# The Derivation and Validation of the Ottawa 3D and Ottawa 3DY Three- and Four-Question Screens for Cognitive Impairment

Frank J. Molnar[1,2,3,4,6], George A. Wells[5] and Ian McDowell[5]

[1]The CT Lamont Centre, Élisabeth-Bruyère Research Institute, Ottawa. [2]The Division of Geriatric Medicine, Department of Internal Medicine, University of Ottawa. [3]The Clinical Epidemiology Program, the University of Ottawa Health Research Institute, Ottawa. [4]DementiaNET—a CIHR New Emerging Team. [5]The Department of Epidemiology and Community Medicine, University of Ottawa. [6]Regional Geriatric Program of Eastern Ontario.

## Abstract

**Objective:** To derive and validate simple screens of 2 to 4 questions to identify cognitive impairment.

**Design:** Cross-sectional database analysis.

**Setting and Participants:** Community-dwelling participants in the Canadian Study of Health and Aging with mild to moderate cognitive impairment (N = 958), or normal cognition (N = 602).

**Measurements:** Cognitive questions not requiring paper, pen, cue cards, props or more than 30 seconds to answer were selected from the Modified Mini-Mental State Examination (3MS). A sequential weighting approach was applied to logistic regression analyses to create scales of equally-weighted questions using the first 2, 3 or 4 questions from the regression equations.

Sensitivities and specificities were calculated for all cutoffs. Two sets of questions, which approximated the psychometric properties of the 3MS, were validated in a second database.

**Results:** The two tests whose properties approached those of the 3MS (sensitivity 84%, specificity 62%) were the Ottawa **3D** test: **D**ay, **D**ate, **D**LROW (sensitivity 76%, specificity 62%) and the Ottawa **3DY** test: **D**ay, **D**ate, **D**LROW, **Y**ear (sensitivity 80%, specificity 61%).

**Conclusions:** The Ottawa 3D and 3DY tests show promising psychometric properties and are easy enough to employ to promote widespread use but must be revalidated in the target groups for which they are intended.

The full potential value of the Ottawa 3D and 3DY tests can best be understood in the context of '*serial trichotomization*' cognitive screening or case-finding algorithms. Dementia researchers focusing on biomarkers and neuroimaging should consider similar trichotomization approaches as these may improve the sensitivities and specificities of their tests.

**Keywords:** cognitive impairment, dementia, screen, alzheimer

## The Derivation and Validation of the Ottawa 3D and Ottawa 3DY Three- and Four-Question Screens for Cognitive Impairment

Cognitive impairment is a common presenting symptom of a number of increasingly prevalent conditions such as dementia, delirium and, occasionally, depression. It is predicted that the prevalence of dementia in North America will increase from 3.4 million people in 2001 to 5.1 million in 2020 and in Western Europe will increase from 4.9 million in 2001 to 6.9 million in 2020.[1] Worldwide it is estimated that 24.3 million people currently suffer from dementia and that, with an estimated 4.6 million new cases every year, the prevalence will increase to 42.3 million in 2020.[1] Depression and delirium can also be anticipated to rise correspondingly as persons with dementia are at increased risk for these conditions.

**Correspondence:** Dr. F.J. Molnar, Associate Professor of Medicine and Director of Research, University of Ottawa Division of Geriatric Medicine, The Ottawa Hospital, Civic Campus, 1053 Carling Ave., Ottawa Ontario, Canada K1Y 4E9. Tel: (613) 798-5555; Ext: 16486; Fax: (613) 761-5334; Email: fmolnar@ottawahospital.on.ca

Early detection of delirium and depression (presenting with cognitive symptoms) may limit the duration of the disorders by triggering earlier treatment. Earlier detection of cognitive impairment of any etiology may benefit patients and families in several ways. It advances the time when they seek assistance, perhaps prior to the development of avoidable stress and related medical disorders for the caregivers. It stimulates planning for the future such as establishing power of attorney, wills and advanced directives while patients are still capable of contributing to decisions. Other benefits include addressing safety issues such as the person's ability to live independently, their wandering risk, fire risk, medication error risk and fitness to drive in order to prevent avoidable morbidity and mortality.[2] There is also some evidence that early initiation of pharmacotherapy for dementia may prolong independence and reduce costs by delaying nursing home placement.[3] For these reasons, and even in the absence of a cure, there is a growing consensus that persons over age 75 should be routinely screened for cognitive impairment.[4–7]

Previous research has demonstrated that physicians miss cognitive impairment in over 50% of cases: the gestalt method (based on general impression or the physician's intuition) is inadequate.[8–13] Plausibly, early or mild cases may be missed while more advanced cases are preferentially detected. Formal screening is not routinely employed. Only 39% of Australian general practitioners (GPs) and 26% of Canadian GPs regularly screen for dementia.[4,14]

In order to increase screening rates we must understand the barriers to screening. The time to administer screening tests is often cited as a barrier, especially in high volume specialties with marked time pressures such as emergency medicine, general practice or family medicine.[4,15,16] Here, the Mini-Mental State Examination (MMSE)[17] has been described as impractical as it takes 10 minutes to administer,[1,18,19] and the need for shorter screening instruments has been highlighted.[4]

