognitive impairment (eg, patients with dementia), patients already enrolled in the study and patients unable to provide full, written, informed consent in English. No incentives for participation were offered.

Sample size

Prior work has demonstrated prevalence rates of 13.4%–37% for geriatric patients with CD presenting to the

ED.^{16 37 38} To detect a screening tool sensitivity of 98.0%, with a 30% prevalence rate for CD in our study population and an alpha of 0.05, we required a minimum sample size of 101 patients in this study.³⁹

Training and reliability of data collectors

The study protocol was executed by research assistants who were trained by an experienced staff emergency physician (DB). Physicians and other test administrators were familiar with the tools, their administration and the non-physician test administrators were familiarised with the process thoroughly by the Principal Investigator (PI). Preparation of the research assistants was ensured by institutional research methods and ethics training under a faculty advisor supervisor. This included 4 hours of in person, then observed administration of the study tools and weekly quality assurance communication. Research assistants had formal training in the ethical conduct of research as mandated by our local institutional review board and in the procedures outlined below.

Data collection methods and management

All data were immediately entered into an Excel spreadsheet (Microsoft) at the bedside on portable tablet devices (iPad Air, Apple). All data were stored in a secure offline database accessible to the primary investigator and to the research assistants.

Screening instruments

The O3DY was derived from the Canadian Study of Health and Aging, and subjects with severe dementia, non-English speakers and those with vision or hearing impairment were excluded. The criterion standard for dementia was a consensus of neuropsychologist, nurse and physician using bedside screening instruments, historical information, physical examination and normative data for the population. Variables for the O3DY were abstracted from the modified Mini-Mental State Exam (MMSE).¹⁶ The O3DY (day, date, dlrow ('world' spelled backward), year) is a four-question instrument designed to assess attention and orientation.⁴⁰ If patients do not correctly answer all four questions, it is considered to be a positive test for CD (see online supplementary appendix 2) The O3DY is advantageous for ED use because it takes minimal time to perform and is easy to remember and score. It does not require pen or paper, and we assume that physicians are able to remember four short phrases of an assessment. Therefore, we did not perform an additional assessment of physician recall.

The SBT originated from the Blessed Mental Status Test and was validated by Katzman *et al* in geriatric community care settings, with further implementation in ED settings.³⁸ The SBT is a weighted six-item instrument to evaluate orientation, registration and attention¹⁶ (see online supplementary appendix 2). The SBT was originally validated on patients in a skilled nursing facility and with active community-dwelling elderly. The SBT has demonstrated excellent reliability.¹⁶ A weighted error

score of more than four constitutes an abnormal result. The SBT is advantageous for ED use because it takes less than 2 min to complete.^{33 37} Despite this, the complicated scoring mechanism for the SBT limits its clinical feasibility in the ED.³⁷ For the SBT, we used a cut-off of >4 out of a maximum weighted error score of 28 to indicate CD. Scores >4 on the SBT have previously been found to correlate with questionable impairment.⁴¹

Similar to prior studies on this topic, the MMSE⁴² was our criterion standard against which we tested the O3DY and SBT (see online supplementary appendix 2).^{37 42} The MMSE assesses five areas of cognitive function: orientation, memory, language, attention and visuospatial. Similar to prior work,^{37 42} a score of ≤ 24 out of 30 was designated as indicative of CI. The MMSE is not regularly used by ED clinicians to screen for CD since it requires too much time, it requires patients to have their glasses, writing materials and a free arm (unobstructed by intravenous infusions) to write with.^{43 44} Furthermore, the MMSE is now copyrighted and requires a fee for its use.³⁵ The MMSE was chosen over the Montreal Cognitive Assessment (MoCA) due to a limited number of studies assessing cognition in the ED with the MoCA and none involving the O3DY and SBT.⁴⁵

Procedures

For all patients enrolled in this study, research assistants administered the O3DY first, followed by the SBT and MMSE (see online supplementary appendix 3). For a subset of 10% of patients, two research assistants administered the tests twice within 1 hour to calculate inter-rater reliability. All test administration occurred in the ED during the enrolment periods previously described. All the patients screened received routine care in the ED. Treating emergency physicians and nurses were blinded to the results of our study. Clinical information and reference standard results were not available to the performers/readers of the index test; clinical information and index test results were not available to the assessors of the reference standard.

