

## The clock drawing test is a poor screening tool for postoperative delirium and cognitive dysfunction after aortic repair

## Le test du dessin de l'horloge est un outil de dépistage médiocre pour le delirium postopératoire et le dysfonctionnement cognitif après une chirurgie de l'aorte

Gregory L. Bryson, MD · Anna Wyand, MD · Denise Wozny, BA ·  
Laura Rees, PhD · Monica Taljaard, PhD · Howard Nathan, MD

Received: 25 September 2010 / Accepted: 13 December 2010 / Published online: 31 December 2010  
© Canadian Anesthesiologists' Society 2010

### Abstract

**Background** The Clock Drawing Test (CDT) is a screening tool for dementia that tests a variety of cognitive domains. The CDT takes a maximum of two minutes to complete and might be helpful in identifying postoperative cognitive disorders at the bedside. The objective of this study was to evaluate the accuracy of the CDT in a population at high risk for postoperative cognitive disorders

**Methods** In this prospective observational cohort study, patients were recruited who were  $\geq 60$  yr of age and scheduled for elective open repair of the abdominal aorta. Delirium was assessed using the Confusion Assessment Method (CAM) on postoperative days (POD) 2 and 4 and at discharge. Cognitive function was assessed with neuropsychometric tests before surgery and at discharge. Postoperative cognitive dysfunction (POCD) was determined using the Reliable Change Index. Clock Drawing Tests were administered at all time points. Agreement between the CDT and test for delirium or POCD was assessed with Cohen's Kappa statistic.

**Results** Delirium was noted in 30 of 83 patients (36%; 95% confidence interval [CI] 26 to 46%) during their hospital stay, while POCD was noted in 48 of 78 patients (60%; 95% CI 51 to 72%) at discharge. Agreement between the CDT and CAM was poor at three intervals (Kappa 0.06 to 0.29), as was POCD at discharge (Kappa 0.46). Sensitivity of the CDT was  $<0.71$  for both delirium and POCD at all intervals. False positives and negatives were common.

**Conclusion** Agreement between CDT and tests for delirium and POCD was poor; sensitivity was inadequate for a screening test. (ClinicalTrials.gov number, NCT00911677).

### Résumé

**Contexte** Le test du dessin de l'horloge (CDT) est un outil de dépistage de la démence qui évalue plusieurs domaines cognitifs. Le CDT prend au maximum deux minutes à compléter et pourrait être utile pour identifier les troubles cognitifs postopératoires directement au chevet du patient. L'objectif de cette étude était d'évaluer l'exactitude du CDT chez une population présentant un risque élevé de troubles cognitifs postopératoires.

**Méthode** Dans cette étude de cohorte prospective et observationnelle, des patients âgés  $> 60$  ans subissant une réparation non urgente ouverte de l'aorte abdominale ont été recrutés. Le delirium a été évalué aux jours postopératoires deux et quatre ainsi qu'au congé à l'aide de la Méthode d'évaluation de la confusion (CAM). Le fonctionnement cognitif a été évalué à l'aide de tests neuropsychométriques avant la chirurgie et lors du congé. Le dysfonctionnement cognitif postopératoire (DCPO) a été déterminé à l'aide d'un indice de changement fiable. Des tests de dessin de l'horloge ont été administrés à tous les points de mesure dans le temps. La correspondance

---

G. L. Bryson, MD (✉) · A. Wyand, MD · D. Wozny, BA ·  
H. Nathan, MD  
Department of Anesthesiology, The Ottawa Hospital,  
1053 Carling Avenue, Box 249C, Ottawa,  
ON K1Y 4E9, Canada  
e-mail: glbryson@ottawahospital.on.ca

L. Rees, PhD  
Department of Psychology, The Ottawa Hospital, Ottawa,  
ON, Canada

M. Taljaard, PhD  
Clinical Epidemiology Program, Ottawa Hospital Research  
Institute, Ottawa, ON, Canada

entre le CDT et le test du delirium ou du DCPO a été évaluée à l'aide de l'analyse statistique kappa.

**Résultats** On a observé du delirium chez 30 des 83 patients (36%; intervalle de confiance [IC] 95%, 26 à 46%) pendant leur séjour à l'hôpital, et le DCPO a été noté chez 48 de 78 patients (60%; IC 95%, 51 à 72%) au moment du congé. La correspondance entre le CDT et le DCPO à trois intervalles (kappa 0,06 – 0,29) et le DCPO au moment du congé (kappa 0,46) était médiocre. La sensibilité du CDT était inférieure à 0,71 pour le delirium et le DCPO à tous les intervalles. Les faux positifs et les négatifs étaient fréquents.