Another barrier to routine screening is the complexity of administration and scoring. This touches on Feinstein's concept of 'clinical sensibility'[20, 21]—whether a tool is clinically reasonable, quick and simple to apply, score and interpret, and whether it suggests a course of action (i.e. categorizes patients rather than merely providing probabilities). Clinical sensibility demands that tools be designed with the realities of busy front-line clinical care in mind. While not explicitly employing the term, Brodaty et al. clearly incorporated the notion of clinical sensibility in their excellent review of dementia screening instruments.[22]

Screening tools often focus on specific diagnoses such as delirium[23], depression[24] or dementia.[16,22] However, in first-contact care settings there may be value in also using very simple screens for the more general symptom of cognitive impairment rather than (or before) applying screens for the more specific causes of impairment, such as delirium, depression, or dementia. Because a screening tool for one of these specific disorders may miss the presence of the other two it may prove optimal to first screen for the common denominator amongst these disorders—cognitive impairment. Screening for cognitive impairment and screening for specific causes are not necessarily mutually exclusive, but may represent complementary approaches that can enhance clinical care.

A number of tools to identify impaired cognition exist. These have been reviewed by Lorentz,[16] Burns,[18] McDowell,[19] and Brodaty et al.[22] The shortest tests on McDowell's list are the Clock Drawing,[25–27] Kahn's Mental Status Questionnaire (MSQ),[28] Pfeiffer's Short Portable Mental Status Questionnaire (SPMSQ)[29] and the Clifton Assessment Procedures for the Elderly (CAPE).[30] Burns indicated that three tests were short enough to be potentially useful in the primary care setting: the Folstein MMSE[17], the Abbreviated Mental Test Score (ABMTS),[31] and the Clock Drawing. With the exception of the Clock Drawing, these tests take several minutes to apply and hence may be too long for many front-line clinicians to easily recall and routinely apply in a busy clinical practice.

The Clock Drawing test merits further discussion as research on this test is often misunderstood. This test is easy to apply but is difficult to score—so the simplicity of the clock drawing is somewhat illusory. Our discussions with specialists and primary care physicians indicate that it is very uncommon for physicians to employ any of the more than 10 scoring systems[19] for the Clock Drawing upon which published psychometric data are based. Indeed, many physicians seem unaware of these scoring systems, and most employ general impression (i.e. gestalt) to rate the clock as either correct or incorrect. This should readily detect

moderate to severe cognitive impairment but may miss subtle drawing errors suggestive of milder impairments. The sensitivity and specificity of such a subjective approach are likely highly physician dependent and hence variable.

Lorentz et al.[16] and Brodaty et al.[22] after performing independent in-depth assessments of psychometric properties and clinical sensibility features, recommended the same three instruments for general practitioners to screen for dementia: the General Practitioner Assessment of Cognition (GPCOG),[32] the Mini-Cog,[33] and the Memory Impairment Screen (MIS).[34] Despite the fact that these tests take 3 to 4½ minutes to apply, they are not routinely used on a widespread basis most likely because they may still be too long and too difficult to score for many clinicians to consider applying them on a routine daily basis.

In the search for even shorter screening tools, Siu has shown that no single cognitive question is adequate as a screen for cognitive impairment.[35] He concluded that a brief combination of screening questions could best determine the need for additional mental status examination.

The goal of our study was to determine if subsets of two, three or four cognitive questions could demonstrate adequate sensitivity and specificity to justify further validation research. Some researchers have pursued this line of inquiry using multivariate analysis, producing predictive formulae. We feel it is unrealistic to expect physicians to routinely incorporate complex equations into clinical practice. This contravenes the principle of clinical sensibility, and therefore multivariable equations are rarely, if ever, used in routine front-line clinical care. We decided a priori that while we would use multivariate analyses to select cognitive questions, the final product of this research would not require any calculation.

## Methods

### Sample and diagnostic classification

The data analyzed were collected as part of the Canadian Study of Health and Aging (CSHA, www.csha.ca ), a national, multicentre epidemiological study of dementia.[36] The first wave (CSHA-1), performed in 1991–1992, drew randomly selected samples of people 65 years of age and older throughout Canada. Of the 10,263 people surveyed, 9008 were living in the community and 1255 in long-term care institutions. The Modified Mini-Mental State Examination (3MS)[37] was used in the community sample as a cognitive screen. The 1600 participants who screened positive for cognitive impairment as defined by a 3MS score ≤77, plus a random sample of 494 who screened negative (3MS score >77) received a full clinical evaluation that included extensive medical, neurological and neuropsychological examinations. Participants who were still alive were re-contacted in 1996 for the follow-up study (CSHA-2), following the same protocol as in CSHA-1.

Based on 12 neuropsychological tests, background information, and established normative information, the study neuropsychologists made a cognitive assessment. The assessments by the physicians included a mental status evaluation (including the 3MS results), as well as general physical and neurological examinations. The physician made an independent preliminary diagnosis. A nurse interviewed a knowledgeable informant who supplied historical information via the CAMDEX.[38] The study diagnosis was established at a clinical consensus meeting involving physician, neuropsychologist and nurse. Subjects were classified as having no cognitive impairment, cognitive impairment but not dementia, or dementia. Dementia was rated as mild, moderate or severe and an etiological diagnosis was made.