Statistical analyses

Descriptive statistics using counts, medians, means and IQR were calculated to describe the cohort. Sensitivities, specificities, likelihood ratios, predictive values and per cent agreement for the O3DY and SBT compared with the MMSE were calculated in STATA V.11.2. Inter-rater reliability was calculated using Cohen's Kappa method for the O3DY and SBT.⁴⁶

RESULTS

Between June and November 2016, we approached 164 patients who met our inclusion criteria (figure 1), and 117 were enrolled. Table 1 shows the demographics of enrolled participants. The median age was 81.9 years (IQR 77–85), and 44.8% were female. The most common presenting complaint and comorbidity were

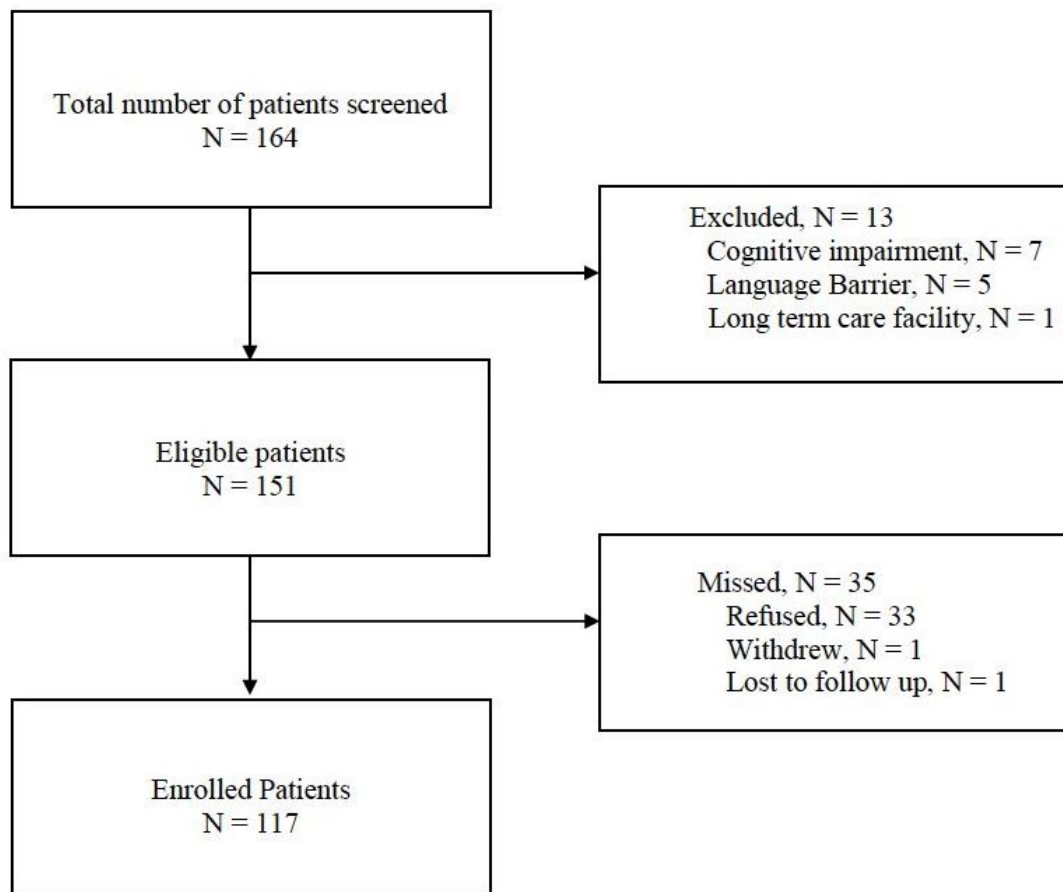


Figure 1 STARD flow diagram of patients in the study.

cardiac (13.8%) and hypertension (57.8%), respectively. The median length of stay in the ED at the time of testing completion was 1:40 (IQR 1:34–1:46). The characteristics of participants who were eligible, yet who declined to participate, were similar to the study population.