**Conclusion** La correspondance entre le CDT et les tests réalisés pour déterminer la présence de delirium et de DCPO était médiocre; sa sensibilité était inadaptée pour en faire un test de dépistage. (Numéro de ClinicalTrials.gov, NCT00911677).

Delirium is a transient fluctuating disturbance of consciousness, attention, cognition, and perception<sup>1</sup> that complicates the postoperative course of 5 to 52% of patients undergoing non-cardiac surgery.<sup>2</sup> New onset delirium in hospital is associated with increased length of hospital stay,<sup>3</sup> greater rates of nursing home placement,<sup>4</sup> and mortality rates approaching 30%.<sup>5</sup> Postoperative cognitive dysfunction (POCD) is a subtle disorder of intellectual function identified with detailed neuropsychometric testing.<sup>6</sup> Postoperative cognitive dysfunction can be found in 30 to 40% of surgical inpatients at the time of discharge<sup>7</sup> and is associated with problems in the workplace and long-term mortality.<sup>8</sup>

Despite their prevalence among surgical patients and their association with adverse outcomes, these disorders of cognition are often overlooked by clinicians at the bedside. Delirium is under-recognized by both nurses<sup>9</sup> and physicians,<sup>10</sup> and the detailed neuropsychometric testing batteries used to identify POCD require both specialized training and as long as 60 min to administer.<sup>11</sup> If clinicians are to identify and manage these common cognitive disorders on a busy surgical service, they require a simple, fast, and easily administered bedside diagnostic tool.

The Clock Drawing Test (CDT) is a well-described screening test for dementia that has been demonstrated to be >85% sensitive for cognitive impairment identified on the Mini-Mental® State Examination.<sup>9</sup> While different variants of the CDT exist, a common adaptation involves asking the subject to draw and number a clock face on a blank sheet of paper and to set the hands of the clock at “ten minutes past eleven.” The CDT tests a variety of cognitive domains and takes a maximum of two minutes to

complete. It would seem, therefore, that the CDT might be helpful in identifying postoperative cognitive disorders at the bedside.

The objective of this study was to evaluate the CDT in a population at high risk for postoperative delirium and cognitive dysfunction by comparing it with tests commonly used in clinical research.<sup>12</sup> The hypothesis of this study was that an abnormal CDT would identify patients with either delirium or POCD.

## Methods

This prospective observational cohort study complies with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) initiative.<sup>10</sup> Following approval of The Ottawa Hospital Research Ethics Board (Protocol OHREB 2004800-01H), patients were assessed for study eligibility if they were  $\geq 60$  yr of age and scheduled for elective open repair of the abdominal aorta at The Ottawa Hospital, a 1,000 bed academic tertiary care centre. Screening for the trial began in August 2005, and the final assessment was completed in March 2008. Exclusion criteria included: 1) planned endovascular repair; 2) emergency surgery; 3) previous diagnosis of dementia, Parkinson's disease, or psychiatric illness; 4) an undiagnosed preoperative cognitive disorder indicated by a score of  $\leq 23$  on the Mini-Mental State Examination<sup>11</sup>; 5) active alcohol or substance abuse; and 6) physical inability to complete psychometric testing. Informed written consent was obtained from eligible consenting patients, and baseline measurements were obtained.

The present study evaluating the CDT was a sub-study of a trial whose primary purpose was to explore the relationships among delirium, POCD, and the apolipoprotein  $\epsilon 4$  genotype. The methods of this trial are described in detail elsewhere<sup>13</sup> and are summarized in the subsequent paragraphs.

Patients enrolled in this study received usual perioperative care. Monitoring, anesthetic technique, and postoperative analgesia were at the discretion of the attending anesthesiologist. On postoperative days (POD) 2 and 4 and at discharge, delirium was assessed in hospital using the Confusion Assessment Method (CAM).<sup>12</sup> The CAM poses nine questions to an observer regarding the mental processes of the patient examined. The CAM then identifies four diagnostic features of delirium: 1) acute onset and fluctuating course; 2) inattention; 3) disorganized thinking; and 4) altered level of consciousness. The CAM provides a dichotomous outcome – delirium is either present or absent – based on the presence of features 1 and 2 plus either 3 or 4.