As physicians do not require screens to detect persons with severe cognitive impairment, and as the inclusion of severely cognitively impaired participants in the analysis would artificially inflate sensitivity and specificity, the data from persons with severe cognitive impairment were removed. The institutionalized sample was not included as it may represent a spectrum of disease not encountered in the community. Other exclusion criteria included primary language other than English, and deficits in hearing or vision severe enough to affect cognitive testing.

### Analysis

Diagnoses were collapsed into cognitively normal vs cognitively impaired. As we were developing a general cognitive screen, the latter included persons with any etiology of cognitive impairment including ethanol induced impairment, cerebrovascular disease, pre-clinical dementias (some of whom would now be diagnosed as having mild

cognitive impairment), etc. Individual cognitive questions were drawn from the 3MS, which provides a larger selection of items than the 30-point MMSE, although virtually all of the MMSE items are embedded in the 3MS. To maximize practicality, we discarded questions that required the use of paper and pen, cue cards or props, and those requiring more than 30 seconds to answer. We simplified scores for each item into a dichotomous score of right or wrong to promote ease of recall and scoring.

Using chi-square ($\chi^2$) univariate analyses of the dichotomously scored questions, we tested the association between each question and cognitive impairment. Questions showing statistically significant associations were then entered into a forward logistic regression algorithm, using cognitive status (i.e. impaired vs. normal) as the dependent variable.

A characteristic of standard logistic regression algorithms is that they do not maximize either sensitivity or specificity, but instead attempt to maximize the overall correct classification—a combination of sensitivity and specificity. This could restrict the ability of regression analyses to yield maximally sensitive or maximally specific screens. To overcome this limitation, we employed a 'logistic regression serial weighting algorithm' by successively weighted cases (cognitively impaired) from 1 to 10 while keeping cognitively normal controls fixed at a weight of 1 in order to generate increasingly sensitive but less specific logistic regression equations. We then serially weighted the controls from 1 to 10 while keeping cases fixed at a weight of 1 in order to generate increasingly specific but less sensitive equations.

Simple scales of equally-weighted questions were created using the first 2, 3 or 4 questions from the logistic regression equations. In multivariate analyses each subsequent variable added to an equation accounts for less variance, so selecting the earliest variables is appropriate.

Sensitivities and specificities were calculated for all possible cutoffs of the sets of 2, 3 and 4 questions generated and areas under the receiver operating characteristic (ROC) curve (AUC) were calculated. The nonparametric trapezoidal ROC analysis of Hanley and McNeil was employed for calculation of the AUC as the brief scale length, plus previous empiric evidence indicate that scores would not assume normal distributions.[39, 40]

The initial screening tools were derived from the CSHA-1. Promising combinations of cognitive questions with sensitivity and specificity similar to those of the full 3MS were selected for validation in data from the second CSHA wave performed in 1996–1997 of surviving CSHA-1 participants. This involved testing the sensitivity and specificity of the brief screening scales developed from the CSHA-1 data in the equivalent data collected five years later, at CSHA-2.

## Results

The selection criteria generated 1560 CSHA-1 community participants who had undergone full medical and neuropsychological evaluation with cognitive status determined at a consensus conference (958 were mildly to moderately cognitively impaired and 602 were not cognitively impaired). The mean age was 79.5 years old (SD 6.7) with 60.8% of participants being female. Twenty-four percent had completed high school and 10% had additional post-secondary education. The mean 3MS score was 76.8, which translated into a calculated MMSE score between 23 and 24.

The complete 3MS had a sensitivity of 84% and specificity of 62% in distinguishing between these two groups, using a cutoff score between 77 and 78 on the 100-point scale.

Based on practical clinical sensibility criteria, 21 candidate questions were selected for analysis of association with cognitive impairment. Consistent with Siu's findings,[35] Chi-square analyses demonstrated that no single question had both sensitivity and specificity approaching those of the complete 3MS. The two most promising single questions were; 'Count from 5 to 1 backwards' and 'Spell WORLD backwards' (Table 1). Eighteen questions demonstrated a statistically significant association with cognitive status. Since the three non-significant questions were close in terms of sensitivities and specificities to the other questions, we ran the logistic regression analyses with and without the three non-significant variables. Including the non-significant variables did not change the results of multivariate analyses.

The serially weighted logistic regression analyses yielded 8 scales ranging from 2 to 4 questions in length (Table 2). The two tests whose psychometric properties most closely resembled those of the complete 3MS in the CSHA-1 derivation data were set 3–1 (which we named the Ottawa **3D** test: **D**ay,

**Table 1.** The sensitivity and specificity of individual screening questions in distinguishing between cognitive impairment and no cognitive impairment. Results of univariate ($\chi^2$) analysis, Canadian Study of Health and Aging.