The prevalence of CD among enrolled participants based on an MMSE <24 was 12.0% (95% CI 6.1% to 17.9%). The O3DY and MMSE agreed in 58.1% of cases. The sensitivity and specificity of the O3DY was 71.4% (95% CI 47.8% to 95.1%) and 56.3% (95% CI 46.7% to 65.9%), respectively. The SBT agreed with the MMSE in 61.5% of cases. The SBT had a sensitivity of 85.7% (95% CI 67.8% to 100%) and a specificity of 58.3% (95% CI 48.7% to 67.8%) (table 2).

The receiver operating characteristic area under the curve was calculated for the O3DY (0.51; 95% CI 0.42 to 0.61) and SBT (0.52; 95% CI 0.43 to 0.61) relative to the MMSE (figure 2).

Inter-rater reliability for the O3DY and SBT were calculated for a subset of 9.4% of participants (n=11). Cohen's Kappa for the interpretation of 'normal versus abnormal' of the O3DY test was $k=0.64$ (95% CI 0.18 to 1.00) and $k=0.63$ (95% CI 0.15 to 1.00) for the SBT were good (table 3).

CONCLUSIONS

In this prospective sample of geriatric patients presenting to a Canadian inner-city academic ED, we measured the diagnostic accuracies of the O3DY and SBT tests for rapid screening for CD compared with the criterion standard of the MMSE. The O3DY and SBT both displayed moderate sensitivity and specificity for detecting CD, but the SBT performed superiorly to the O3DY. Our study is the first Canadian ED study to assess the diagnostic accuracy of these tests when administered by research assistants and not geriatrics nurses. Our study is also notable since it is the first prospective ED study to report the inter-rater reliability of the O3DY and SBT for screening for CD in the ED, for which we report good agreement between assessors.⁴⁷ The results of our study provide further evidence that these tests lack sufficient sensitivity and specificity to screen for CD in geriatric patients presenting to the ED.

The O3DY and SBT tests have been previously studied in geriatric patients in the ED. In a study of patients aged 75 years or older at two Canadian academic EDs, the O3DY demonstrated a sensitivity of 93.8% and specificity of 72.8%.³⁷ In a similar study of patients aged 65 years or older from an academic ED in the USA, the O3DY demonstrated a sensitivity of 95% and specificity of 51%; the sensitivity and specificity of the SBT was 95% and 65%, respectively.¹⁶ The results from our study show lower

Table 1 Patient characteristics

Characteristic	Enrolled patients (n=117)
Female gender (%)	52 (44.8)
Mean age (SD)	81.9 (5.7)
Median age (IQR)	81 (8)
Presenting complaint (%)	
Cardiac	16 (13.8)
Minor injuries	15 (12.9)
GI	13 (11.2)
Pain	9 (7.8)
Minor ailments	9 (7.8)
Undifferentiated chest pain	8 (6.9)
Respiratory	8 (6.9)
Swelling	8 (6.9)
Dizziness/vertigo	7 (6.0)
Urinary	7 (6.0)
Miscellaneous	6 (5.2)
Weakness	5 (4.3)
Other	5 (4.3)
Pre-existing conditions (%)	
Hypertension	67 (57.8)
Diabetes	18 (15.5)
CHF	6 (5.2)
Pacemaker	8 (6.9)
Coronary artery disease/MI	31 (26.7)
Stroke/CVA	6 (5.2)
Cancer	10 (8.6)
COPD	12 (10.3)
Mean length of stay at time of study	1:40

CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident; GI, gastrointestinal; MI, myocardial infarction.