The assessment of POCD was made by trained psychometrists before surgery and at hospital discharge and, whenever possible, by the same evaluator. Detailed neuropsychometric testing was performed using a battery of nine tests compliant with the consensus guidelines for cognitive testing following cardiac surgery<sup>14</sup> and similar in content to the International Study of Post-Operative Cognitive Dysfunction (ISPOCD) test battery.<sup>6</sup> The nine-item neuropsychometric test battery includes: the Rey Auditory Verbal Learning Test; the Trail Making Test (parts A and B); the Grooved Pegboard Test; the Symbol Digit Modalities Test (oral administration); the Wechsler Adult Intelligence Scale (Digit Span); the Wechsler Memory Scale-III (Mental Control Subtest); and letter and category fluency tasks. The Reliable Change Index method<sup>6</sup> was chosen to categorize the presence or absence of POCD from the neuropsychometric test battery. The neuropsychometric test scores from a cohort of 50 non-surgical control patients tested at similar intervals were used to calculate the Reliable Change Index.

The CDT, as described by Roth,<sup>15</sup> was administered at all assessments of either delirium or POCD. The CDT drawings were scored using both the Clock Drawing Interpretation Scale (CDIS)<sup>16</sup> and the Cambridge Mental Disorders of the Elderly Examination (CAMDEX) scoring system.<sup>15</sup> The CDIS is a 20-item score with three items describing the general shape of the drawing, 12 items describing the numbering, and five items describing placement of the hands on the clock face. Each item scores a single point with scores of  $\leq 18$  indicating an abnormal drawing. The CAMDEX scoring system assesses three domains: correctly drawn clock shape, all numbers in the correct position, and “hands of the clock set to the correct time” – each domain scoring one point if completed properly. Nishiwaki<sup>17</sup> suggested adding a fourth domain that would be given a score of one point if absent, i.e., a description of a drawing as “very disorganized, bizarre, or otherwise an abnormal representation of a clock”. Using this variation, the CDT was scored from 0 to 4, with higher scores indicating better performance and scores  $\leq 2$  indicating an abnormal drawing.<sup>17</sup> Drawings were scored by a single investigator (A.W.) who was unaware of the results of the other cognitive assessments.

A convenience sample of patients enrolled in the main study served as the study subjects for this assessment of the CDT. Patient characteristics are described using proportions and means with standard deviations. Estimates of the rates of delirium and POCD are described as proportions with 95% confidence intervals (CI). Assessment of agreement between the CDIS and CAMDEX scoring systems was made using an unweighted Cohen’s Kappa with 95% CI. Given the well-known shortcomings of Kappa and the wide disagreement about the usefulness of Kappa statistics

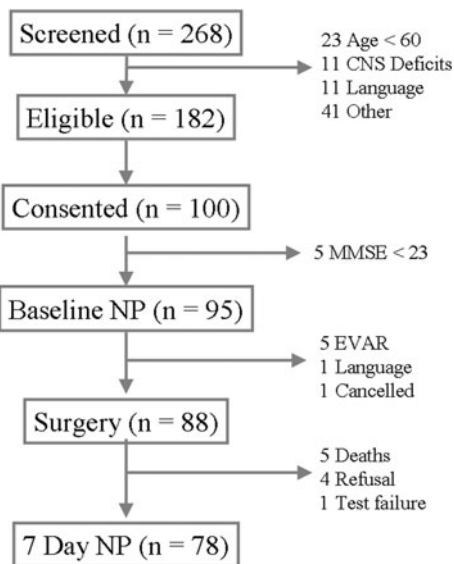
to assess agreement,<sup>18</sup> we additionally calculated the proportions of specific positive and negative agreement, as recommended in the literature, together with 95% CI.<sup>19</sup> The positive agreement index estimates the conditional probability of a positive diagnosis on one scoring system given a positive diagnosis on the other; likewise, the negative agreement index estimates the conditional probability of a negative diagnosis on one scoring system given a negative diagnosis on the other. These indices are analogous to sensitivity and specificity in the presence of a gold standard classification. Assessment of agreement between the CDT and the clinical assessments of both delirium and POCD (considered as gold standard classifications) was made by calculating Kappa with 95% CI as well as sensitivity and specificity with corresponding 95% confidence limits. Exact binomial methods were used in the case of small cell frequencies.

## Results

We screened 268 patients scheduled for open repair of the abdominal aorta. Eighty-eight of the 100 patients who consented to neuropsychometric testing underwent open aortic repair. Patient flow through the trial is summarized in Figure 1. Demographic characteristics of the patients undergoing surgery ( $n = 88$ ) and those completing neuropsychometric testing at discharge ( $n = 78$ ) are shown in Table 1. Fifteen cases of delirium were identified on POD 2; 20 cases (13 new and seven preexisting) were noted on POD 4, and three patients were delirious at the time of discharge testing (one new, two preexisting). In total, delirium was noted in 30 of 83 patients (36%; 95% CI 26 to 46%) completing the CAM assessment at any time during their hospital stay. The diagnosis of POCD was noted in 48 of 78 patients (60%; 95% CI 51 to 72%) completing neuropsychometric testing prior to discharge.