| 3MS Question | Sensitivity (%) | Specificity (%) |
|---|---|---|
| 1. In which year were you born? | 8 | 98 |
| 2. What is your birth date? | 6 | 99 |
| 3. In which month were you born? | 3 | 99 |
| 4. Repeat these 3 words (registration). | 17 | 93 |
| 5. **Count from 5 to 1 backwards**. | **61** | **69** |
| 6. **Spell WORLD backwards** | **61** | **70** |
| 7. Can you remember the three words? | 95 | 18 |
| 8. What is today's date? | 53 | 78 |
| 9. Which month are we in? | 15 | 99 |
| 10. What year is it? | 23 | 97 |
| 11. What season are we in? | 12 | 97 |
| 12. What day of the week is it? | 24 | 98 |
| 13. What province (state) are we in? | 6 | 99 |
| 14. What country are we in? | 13 | 99 |
| 15. What city are we in? | 6 | 98 |
| 16. What building are we in?* | 1 | 99 |
| 17. What is this (point to forehead)?* | 4 | 97 |
| 18. What is this (point to chin)? | 3 | 99 |
| 19. What is this (point to shoulder)?* | 4 | 98 |
| 20. What is this (point to elbow)? | 4 | 98 |
| 21. What is this (point to knuckle)? | 23 | 87 |

*$\chi 2$ (1 d.f.) $p > 0.05$.

**D**ate, **D**LROW) and set 4–1 (the Ottawa **3DY** test: **D**ay, **D**ate, **D**LROW, **Y**ear). Taking one or more errors to indicate cognitive impairment, the Ottawa 3D test demonstrated a sensitivity of 80% and a specificity of 56%. With a cutoff of one or more errors the Ottawa 3DY test demonstrated a sensitivity of 82% and a specificity of 55%.

The sensitivity and specificity of the Ottawa 3D and 3DY in the CSHA-2 validation database are shown in Table 3. The final results were: Ottawa 3D sensitivity = 76% and specificity = 62%; Ottawa 3DY sensitivity = 80% and specificity = 61%).

## Discussion
In the clinical care of persons at risk for dementia and delirium, it is likely misguided to believe that one cognitive screening tool will be acceptable to all clinicians on all occasions in all settings. Rather, we should respect physician heterogeneity and the rapidly changing and challenging clinical scenarios they face. Clinicians should have a range of useful tools to draw on depending on the demands of the circumstance—a *cognitive screening toolbox*. Some physicians will want screening tools of varying length and complexity to detect general cognitive impairment while others will want a range of tools to screen for specific diagnoses such as dementia, delirium or depression. Some may resist screening altogether, feeling their general impression is adequate. The latter group might become more receptive to formal screening if a well-stocked toolbox of cognitive screening tests were readily available (perhaps with downloadable forms on national medical association or dementia websites).

The shortest of the previously published tests requires several minutes to apply and would likely

**Table 2.** Candidate sets of screening questions, derived from logistic regression analyses, to distinguish between cognitive impairment and no cognitive impairment. Canadian Study of Health and Aging—1. (all questions weighted equally—1 point).

| Set | Cognitive questions | AUC* | Cut-off (# of errors) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|---|---|
| 2–1 | What is the day of the week? | 0.709 | 2 | 16 | 99 |
| | Spell WORLD backwards | | 1 | 69 | 69 |
| 2–2 | What is the day of the week? | 0.646 | 2 | 13 | 100 |
| | What is the year? | | 1 | 34 | 95 |
| 2–3 | Spell WORLD backwards | 0.716 | 2 | 34 | 90 |
| | What is the date? | | 1 | 79 | 57 |
| **3–1** | (**Ottawa 3D**—**D**ay/**D**LROW/**D**ate) | 0.742 | 3 | 13 | 99 |
| | What is the **D**ay of the week? | | 2 | 44 | 89 |
| | Spell WORL**D** backwards | | **1** | **80** | **56** |
| | What is the **D**ate? | | | | |
| 3–2 | What is the day of the week? | 0.730 | 3 | 9 | 100 |
| | Spell WORLD backwards | | 2 | 26 | 97 |
| | What is the year? | | 1 | 72 | 67 |
| 3–3 | What is the day of the week? | 0.698 | 3 | 12 | 100 |
| | What is the year? | | 2 | 28 | 99 |
| | What is the date? | | 1 | 59 | 74 |
| **4–1** | (**Ottawa 3DY**—**D**ay/**D**LROW/**D**ate/**Y**ear) | 0.752 | 4 | 9 | 100 |
| | What is the **D**ay of the week? | | 3 | 22 | 99 |
| | Spell WORL**D** backwards | | 2 | 48 | 88 |
| | What is the **D**ate? | | **1** | **82** | **55** |
| | What is the **Y**ear? | | | | |
| 4–2 | What is the day of the week? | 0.759 | 4 | 12 | 99 |
| | Spell WORLD backwards | | 3 | 40 | 91 |
| | What is the date? | | 2 | 76 | 65 |
| | What is the difference between laughing and crying? | | 1 | 95 | 24 |
| 3MS | Using the published 77/78 cut-off | 0.800 | 23 | **84** | **62** |

*Area under receiver operating characteristic curve.

be too complex for most busy MDs to readily recall. With this in mind, our intention was not to develop screening tools to compete with existing tests but rather to add to the armamentarium of practical or clinically sensible tests available to clinicians—to augment their *cognitive screening toolbox*—by exploring the possibility of even shorter tests. The Ottawa 3D and Ottawa 3DY tests performed as well as the 3MS and, due to their ease of recall and scoring, minimize barriers to screening and thereby maximize the feasibility of widespread application.