sensitivity and specificity for the O3DY and SBT compared with the MMSE as a criterion standard. One possible explanation for this discrepancy is both prior studies^{16,37} integrated the O3DY and the MMSE and administered them as one test. While this avoids the problem of recall bias, it may cause incorporation bias which can falsely increase the test's sensitivity.⁴⁸ We administered the O3DY, SBT and MMSE consecutively, which could have

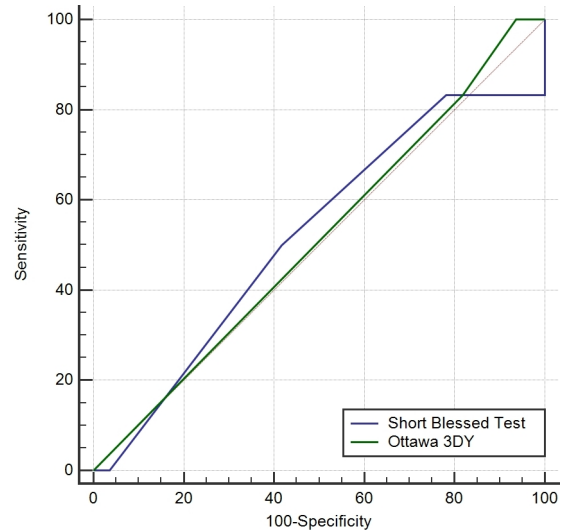


Figure 2 ROC curve of O3DY and SBT compared with MMSE. MMSE, Mini-Mental State Exam; O3DY, Ottawa 3DY; ROC, receiver operating characteristic; SBT, Short Blessed Test.

resulted in response fatigue due to a carry-over effect.⁴⁹ This may have lowered scores on the MMSE. However, consecutive testing may also have introduced recall bias, leading to practice effects and improved MMSE scores.⁵⁰ Another potential explanation for the different results of the three studies is the different rates of CD, as defined by MMSE, across the three studies (our study 12%, Wilding *et al* 13.4%, Carpenter *et al* 37%).^{16,37} The patient populations included in the three studies also differed. A further explanation is that the ages of enrolment (65 vs 75 years) and the cut-off score for the definition of CD with the MMSE differed between the three studies. It is difficult to determine the direction or magnitude of effect these two differences may have on the final results of each study. Another key difference that might explain our study results is the median ED LOS for patients included in our study was 1 hour and 40 min, notably less than the 9 hours and 54 min observed in the study by Wilding *et al*. This raises the possibility that the study by Wilding *et al* detected both incident and prevalent CD in their study population.

A notable difference between the three studies is that our study and the study by Carpenter *et al* used trained research assistants to enrol patients, perform the screening tests and collect data.¹⁶ In the study by Wilding *et al*, ED geriatric nurses completed these tasks.³⁷ ED geriatric nurses receive extensive training on caring for the

Table 2 Diagnostic statistics of the O3DY and SBT

Test	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio	Negative likelihood ratio
O3DY	71 (0.48 to 0.95)	56 (0.47 to 0.66)	1.63 (1.10–2.43)	0.51 (0.22–1.18)
SBT	86 (0.67 to 1.00)	58 (0.49 to 0.68)	2.05 (1.50–2.81)	0.25 (0.07–0.89)

O3DY, Ottawa 3DY; SBT, Short Blessed Test.

Table 3 Inter-rater reliability of the O3DY and SBT

Test	Percent agreement	Kappa (95% CI)
O3DY	83.3	0.64 (0.18 to 1.00)
SBT	83.3	0.63 (0.15 to 1.00)

O3DY, Ottawa 3DY; SBT, Short Blessed Test.

elderly and screening for CD. The difference observed in our study suggests that in academic EDs where geriatric nurses are more common, the diagnostic accuracy of these two cognitive screening tests may be enhanced. In suburban, community and rural hospitals, which often do not have access to geriatric nurses, the results of our study and those of Carpenter *et al* may more reflect the diagnostic accuracy of these screening tests for busy clinicians and those not as comfortable caring for the geriatric patient.¹⁶