Figure 2 indicates group CDT scores on both the CDIS and CAMDEX scoring systems over the course of the study. In Table 2, a representative series of CDTs from a single patient is presented with related CDIS, CAMDEX, delirium, and POCD assessments. Transition between a normal CDT at baseline to grossly abnormal results on POD 2 and POD 4 is readily apparent. Interestingly this patient was not identified as being delirious on POD 2 despite the highly abnormal CDT. Similarly, the subtle abnormality in the CDT noted at discharge underestimated a significant degree of POCD demonstrated by neuropsychometric testing.

Comparisons of the CDIS and CAMDEX clock scoring methods are summarized in Table 3. Agreement between CDIS and CAMDEX was fair (Kappa 0.39 to 0.65). Positive agreement ranged from 0.76 on POD 2 to 0.44 at



**Fig. 1** Patient flow. CNS = central nervous system; MMSE = Mini-Mental State Examination; NP = neuropsychometric; EVAR = endovascular aortic repair

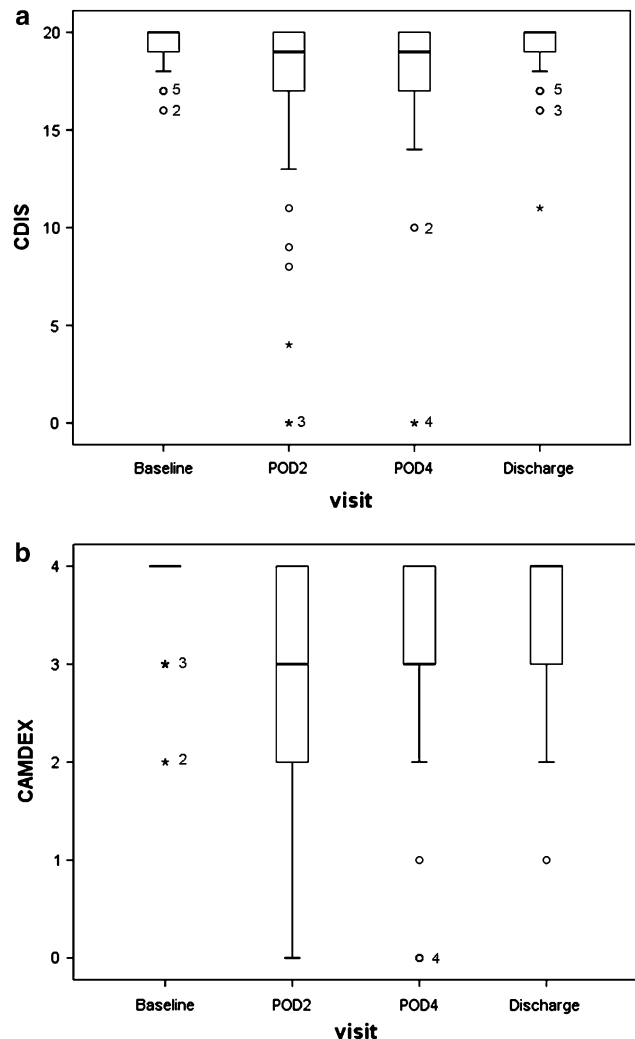
**Table 1** Characteristics of patients at baseline and at hospital discharge

Preoperative Characteristic	Those Undergoing Surgery (n = 88)	Those Completing NP Testing at Discharge (n = 78)
Age at surgery, mean (SD)	71 (6)	71 (6)
Male, n (%)	64 (73%)	61 (78%)
Duke activity status index, mean (SD)	30 (15)	31 (14)
RCRI Class, n (%)		
2	43 (49%)	36 (46%)
3	33 (38%)	31 (40%)
4	12 (14%)	11 (14%)
ASA status, n (%)		
II	7 (8%)	7 (9%)
III	65 (74%)	57 (73%)
IV	16 (18%)	14 (18%)
Mini-Mental State Examination, mean (SD)	29 (1)	29 (1)

SD = standard deviation; NP = neuropsychometric; RCRI = revised cardiac risk index; ASA = American Society of Anesthesiologists

discharge, while negative agreement remained constant at approximately 0.9. The CDIS identified a total of 28 abnormal CDT scores that were not identified by the CAMDEX. Being the more sensitive scoring system, the CDIS was chosen as the scoring tool for comparison with the delirium and POCD assessments.