Some have questioned why we attempted to match the sensitivity and specificity of the 3MS rather than aiming for 100% sensitivity and 100% specificity. Cognitive decline is a continuum that is conveniently, but somewhat arbitrarily, divided into those with normal cognition and those with cognitive impairment. Because of individual variations, test scores for these groups commonly

overlap (Fig. 1A), so reducing sensitivity and specificity. Due to commonly overlapping distributions of test scores for persons with normal cognition vs those with cognitive impairment (Fig. 1A) there is always a trade-off between sensitivity and specificity when employing dichotomization (as sensitivity increases specificity decreases—see Fig. 1C and Tables 2 and 3). Although less apparent, this reciprocal relationship between sensitivity and specificity is also reflected in standard ROC curves (Fig. 1D). The result is that no cognitive test can ever simultaneously approach perfect sensitivity and specificity. It is therefore more informative to judge the Ottawa 3D and 3DY tests against existing validated tests such as the 3MS than against an unattainable ideal.

Although they performed well in comparison to the 3MS, the Ottawa 3D and 3DY tests are not ready for immediate clinical use. This study should be viewed as an exploratory data analysis with a preliminary validation, with several limitations that mandate that the findings be more thoroughly
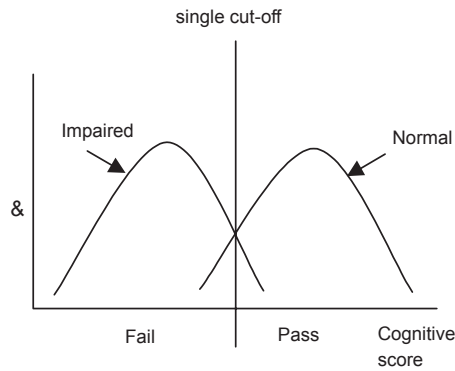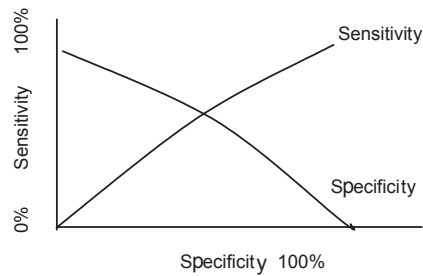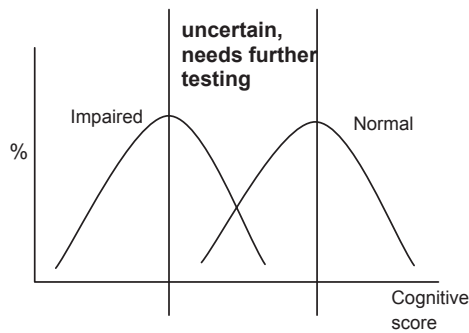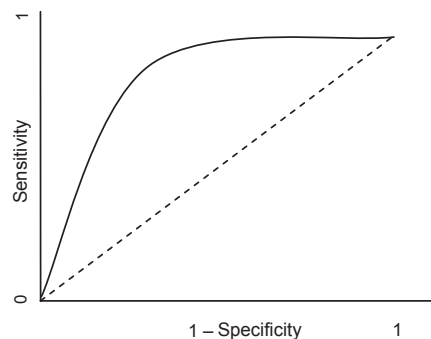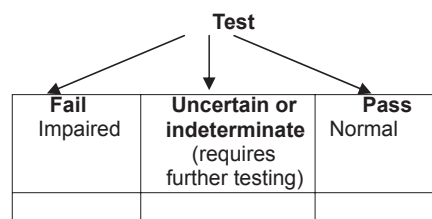


**Figure 1.** Features of cut-off scores.

**Table 3.** Validation of screening sets 3–1 and 4–1 on the second wave of the Canadian Study of Health and Aging data.

| Set | Cognitive questions | Cut-off (# of errors) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|---|
| 3–1 | (**Ottawa 3D**—**D**ay / **D**lrow / **D**ate) | 3 | 9 | 100 |
| | What is the **D**ay of the week? | 2 | 35 | 93 |
| | Spell WORL**D** backwards | **1** | **76** | **62** |
| | What is the **D**ate? | | | |
| 4–1 | (**Ottawa 3DY**—**D**ay / **D**lrow / **D**ate/**Y**ear) | 4 | 5 | 100 |
| | What is the **D**ay of the week? | 3 | 16 | 99 |
| | Spell WORL**D** backwards | 2 | 41 | 93 |
| | What is the **D**ate? | **1** | **80** | **61** |
| | What is the **Y**ear? | | | |

validated before widespread clinical use is recommended. The limitations include the fact that the derivation (CSHA-1) and validation (CSHA-2) databases are not completely independent, and the cognitive questions studied (including the entire 3MS) formed a small part of the data employed to achieve a consensus diagnosis of cognitive status. These factors can artificially inflate both sensitivity and specificity.

Since those who screened positive for cognitive impairment (plus a random subset of those who screened negative) were selected for full clinical evaluation, verification bias (a.k.a. work-up bias or sequential-ordering bias) exists and the risk that the sensitivities and specificities reported are inaccurate must be recognized. Because of their shared content, it is likely that any verification bias will act in a similar fashion on the Ottawa 3D, 3DY and 3MS. If so, then the comparison between them should remain constant in other datasets (i.e. the psychometric properties of the Ottawa 3D and 3DY should remain close to those of the 3MS). More extensive validation research is required to support or refute this possibility. Consequently, we cannot definitively predict the psychometric properties of the tests in actual clinical practice and limit our conclusion to stating that the Ottawa 3D and 3DY screens appear to have psychometric properties similar to the complete 3MS, a test that has been found to be superior to the MMSE, and that these new tests therefore merit further study.