Our study only included patients able to provide full, written informed consent in English, and this was supported by the Research Ethics Board approval at our institution. However, this raises the important challenge that those patients identified as displaying CD through the screening tests in our study may have only been capable of providing informed assent—a key ethical distinction.⁵¹ The systematic exclusion of patients with CD may contribute to the dearth of evidence for this vulnerable population and make the provision of evidence-based care by emergency clinicians even more challenging.⁵²

These screening tests (O3DY and SBT) cannot determine the aetiology of CD which is required to appropriately manage these patients, yet the role of an initial screening test is to highlight an abnormality. The definitive aetiology of the underlying problem causing this result can be identified through focused physical examination, laboratory and radiological testing.^{14 29} Our study did not assess whether the ED management of patients or their outcomes would change with the application of these cognitive screening tests. Prior work has demonstrated that ED management does not change when CD is identified in research settings,⁵³ so developing accurate and reliability CD screening instruments is only the first step towards improving outcomes for these potentially vulnerable patients. The failure to adapt ED care when CD was identified was noted in a single-centre study almost two decades ago and may have been due to a quality of care issue in which physicians did not realise that CD was a potentially serious and ED relevant medical problem, or physicians in this study may not have trusted the validity of the screening tests. However, the move to develop a more patient-centric approach to ED geriatric patients and the evaluation of CD in the ED have evolved rapidly in the last 15 years.^{22 23} Screening for CD in ED geriatric patients has until recently been subject to the trade-off between tool validity and applicability in the ED setting. The priority for future research on this topic is to derive and validate ED specific screening tools using modern psychometric

techniques and an iterative process which will hopefully result in more robust likelihood ratios for the screening of CD in ED geriatric patients. Modern psychometric methods are now commonly used across health sectors to evaluate the extent to which screening tools are fit for purpose for the context of use.^{54–57} Unlike classical psychometric methods, these new methods provide critical information to describe the extent to which (a) the full construct is captured, (b) the items target the population under investigation, (c) is unidimensional (measures only one construct) and (d) produces a total score that is robust and clinically meaningful.

There are important limitations to our study. First, our study lacked a gold criterion standard. Ideally, a comprehensive interview with short functional screening testing would have been possible.⁵⁸ Due to a limited study budget, this was not feasible for our study. As well, the busy ED environment would make such a comprehensive assessment nearly impossible. Consequently, we used the MMSE as our criterion standard.⁵⁹ As a result, our study may be subject to imperfect gold standard bias, which may have resulted in artificial estimates of the diagnostic accuracy of the screening tools tested in this study.^{48 60} We also recognise that despite the MMSE being widely accepted in the ED, the lack of a criterion standard may have led to misclassification bias.⁶⁰ In addition, there is evidence that the MMSE displays high rates of false positive results for CD in elderly patients with low education achievement and different ethnic backgrounds.^{61–65} A further limitation is that we enrolled a convenience sample of patients when research assistants were available. Due to time restrictions, data collection was limited to 7 hours per day on weekdays. This may have increased the risk that our study is subject to selection bias.^{60 66} Another possible limitation is that we did not specifically power this study for the determination of inter-rater reliability of the study tools being tested.⁶⁷ Consequently, our estimates of inter-rater reliability are likely under powered. Finally, another possible limitation of our study was that the application of the three tests to each participant was not randomised. This may have introduced the possibility of practice effects due to repeat testing.⁵⁰ However, the direction and magnitude of this potential bias is difficult to determine since a recent meta-analysis of the neuropsychological assessment literature demonstrated older test subjects, those with longer retest intervals and those in clinical settings (compared with those in non-clinical settings) demonstrated the least benefit of practice effects on repeat test scores.⁶⁸

In a prospective cohort of geriatric patients presenting to an inner-city academic ED, our study demonstrates that the O3DY and SBT have moderate sensitivity and specificity compared with the MMSE, and the internal inter-rater reliability of these two screening tools was moderate. The results of our study provide further evidence that these tests lack sufficient sensitivity and specificity to screen for CD in geriatric patients presenting to the ED.