Comparison of CDIS scores, CDT scores, and delirium, as assessed by the CAM, are shown in Table 4. Agreement



**Fig. 2 a, b** Clock Drawing Test scores over time. POD = postoperative day. Upper and lower margins of box indicate 75th and 25th percentiles, while line indicates median. Whiskers indicate the two standard deviations from mean. O represents values 1.5 to 3.0 times the interquartile range, while \* indicates values > 3.0 times the interquartile range. Numbers indicate total number of values represented at each O or \*

between CDT and CAM assessments was poor, with Kappa consistently < 0.3 noted at any observation interval. If we assume that the CAM is diagnostic of delirium, then the sensitivity of the CDT in identifying delirium ranges from 0.33 at discharge to 0.59 on POD 4, whereas specificity ranges from 0.65 on POD 2 to 0.83 at discharge. Abnormal CDTs were reported with significantly greater frequency than delirium, leading to a number of “false positives.” It should be noted that several patients (three on each of POD 2 and POD 4) were noted to be delirious on CAM assessments but were unable to attempt a CDT. As neither CDIS nor CAMDEX scoring systems assigned a score to clocks “not drawn”, these CDT assessments were excluded. If we assume that absent CDTs were “positive” in the

**Table 2** Representative Clock Drawing Tests and related scores in one patient

	Baseline	POD 2	POD 4	Discharge
				
CDIS	20	9	10	17
CAMDEX	4	0	0	2
CAM		Negative	Positive	Negative
RCI				-10.24

All clocks reduced to 50% original size, except discharge, which was reduced to 25% original size

POD = postoperative day; CDIS = Clock Drawing Interpretation Scale (score 0 to 20, with scores ≤ 18 indicating an abnormal drawing); CAMDEX = Cambridge Mental Disorders of the Elderly Examination scoring system (score 0 to 4, with scores ≤ 2 indicating an abnormal drawing); CAM = Confusion Assessment Method (results expressed as delirium testing positive or negative); RCI = Reliable Change Index (results expressed as Z score (a score < -1.96 indicating postoperative cognitive dysfunction))

**Table 3** Comparison of CDIS and CAMDEX scores

	POD 2 (n = 77)		POD 4 (n = 2)		Discharge (n = 8)	
	CAMDEX-	CAMDEX+	CAMDEX-	CAMDEX+	CAMDEX-	CAMDEX+
CDIS-	46	1	48	0	64	0
CDIS+	11	19	12	12	10	4
Kappa (95% CI)	0.65 (0.47 to 0.82)		0.57 (0.37 to 0.77)		0.40 (0.12 to 0.67)	
PA (95% CI)	0.76 (0.63 to 0.89)		0.67 (0.49 to 0.84)		0.44 (0.16 to 0.73)	
NA (95% CI)	0.88 (0.82 to 0.95)		0.89 (0.83 to 0.95)		0.93 (0.88 to 0.97)	

POD = postoperative day; CAMDEX = Cambridge Mental Disorders of the Elderly Examination scoring system (+ indicates positive result, - indicates negative result); CDIS = Clock Drawing Interpretation Scale score (+ indicates positive result, - indicates negative result); CI = confidence interval; PA = index of positive agreement; NA = index of negative agreement

**Table 4** CDIS Scoring of Clock Drawing and Delirium

	POD 2 (n = 77)		POD 4 (n = 72)		Discharge (n = 78)	
	CAM-	CAM+	CAM-	CAM+	CAM-	CAM+
CDIS-	42	5	41	7	62	2
CDIS+	23	7	14	10	13	1
Kappa (95% CI)	0.14 (-0.05 to 0.34)		0.29 (0.06 to 0.52)		0.06 (-0.14 to 0.26)	
Sensitivity (95% CI)	0.58 (0.28 to 0.85)		0.59 (0.33 to 0.82)		0.33 (0.01 to 0.91)	
Specificity (95% CI)	0.65 (0.52 to 0.76)		0.75 (0.61 to 0.85)		0.83 (0.72 to 0.90)	

POD = postoperative day; CI = confidence interval; CAM = Confusion Assessment Method (+ indicates positive result, - indicates negative result); CDIS = Clock Drawing Interpretation Scale score (+ indicates positive result, - indicates negative result)

presence of delirium, the sensitivity of the CDT at POD 2 increases slightly to 0.66 (95% CI 0.38 to 0.85) with little change in specificity (0.64; 95% CI 0.52 to 0.76). Similarly, sensitivity and specificity of the CDT on POD 4 (0.56; 95% CI 0.41 to 0.83 and 0.75; 95% CI 0.61 to 0.85, respectively) remain largely unchanged when absent CDTs are considered true positives. Assessment of agreement

with the less sensitive CAMDEX score was similarly poor (data not reported).