This study has several strengths that are worth highlighting. The Ottawa 3D and 3DY tests are easy to recall, apply and score and hence can be applied anywhere and at any time. The use of a comprehensive multidisciplinary consensus diagnosis based on extensive medical and neuropsychological assessments provided a high quality diagnostic criterion. The removal of institutionalized participants and persons with severe dementia from analysis creates more realistic estimates of the performance of the tests in community-based settings

## Future research

It is unfortunately common for cognitive tests to be promoted for clinical application before proper validation research has been completed. This tendency to prematurely apply new tests to clinical practice results in these new tests failing to live up to their promise, showing reduced sensitivity and specificity when applied to real clinical populations. With this in mind, while we encourage clinicians to try the Ottawa 3D and 3DY screens, we recommend they do so in conjunction with established tests such as the MMSE. The Ottawa 3D and 3DY screens should not be used in isolation before proper validation research is performed on an appropriate spectrum of patients (mild to moderate cognitive impairment vs. cognitively normal) in the intended setting (primary care, emergency departments or community based settings). Furthermore, psychometric properties of cognitive tests will depend on whether the tests are used for screening (i.e. applied to all patients in the clinical setting of interest) or case finding (i.e. applied only to patients with a high risk of dementia such as

those over 75 years old with vascular risk factors and/or a family history of dementia). With these considerations in mind, validation against a high-level cognitive assessment, as employed in the CSHA, in a primary care or community-based sample of participants should be performed. Where such databases already exist, we encourage researchers holding the data to evaluate the psychometric properties of Ottawa 3D and 3DY tests.

The full potential value of the Ottawa 3D and 3DY tests can best be understood in the context of a serial testing approach whereby short tests requiring very little time are applied to large sets of patients and longer tests are applied to increasingly smaller subsets of patients that are selected by the shorter tests (Fig. 2). This approach improves efficiency by increasing case finding while minimizing time investment. Due to selection of increasingly smaller subsets of patients, the sensitivities and specificities of secondary and tertiary screens would need to be reevaluated in the context of the screening algorithms.

Sequential screening approaches can be further refined to maximize efficiency. When employing a single cut-off (i.e. dichotomization), screening
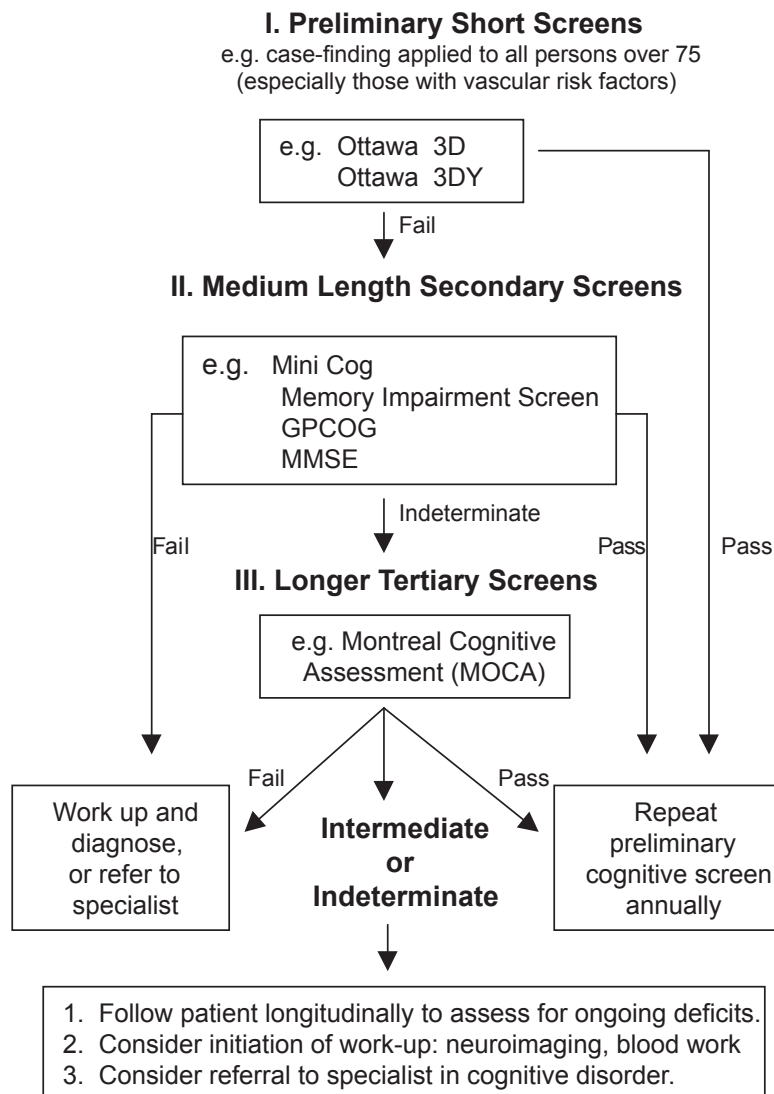
### I. Preliminary Short Screens
e.g. case-finding applied to all persons over 75
(especially those with vascular risk factors)

e.g.  Ottawa  3D
       Ottawa  3DY

↓ Fail

### II. Medium Length Secondary Screens

e.g.   Mini Cog
        Memory Impairment Screen
        GPCOG
        MMSE

Fail          Indeterminate          Pass          Pass

### III. Longer Tertiary Screens

e.g. Montreal Cognitive Assessment (MOCA)