Acknowledgements The authors would like to thank all of the patients who gave of their time to participate in this study, the Geriatrics Liaison Nurses of St Paul's Hospital in their assistance planning this study and the Department of Emergency Medicine and St Paul's Hospital Foundation for their support during the conduct of this study.

Contributors DB, BK, QS and SPB conceived and designed the study. BK and QS enrolled participants and collected data. DB, BK, QS, SPB, KK and CC analysed the data. All authors made substantial contributions to the final manuscript.

Funding This study was funded by the St Paul's Hospital Department of Emergency Medicine and the St Paul's Hospital Foundation.

Disclaimer The funders of this study (Department of Emergency Medicine and St Paul's Hospital Foundation) did not participate in study conduct, have access to data or have influence over the final analysis.

Competing interests None declared.

Patient consent Not required.

Ethics approval Providence Health Care Research Institute Ethics Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Unpublished data are not available at this time.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Ferrucci L, Giallauria F, Guralnik JM. Epidemiology of aging. *Radiol Clin North Am* 2008;46:643–52.
- Cusack BKK. Surfing the “silver tsunami”: National Black Nurses Association News. Vol 262016.
- CANSIM. Projected population, by projection scenario, age and sex, as of July 1, Canada, provinces and territories. <http://www5.statcan.gc.ca/cansim/a26?lang=eng&id=520005> (accessed 7 Feb 2017).
- Samaras N, Chevalley T, Samaras D, et al. Older patients in the emergency department: a review. *Ann Emerg Med* 2010;56:261–9.
- Berua BRF. *Waiting your turn: wait times for health care in Canada*. Canada: Fraser Institute, 2016.
- Information CifH. *Health care in Canada, 2012: a focus on wait times*. Ottawa, Canada: Canadian Institute for Health Information 2012.
- Schumacher JG. Emergency medicine and older adults: continuing challenges and opportunities. *Am J Emerg Med* 2005;23:556–60.
- Statistics Canada GoC. Access to a regular medical doctor. 2014:82–625 (accessed 3 Jul 2017).
- Schnitker LM, Beattie ER, Martin-Khan M, et al. Characteristics of older people with cognitive impairment attending emergency departments: a descriptive study. *Australas Emerg Nurs J* 2016;19:118–26.
- Xiao HY, Wang YX, Xu TD, et al. Evaluation and treatment of altered mental status patients in the emergency department: life in the fast lane. *World J Emerg Med* 2012;3:270–7.
- Nelson JM, Robinson MV. Hyponatremia in older adults presenting to the emergency department. *Int Emerg Nurs* 2012;20:251–4.
- Mistovich JJ, Krost WS, Limmer DD. Beyond the basics: interpreting altered mental status assessment findings. *EMS Mag* 2008;37:90–7.
- Thiruganasambandamoorthy V, Wells GA, Hess EP, et al. Derivation of a risk scale and quantification of risk factors for serious adverse events in adult emergency department syncope patients. *CJEM* 2014;16:120–30.
- Han JH, Bryce SN, Ely EW, et al. The effect of cognitive impairment on the accuracy of the presenting complaint and discharge instruction comprehension in older emergency department patients. *Ann Emerg Med* 2011;57:662–71.
- Litovitz GL, Hedberg M, Wise TN, et al. Recognition of psychological and cognitive impairments in the emergency department. *Am J Emerg Med* 1985;3:400–2.
- Carpenter CR, Bassett ER, Fischer GM, et al. Four sensitive screening tools to detect cognitive dysfunction in geriatric emergency department patients: brief Alzheimer's screen, short blessed test, Ottawa 3DY, and the caregiver-completed AD8. *Acad Emerg Med* 2011;18:374–84.
- Sklar DP, Crandall CS, Loeliger E, et al. Unanticipated death after discharge home from the emergency department. *Ann Emerg Med* 2007;49:735–45.
- McCusker J, Healey E, Bellavance F, et al. Predictors of repeat emergency department visits by elders. *Acad Emerg Med* 1997;4:581–8.