Table 5 demonstrates the assessment of agreement between CDIS scores of CDT and POCD assessed by neuropsychometric testing at hospital discharge. Agreement as assessed by Cohen’s Kappa was marginally better (0.46), as was sensitivity (0.71) and specificity (0.75). The

**Table 5** CDIS Scoring of Clock Drawing and POCD

	POCD at discharge ( <i>n</i> = 77)	
	POCD–	POCD+
CDIS–	27	12
CDIS+	9	30
Kappa (95% CI)	0.46 (0.27 to 0.66)	
Sensitivity (95% CI)	0.71 (0.55 to 0.84)	
Specificity (95% CI)	0.75 (0.58 to 0.88)	

POCD = postoperative cognitive dysfunction (+ indicates positive result, – indicates negative result); CDIS = Clock Drawing Interpretation Scale score (+ indicates positive result, – indicates negative result); CI = confidence interval

numbers of “false positives” and “false negatives” generated by CDTs were approximately similar when identifying POCD at discharge.

## Discussion

The results of this study indicate that the CDT cannot be recommended for bedside screening of delirium or POCD. Agreement between scoring systems was acceptable, with the CDIS providing a more refined tool that identified cases of delirium more consistently. Agreement between the CDT and clinical assessments of delirium and POCD was relatively poor. As the CDT proved insensitive at detecting delirium or POCD, it would fail to identify a substantial number of “at risk” patients for more rigorous cognitive assessment.

When two tests or observers evaluate the same condition, one can compare test performance between tests or observers using a test of agreement, the most common of which is Cohen’s Kappa. Interpretation of the value of the Kappa coefficient as “good”, “fair”, or “poor” based on categorized ranges is controversial and is actively discouraged by some authors.<sup>A</sup> Nevertheless, a Kappa > 0.6 would suggest to the investigators of this trial that further evaluation of the CDT would be warranted. In evaluating the performance of the CDIS and CAMDEX scoring systems, measures of positive and negative agreement were added to facilitate assessment of agreement. Positive agreement between two scoring systems estimates the probability of a “positive” test based on one scoring system, given a positive test based on the other. As the prevalence of positive tests decreased, positive agreement of CDIS and CAMDEX was variable and less probable. On the other hand, negative agreement was significantly more likely and more consistent. In other words, the CDIS and CAMDEX scoring systems consistently agree when CDTs

are normal, but they agree less consistently when CDTs become more abnormal. Given that the CDIS was more likely to identify an abnormal CDT and reliably defined a normal CDT, its use would be preferred for perioperative research and care.

At this point, it would be helpful to highlight the characteristics of diagnostic tests and their measurement for the clinician as reviewed by Haynes *et al.*<sup>20</sup> Any diagnostic test aims to determine the presence or absence of a condition. Test results can be positive or negative; the condition can be present or absent. A test is said to be sensitive when it correctly identifies a large proportion of those with the condition. A test is said to be specific when it correctly identifies those without the condition. When screening for disease, it is useful to have a test that is maximally sensitive to ensure that all potential cases are identified. When confirming a diagnosis, a highly specific test is preferred to ensure that healthy individuals are not labelled incorrectly as having the disease.

These common measures of test performance assume that we have proof of the presence or absence of the condition from a so-called reference standard. The tools we have selected, i.e., the CAM and the ISPOCD’s Reliable Change Index method, are widely used and accepted tools for perioperative research; however, they are also tests with their own inherent errors and biases.

To place the clinical utility of the CDT in context, it would be helpful to consider the diagnostic characteristics of the CAM. When compared with an evaluation by trained experts using diagnostic criteria for delirium according to the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV)*, the CAM demonstrated a sensitivity > 0.9 while recording interobserver Kappa > 0.9.<sup>12,21</sup> The sensitivity of the CDT in the present study was ≤ 0.60 and should have been ≥ 0.9 to supplant the CAM. The CAM and the CAM-ICU, a modification of the original CAM instrument for use in the intensive care unit, are simple and quick to administer, making them an attractive tool for daily clinical care. Freidman *et al.* reported that the use of the CAM-ICU added only 2.3 min to their daily acute pain service assessments but identified seven of eight cases of delirium not evident to attending ward staff.<sup>22</sup> It would seem, therefore, that the clinician seeking a bedside tool to assess delirium would be better off using the CAM or CAM-ICU rather than the CDT. These findings are corroborated by Adamis *et al.* who concluded that the CDT “is not suitable to specifically detect or monitor delirium among this patient group” (i.e., 94 inpatients on a specialty geriatric service).<sup>23</sup>