Fail          Pass

**Work up and diagnose, or refer to specialist**

**Intermediate or Indeterminate**

**Repeat preliminary cognitive screen annually**

1. Follow patient longitudinally to assess for ongoing deficits.
2. Consider initiation of work-up: neuroimaging, blood work
3. Consider referral to specialist in cognitive disorder.

**Figure 2. Possible clinical algorithm for Serial Testing and Trichotomization** (N.B. This is not the only possible approach but should be viewed as one of many potential approaches).

and diagnostic tests often demonstrate limited psychometric properties (e.g. poor sensitivity, specificity, positive predictive value, and/or negative predictive value) due to the frequent overlap between the distributions of test results of persons who are cognitively normal and those with impairment (Fig. 1A). Trichotomization, employing two cut-points, can sometimes overcome the limited sensitivities and specificities generated by dichotomization—employing a single cut-point (Fig. 1B). By combining the sequential testing and trichotomization approaches we can develop increasingly efficient screening and case-finding algorithms as depicted in Figure 2. Very simple and rapid screening tools such as the Ottawa 3D and 3DY tests that can be readily applied to large numbers of patients without disrupting clinical practice could serve as triggers for such high efficiency screening algorithms.

The concept of 'serial trichotomization' is not new to the field of dementia as it has been recommended in the assessment of fitness-to-drive in dementia.[41] Dementia researchers should consider reanalyzing neuropsychological tests, biomarkers and neuroimaging using serial trichotomization to determine if this approach will generate more promising sensitivities and specificities.

## Conclusions

The Ottawa 3D and 3DY tests are promising cognitive screens that should be simple enough to employ to promote widespread use. While the results are very promising they must be validated in the target group for which the tests are intended before we will know the true sensitivity, specificity and utility of the tests. Despite these cautionary notes the findings of this study are very promising and, if the derived tests live up to their promise in future validation studies, we would have a major addition to the armamentarium of practicing clinicians.

## Acknowledgements, Funding, Conflicts of Interest, Competing Interests

## References

[1]  Ferri, C.P., Prince, M., Brayne, C. et al. 2005. Global prevalence of dementia: a Delphi consensus study. *Lancet*, 366:2112–17.

[2]  Boise, L., Camicioli, R., Morgan, D.L. et al. 1999. Diagnosing dementia: perspectives of primary care physicians. *Gerontologist*, 39:457–64.

[3]  Doraiswamy, P.M., Krishnan, K.R.R., Anand, R. et al. 2002. Long-term effects of rivastigmine in moderately severe Alzheimer's disease: does early initiation of therapy offer sustained benefits? *Prog. Neuropsychopharmacol.*, 26:705–12.

[4]  Brodaty, H., Clarke, J., Ganguli, M. et al. 1998. Screening for cognitive impairment in general practice: toward a consensus. *Alzheimer Dis. Assoc. Disord.*, 12:1–13.

[5]  Brodaty, H., Howarth, G.C., Mant, A. et al. 1994. General practice and dementia. A national survey of Australian GPs. *Med. J. Aust*, 160:10–4.

[6]  Doraiswamy, P.M., Steffens, D.C., Pitchumoni, S. et al. 1998. Early recognition of Alzheimer's disease: what is consensual? What is controversial? What is practical? *J. Clin. Psychiatry*, 59:6–18.

[7]  Small, G.W., Rabins, P.V., Barry, P.P. et al. 1997. Diagnosis and treatment of Alzheimer disease and related disorders. Consensus statement of the American Association for Geriatric Psychiatry, the Alzheimer's Association, and the American Geriatrics Society. *JAMA*, 278:1363–71.

[8]  Valcour, V.G., Masaki, K.H., Curb, J.D. et al. 2000. The detection of dementia in the primary care setting. *Arch. Intern. Med.*, 160:2964–68.

[9]  Illife, S., Booroff, A., Gallivan, S. et al. 1990. A screening for cognitive impairment in the elderly using the Mini-Mental State Examination. *Br. J. Gen. Pract.*, 40:277–9.

[10]  O'Connor, D.W., Pollitt, P.A., Hyde, J.B. et al. 1988. Do general practitioners miss dementia in elderly patients? *BMJ*, 297:1107–10.

[11]  Bowers, J., Jorm, A.F., Henderson, S. and Harris, P. 1990. General practitioners' detection of depression and dementia in elderly patients. *Med. J. Aust.*, 153:192–6.

[12]  McMartney, J.R. and Palmateer, L.M. 1985. Assessment of cognitive deficit in geriatric patients. *JAGS*, 33:483–8.

[13]  Somerfield, M.R., Weisman, C.S., Ury, W. et al. 1991. Physician practices in the diagnosis of dementing disorders. *JAGS*, 39:172–5.