- Wofford JL, Loehr LR, Schwartz E. Acute cognitive impairment in elderly ED patients: etiologies and outcomes. *Am J Emerg Med* 1996;14:649–53.
- Kakuma R, du Fort GG, Arsenault L, et al. Delirium in older emergency department patients discharged home: effect on survival. *J Am Geriatr Soc* 2003;51:443–50.
- Hustey FM, Meldon SW. The prevalence and documentation of impaired mental status in elderly emergency department patients. *Ann Emerg Med* 2002;39:248–53.
- Hogan TM, Losman ED, Carpenter CR, et al. Development of geriatric competencies for emergency medicine residents using an expert consensus process. *Acad Emerg Med* 2010;17:316–24.
- American College of Emergency Physicians SfaEM, The American Geriatrics Society, Emergency Nurses Association. *Geriatric emergency department guidelines*: American College of Emergency Physicians SfaEM, The American Geriatrics Society, Emergency Nurses Association, 2013.
- Schnitker LM, Martin-Khan M, Burkett E, et al. Process quality indicators targeting cognitive impairment to support quality of care for older people with cognitive impairment in emergency departments. *Acad Emerg Med* 2015;22:285–98.
- Schnitker LM, Martin-Khan M, Burkett E, et al. Structural quality indicators to support quality of care for older people with cognitive impairment in emergency departments. *Acad Emerg Med* 2015;22:273–84.
- Kennelly SP, Morley D, Coughlan T, et al. Knowledge, skills and attitudes of doctors towards assessing cognition in older patients in the emergency department. *Postgrad Med J* 2013;89:137–41.
- Elie M, Rousseau F, Cole M, et al. Prevalence and detection of delirium in elderly emergency department patients. *CMAJ* 2000;163:977–81.
- Callahan CM, Tu W, Unroe KT, et al. Transitions in care in a nationally representative sample of older Americans with Dementia. *J Am Geriatr Soc* 2015;63:1495–502.
- Han JH, Zimmerman EE, Cutler N, et al. Delirium in older emergency department patients: recognition, risk factors, and psychomotor subtypes. *Acad Emerg Med* 2009;16:193–200.
- Carpenter CR, DesPain B, Keeling TN, et al. The six-item screener and ad8 for the detection of cognitive impairment in geriatric emergency department patients. *Ann Emerg Med* 2011;57:653–61.
- Power C, Duffy R, Bates H, et al. The detection, diagnosis, and impact of cognitive impairment among inpatients aged 65 years and over in an Irish general hospital: a prospective observational study. *Int Psychogeriatr* 2017;29:1879–88.
- Helm L, Balzer K, Behncke A, et al. Patients with dementia in acute care hospitals: a cross-sectional study of physicians' experiences and attitudes. *Z Gerontol Geriatr* 2017.
- Gerson LW, Counsell SR, Fontanarosa PB, et al. Case finding for cognitive impairment in elderly emergency department patients. *Ann Emerg Med* 1994;23:813–7.
- Cullen B, O'Neill B, Evans JJ, et al. A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry* 2007;78:790–9.
- Carpenter CR. Evidence-based emergency medicine/rational clinical examination abstract. Does this patient have dementia? *Ann Emerg Med* 2008;52:554–6.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
- Wilding L, Eagles D, Molnar F, et al. Prospective validation of the Ottawa 3DY scale by geriatric emergency management nurses to identify impaired cognition in older emergency department patients. *Ann Emerg Med* 2016;67:157–63.
- Katzman R, Brown T, Fuld P, et al. Validation of a short orientation-memory-concentration test of cognitive impairment. *Am J Psychiatry* 1983;140:734–9.
- Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. *Emerg Med J* 2003;20:453–8.
- Molnar FJ, Wells GA, McDowell I. The derivation and validation of the Ottawa 3D and Ottawa 3DY three- and four-question screens for cognitive impairment. *Clinical Medicine: Geriatrics* 2008;2:1–11.
- Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and