Postoperative cognitive dysfunction, on the other hand, is a subtle condition determined only by detailed neuropsychometric tests. The battery of nine tests used in the present study took an hour or more to complete,

<sup>A</sup> A <http://www.john-uebersax.com/stat/kappa.htm#summary>.

highlighting the need for a simple effective bedside screening test for POCD. Observational research demonstrates that POCD is most evident in those tests assessing executive function<sup>7</sup> and attention.<sup>24</sup> Detailed analysis of the cognitive performance of patients in the cohort tested in this study revealed that a decrease of more than one standard deviation in performance was present in only three tests: Trail Making Test (part B), Symbol Digit Modalities Test, and Grooved Pegboard Test.<sup>25</sup> The former two tests are also easy to administer with pen and paper at the bedside, and they are sensitive to changes in visuospatial thought, attention, and executive function. These domains are also highlighted in the CDT<sup>9</sup> and may explain the higher sensitivity of the CDT in the assessment of POCD. If clinicians are willing to accept a substantial number of false positives and negatives, the CDT may provide some utility in the screening of POCD; however, they might also consider using selected tools from the POCD battery.

This study has a number of limitations. First, we must recognize the absence of an agreed standard of reference to base our diagnoses of delirium and POCD. While the DSM-IV provides diagnostic criteria for delirium, it is dependent on a bedside evaluation by an expert to determine “inattention” and “disorganized thinking.” While the individual tests used in our neuropsychometric battery have well-established norms, their synthesis into a diagnosis of POCD is a matter of considerable discussion among researchers. For this reason, we have not applied the Standards for Reporting of Diagnostic Accuracy (STARD) initiative criteria<sup>26</sup> for evaluation of a diagnostic test in reporting our results, and we are hesitant to discuss our findings in terms of sensitivity and specificity alone. Next, we must recognize that our patient population is highly selected and does not reflect the variety of patients presenting for elective non-cardiac surgery. Exclusion of patients with previously diagnosed cognitive disorders or poor performance on the Mini-Mental State Examination ensured that our results could not define the use of the CDT as a preoperative risk prediction tool and that our results would apply only to new onset of delirium or POCD in previously normal individuals. We must also recognize that the perioperative period is dynamic, and a detailed assessment of cognition was performed only at discharge. It is possible that subtle changes in thinking on POD 2 or POD 4 yielded abnormal CDTs that were insufficient to yield a positive CAM. A transient “sub-clinical” change in cognition might lead to a false positive CDT for delirium, but it remains unclear what, if any, impact these changes would have on patient outcome. Finally, our sample was limited, by convenience, to that of the parent study, as we were unable to accurately estimate sample size for the CDT assessments. If we applied the method used by Flahault *et al.*<sup>27</sup> to our observed frequency of POCD, we would

have required 67 cases of POCD and 45 controls to establish a 75% sensitivity of the CDT for detecting POCD with a 95% CI, which would violate a minimally acceptable lower limit of 0.5 with probability < 5%. A sample of 67 cases of delirium and 119 controls would be required if using the same method, sensitivity, and confidence intervals to establish the diagnostic characteristics of the CDT for detecting delirium. Our smaller sample size may have contributed, in part, to the wide confidence limits noted in the estimates of sensitivity and specificity of the CDT for delirium.

In summary, the primary objective of this study was to determine if the CDT was useful in the detection of delirium and POCD following elective non-cardiac surgery. Within the limitation of our sample size, the results of the present study suggest that there is insufficient evidence to adopt the CDT as a screening tool for perioperative care. Knowledge in this field will advance rapidly once investigators agree on a concise set of cognitive domains and tools for both screening and diagnostic evaluation of cognitive function in the perioperative period.

**Funding** This work was funded by the Canadian Anesthesiologists’ Society’s Dr. R.A. Gordon Clinical Research Award (2005) and the University of Ottawa, Department of Anesthesiology’s Chair’s Research Fund. Dr. Bryson was supported by the Ottawa Hospital Anesthesia Alternate Funds Association.

**Disclosure** The authors declare no commercial or non-commercial affiliations that are or may be perceived to be a conflict of interest with the work.