[14]  Bush, C., Kozak, J. and Elmslie, T. 1997. Screening for cognitive impairment in the elderly. *Can. Fam. Physician*, 43:1763–68.

[15]  Deveugele, M., Derese, A., Van den Brink-Muinen, A. et al. 2002. Consultation length in general practice: cross-sectional study. *BMJ*, 325:472–4.

[16]  Lorentz, W.J., Scanlan, J.M. and Borson, S. 2002. Brief screening tests for dementia. *Can. J. Psychiatry*, 47:723–33.

[17]  Folstein, M.F., Folstein, S.E. and Mcugh, P.R. 1975. 'Mini-mental state.' A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.*, 12:189–98.

[18]  Burns, A., Lawlor, B. and Craig, S. Assessment Scales in Old Age Psychiatry. Andover, Thomson Publishing Services, 2004.

[19]  McDowell, I. Measuring Health (third edition). New York, Oxford University Press, 2006.

[20]  Feinstein, A.R., Clinimetrics. New Haven, C.T. Yale University Press, 1987.

[21] Stiell, I.G., Greenberg, G.H., McKnight, R.D. et al. 1993. Decision rules for the use of radiography in acute ankle injuries—refinement and prospective validation. *JAMA*, 269:1127–32.

[22] Brodaty, H., Low, L.F., Gibson, L. and Burns, K. 2006. What is the best dementia screening instrument for General Practitioners to use? *Am. J. Geriatr. Psychiatry*, 14:391–400.

[23] Inouye, S.K., van Dyck, C.H., Alessi, C.A. et al. 1990. Clarifying confusion: the Confusion Assessment method: a new method for detection of delirium. *Ann. Intern. Med.*, 113:941–8.

[24] Yesavage, J.A., Brink, T.L., Rose, T.L. et al. 1983. Development and validation of a Geriatric Depression Screening scale: a preliminary report. *J. Psychiatr. Res.*, 17:37–49.

[25] Shulman, K., Shedletsky, R. and Silver, I.L. 1986. The challenge of time: clock drawing and cognitive function in the elderly. *Int. J. Geriatric. Psychiatry*, 1:135–40.

[26] Sunderland, T., Hill, J.L., Mellow, A.M. et al. 1989. Clock drawing in Alzheimer's disease. *JAGS*, 37:725–9.

[27] Wolf-Klein, G.P., Silverstone, F.A., Levy, A.P. et al. 1989. Screening for Alzheimer's disease by clock drawing. *JAGS*, 37:730–4.

[28] Kahn, R.L., Goldfarb, A.I., Pollack, M. et al. 1960. Brief objective measures for the determination of mental status in the aged. *Am. J. Psychiatry*, 117:326–8.

[29] Pfeiffer, E. 1975. A Short Portable Mental Status Questionnaire for the assessment of organic brain deficit in elderly patients. *JAGS*, 23:433–41.

[30] Pattie, A.H. and Gilleard, C.J. 1975. A brief psychogeriatric assessment schedule: validation against psychiatric diagnosis and discharge from hospital. *Br. J. Psychiatry*, 127:489–93.

[31] Hodkinson, M. 1972. Evaluation of a mental test score for assessment of mental impairment in the elderly. *Age Aging*, 1:233–8.

[32] Brodaty, H., Pond, D., Kemp, N.M. et al. 2002. The GPCOG: a new screening test for dementia designed for general practice. *JAGS*, 50:530–4.

[33] Borson, S., Scanlan, J.M., Chen, P. et al. 2003. The Mini-Cog as a screen for dementia: validation in a population-based sample. *JAGS*, 51:1451–54.

[34] Buschke, H., Kuslansky, G., Katz, M. et al. 1999. Screening for dementia with the Memory Impairment Screen. *Neurology*, 52:231–8.

[35] Siu, A.L. 1991. Screening for dementia and investigating its causes. *Ann. Intern. Med.*, 115:122–32.

[36] Canadian Study of Health and Aging working group, . 1994. Canadian Study of Health and Aging study methods and prevalence of dementia. *Can. Med. Assoc. J.*, 150:899–912.

[37] Teng, E.L. and Chui, H.C. 1987. The Modified Mini-Mental State (3MS) Examination. *J. Clin. Psychiatry*, 48:314–8.

[38] Roth, M., Tym, E., Mountjoy, C.Q. et al. 1986. CAMDEX: a standardised instrument for the diagnosis of mental disorder in elderly with special reference to the early detection of dementia. *Br. J. Psychiatry*, 149:698–709.

[39] McDowell, I., Kristjansson, B., Hill, G.B. et al. 1997. Community screening for dementia: the Mini-mental State Exam (MMSE) and the Modified Mini-Mental State Exam (3MS) compared. *J. Clin. Epidemiol.*, 45:581–6.

[40] Ritchie, K. and Fuhrer, R. 1992. A comparative study of the performance of screening tests for senile dementia using Receiver Operating Characteristics Analysis. *J. Clin. Epidemiol.*, 45:627–37.

[41] Molnar, F.J., Patel, A., Marshall, S.C. et al. 2006. Clinical utility of office-based cognitive predictors of fitness to drive in persons with dementia: a systematic review. *JAGS*, 54:1809–24.