- neuropsychological assessment of Alzheimer's disease. *Neurology* 1989;39:1159–65.
42. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–98.
 43. Carpenter CR, Griffey RT, Stark S, et al. Physician and nurse acceptance of technicians to screen for geriatric syndromes in the emergency department. *West J Emerg Med* 2011;12:489–95.
 44. Dyer AH, Foley T, O'Shea B, et al. Cognitive assessment of older adults in general practice: the collateral history. *Ir J Med Sci* 2017.
 45. Ouellet MC, Sirois MJ, Beaulieu-Bonneau S, et al. Is cognitive function a concern in independent elderly adults discharged home from the emergency department in Canada after a minor injury? *J Am Geriatr Soc* 2014;62:2130–5.
 46. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012;22:276–82.
 47. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
 48. Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. *Acad Emerg Med* 2013;20:1194–206.
 49. T.J.C. Carryover effects in clinical research. *Human experimentation*. Dordrecht: Springer, 1999.
 50. McCaffrey RJ, Duff K, Westervelt HJ. *Practitioner's guide to evaluating change with intellectual assessment instruments*. New York: Kluwer Academic, Plenum Publishers, 2000.
 51. Slaughter S, Cole D, Jennings E, et al. Consent and assent to participate in research from people with dementia. *Nurs Ethics* 2007;14:27–40.
 52. Prusaczyk B, Cherney SM, Carpenter CR, et al. Informed consent to research with cognitively impaired adults: transdisciplinary challenges and opportunities. *Clin Gerontol* 2017;40:63–73.
 53. Hustey FM, Meldon SW, Smith MD, et al. The effect of mental status screening on the care of elderly emergency department patients. *Ann Emerg Med* 2003;41:678–84.
 54. Barbic SP, Cano SJ. The application of rasch measurement theory to psychiatric clinical outcomes research. *BJPsych Bull* 2016;40:243–4.
 55. Cleanthous S, Isenberg DA, Newman SP, et al. Patient Uncertainty Questionnaire-Rheumatology (PUQ-R): development and validation of a new patient-reported outcome instrument for systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) in a mixed methods study. *Health Qual Life Outcomes* 2016;14:33.
 56. Klassen AF, Cano SJ, Alderman A, et al. The BODY-Q: a patient-reported outcome instrument for weight loss and body contouring treatments. *Plast Reconstr Surg Glob Open* 2016;4:e679.
 57. Klassen AF, Cano SJ, Grotting JC, et al. FACE-Q eye module for measuring patient-reported outcomes following cosmetic eye treatments. *JAMA Facial Plast Surg* 2017;19:7–14.
 58. Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* 2004;256:240–6.
 59. Barron EA, Holmes J. Delirium within the emergency care setting, occurrence and detection: a systematic review. *Emerg Med J* 2013;30:263–8.
 60. Mower WR. Evaluating bias and variability in diagnostic test reports. *Ann Emerg Med* 1999;33:85–91.
 61. Escobar JI, Burnam A, Karno M, et al. Use of the Mini-Mental State Examination (MMSE) in a community population of mixed ethnicity. Cultural and linguistic artifacts. *J Nerv Ment Dis* 1986;174:607–14.
 62. Fillenbaum G, Heyman A, Williams K, et al. Sensitivity and specificity of standardized screens of cognitive impairment and dementia among elderly black and white community residents. *J Clin Epidemiol* 1990;43:651–60.
 63. Anthony JC, LeResche L, Niaz U, et al. Limits of the 'Mini-Mental State' as a screening test for dementia and delirium among hospital patients. *Psychol Med* 1982;12:397–408.
 64. Crum RM, Anthony JC, Bassett SS, et al. Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA* 1993;269:2386–91.
 65. Khan F, Tadros G. Complexity in cognitive assessment of elderly British minority ethnic groups: cultural perspective. *Dementia* 2014;13:467–82.
 66. Rutjes AW, Reitsma JB, Di Nisio M, et al. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469–76.
 67. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257–68.
 68. Calamia M, Markon K, Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *Clin Neuropsychol* 2012;26:543–70.