## References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM IV, Forth Edition, Text Revision. Washington, DC: American Psychiatric Association, 2000.
2. Dasgupta M, Dumbrell AC. Preoperative risk assessment for delirium after noncardiac surgery: a systematic review. *J Am Geriatr Soc* 2006; 54: 1578-89.
3. McCusker J, Cole MG, Dendukuri N, Belzile E. Does delirium increase hospital stay? *J Am Geriatr Soc* 2003; 51: 1539-46.
4. Inouye SK, Rushing JT, Foreman MD, Palmer RM, Pompei P. Does delirium contribute to poor hospital outcomes? A three-site epidemiologic study. *J Gen Intern Med* 1998; 13: 234-42.
5. McCusker J, Cole M, Abrahamowicz M, Primeau F, Belzile E. Delirium predicts 12-month mortality. *Arch Intern Med* 2002; 162: 457-63.
6. Rasmussen LS, Larsen K, Houx P, Skovgaard LT, Hanning CD, Moller JT; the ISPOCD group. The assessment of postoperative cognitive function. *Acta Anaesthesiol Scand* 2001; 45: 275-89.
7. Price CC, Garvan CW, Monk TG. Type and severity of cognitive decline in older adults after noncardiac surgery. *Anesthesiology* 2008; 108: 8-17.
8. Steinmetz J, Christensen KB, Lund T, Lohse N, Rasmussen LS; ISPOCD Group. Long-term consequences of postoperative cognitive dysfunction. *Anesthesiology* 2009; 110: 548-55.

9. *Shulman KI*. Clock-drawing: is it the ideal cognitive screening test? *Int J Geriatr Psychiatry* 2000; 15: 548-61.
10. *von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative*. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; 370: 1453-7.
11. *Tombaugh TN, McIntyre NJ*. The mini-mental state examination: a comprehensive review. *J Am Geriatr Soc* 1992; 40: 922-35.
12. *Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegel AP, Horwitz RI*. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med* 1990; 113: 941-8.
13. *Bryson GL, Wyand A, Wozny D, Rees L, Taljaard M, Nathan H*. A prospective cohort study evaluating associations among delirium, postoperative cognitive dysfunction, and apolipoprotein E genotype following open aortic repair. *Can J Anesth* 2011; 58. DOI: [10.1007/s12630-010-9446-6](https://doi.org/10.1007/s12630-010-9446-6).
14. *Murkin JM, Newman SP, Stump DA, Blumenthal JA*. Statement of consensus on assessment of neurobehavioral outcomes after cardiac surgery. *Ann Thorac Surg* 1995; 59: 1289-95.
15. *Roth M, Huppert FA, Mountjoy CQ, Tym E*. *Camdex-R: The Cambridge Examination for Mental Disorders of the Elderly*, 2nd ed. Cambridge, UK: Cambridge University Press, 1998.
16. *Mendez MF, Ala T, Underwood KL*. Development of scoring criteria for the clock drawing task in Alzheimer's disease. *J Am Geriatr Soc* 1992; 40: 1095-9.
17. *Nishiwaki Y, Breeze E, Smeeth L, Bulpitt CJ, Peters R, Fletcher AE*. Validity of the clock-drawing test as a screening tool for cognitive impairment in the elderly. *Am J Epidemiol* 2004; 160: 797-807.
18. *Feinstein AR, Cicchetti DV*. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; 43: 543-9.
19. *Cicchetti DV, Feinstein AR*. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43: 551-8.
20. *Haynes RB, Sackett DL, Guyatt GH, Tugwell P*. *Clinical Epidemiology: How to Do Clinical Practice Research*, 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2005.
21. *Ely EW, Inouye SK, Bernard GR, et al*. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *JAMA* 2001; 286: 2703-10.
22. *Friedman Z, Qin J, Berkenstadt H, Katznelson R*. The confusion assessment method—a tool for delirium detection by the acute pain service. *Pain Pract* 2008; 8: 413-6.
23. *Adams D, Morrison C, Treloar A, Macdonald AJ, Martin FC*. The performance of the clock drawing test in elderly medical inpatients: does it have utility in the identification of delirium? *J Geriatr Psychiatry Neurol* 2005; 18: 129-33.
24. *Silverstein JH, Steinmetz J, Reichenberg A, Harvey PD, Rasmussen LS*. Postoperative cognitive dysfunction in patients with preoperative cognitive impairment: which domains are most vulnerable? *Anesthesiology* 2007; 106: 431-5.
25. *Bryson GL, Wozny D, Rees L, Nathan H*. Distribution of cognitive deficits following open aortic repair. *Can J Anesth* 2009; 56: S6 (abstract).
26. *Bossuyt PM, Reitsma JB, Bruns DE, et al*. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138: W1-12.
27. *Flahault A, Cadilhac M, Thomas G*. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005; 58: 859-62.