Received: 6 June 2017 | Revised: 21 May 2018 | Accepted: 5 June 2018

DOI: 10.1111/jan.13810

REVIEW PAPER

WILEY JAN Leading Global Nursing Research

The clock drawing test: A systematic review and meta-analysis of diagnostic accuracy

JinKyung Park | EunHye Jeong | GyeongAe Seomun

College of Nursing, Korea University, Seoul, Korea

Correspondence
GyeongAe Seomun, College of Nursing, Korea University, 145, Anam-ro, Seongbuk-Gu, Seoul 02841, Korea.
Email: seomun@korea.ac.kr

Abstract

Aim: The aim of this study was to evaluate the accuracy of the clock drawing test and to compare its scoring methods.

Background: Dementia refers to a symptom where an adult demonstrates memory disorder and cognitive impairment. Early diagnosis of dementia is very important for medication management and prognosis. The clock drawing test is one of the most used cognitive screening tools for dementia. However, due to its scoring system, the accuracy of the clock drawing test remains a topic of debate.

Design: A systematic review with meta-analysis following Cochrane's methods and the guidelines of the Agency for Healthcare Research and Quality.

Data sources: A literature search was carried out in the OVID-MEDLINE, OVID-EMBASE and OVID-PsycINFO databases (27 October 2015).

Review method: The quality assessment of diagnostic accuracy studies (QUADAS-2) tool was employed for this review. We used hierarchical regression models to pool the values of diagnostic accuracy in a random effects model.

Results: A total of 18 studies with 5,531 participants were identified for this review. Fifteen of these studies were included for meta-analysis according to each scoring system. The pooled sensitivity and specificity of the clock drawing test using the Shulman system were 82% and 75.7% respectively. In the Sunderland system, these were 72.6% and 87.9% respectively.

Conclusions: The findings indicate that the accuracy of the clock drawing test using the Shulman system was the most studied and highly sensitive. After gaining a better understanding of the clock drawing test through this study, we recommend it for widespread use in the diagnosis of dementia.

KEYWORDS

clock drawing test, cognitive impairment, dementia, diagnostic accuracy, literature review, meta-analysis, sensitivity, specificity, systematic review

1 | INTRODUCTION

Dementia refers to a symptom where an adult with normal cognitive development demonstrates memory disorder and cognitive impairment to the extent that it interferes with their normal activities of daily living. We are living in a globally ageing society, where dementia has become a common illness and disability in everyday life among elders (DESA 2000; Fiest et al., 2016; Kronhaus, Fuller, Zimmerman, & Reed, 2016; Lobo et al., 2000). Early diagnosis of dementia is available and advance warning of this disease may allow time to plan and implement treatment. Furthermore, early diagnosis offers an opportunity for medication management with the hope of

enhancing cognitive function and behavioural strategies (Kronhaus et al., 2016; Laakkonen et al., 2016; Molnar, Patel, Marshall, Man-Son-Hing, & Wilson, 2006). Dementia, delirium, and depression are the most common and challenging diagnoses for older adults. The differential diagnosis is very important and hinges on a careful clinical evaluation because there are no pathognomonic diagnostic tests of these diseases (Downing, Caprio, & Lyness, 2013; Johnson, Sims, & Gottlieb, 1994).

There are many cognitive instruments and diagnostic criteria for the evaluation of cognitive impairment and dementia (Levey, Lah, Goldstein, Steenland, & Bliwise, 2006). The most widely used tool will be the mini-mental state examination (MMSE) (Folstein, Folstein, & McHugh, 1975; Hesson & Pichler, 2016; Shulman et al., 2006). The disadvantage of MMSE, however, is that it is language-based (Spreen & Strauss, 1998) and scores can be lowered if not managed in the native language of the individual. MMSE is also considered to be influenced by the level of education, so lower levels of education, especially language-based items, will have lower scores (Espino, Lichtenstein, Palmer, & Hazuda, 2004; Hesson & Pichler, 2016; Kliegel, Zimprich, & Rott, 2004; Spreen & Strauss, 1998). The clock drawing test (CDT), cognitive screening tool, does not require language ability and performance and is a tool that can compensate for the shortcomings of MMSE because it is hardly affected by the level of education. The CDT is easy to administrate (Shulman et al., 2006) and less affected by depression or dysphoria (Gruber, Varner, Chen, & Lesser, 1997; Herrmann et al., 1998).

1.1 | Background

The CDT is one of the most used cognitive screening instruments for dementia and it has been efficiently accepted among clinicians due to ease of use and reduced administration time (Amodeo, Mainland, Herrmann, & Shulman, 2015; Kim & Chey, 2010; Shulman, 2000; Shulman et al., 2006; Souillard-Mandar et al., 2016; Sugawara et al., 2010). The CDT, which has been in use since the 1950s, is a valid screening tool for dementia because it correlates highly with global cognitive function (Amodeo et al., 2015; Shulman, 2000; Shulman et al., 2006; Wolf-Klein, Silverstone, Levy, & Brod, 1989). A variety of cognitive functions is assessed by the CDT, including orientation, selective and persistent attention, auditory comprehension, verbal memory, numerical knowledge, visual memory and reconstruction, visuospatial organization, and motor performance (praxis) (Royall, 1996; Shulman, 2000; Shulman et al., 2006). The CDT is a valuable instrument because it is simple, acceptable to patients, independent of education/cultural/language difference, easy management, psychometrically robust, and wide range of cognitive domains (Ismail, Rajji, & Shulman, 2010; Shulman, 2000).

The CDT is primarily administered by doctors and nurses in primary care settings (Chen, Leung, & Chen, 2011; Lee, Kim, Choi, Huh, & Park, 2015; Shulman, 2000). Research was conducted in England to test the validity of the CDT as a cognitive impairment screening test for elders. A large-scale prospective trial was performed with 13,557 participants. The findings showed the CDT's sensitivity and specificity as 77% and 87%, respectively, and its diagnostic accuracy was higher when administered by a nurse (Nishiwaki et al., 2004).

Why is this research or review needed?

- Dementia refers to a symptom where an adult demonstrates memory disorder and cognitive impairment. Early diagnosis of dementia is available and advance warning of this disease may allow time to plan and implement treatment. Furthermore, early diagnosis offers an opportunity for medication management with the hope of enhancing cognitive function and behavioural strategies
- The clock drawing test is one of the most used cognitive screening instruments for dementia and it has been efficiently accepted among clinicians due to ease of use and reduced administration time and it can be performed without being influenced by the patient's level of language or education.
- The accuracy of the clock drawing test remains a subject for debate because of its scoring system. However, to date, no systematic review and meta-analysis on the diagnostic accuracy of clock drawing test by scoring system has been conducted.

What are the key findings?

- Through a systematic review and meta-analysis, the diagnostic accuracy of the clock drawing test was found to be appropriately sensitive as a cognitive screening tool for dementia
- The Shulman scoring system could be interpreted as the most accurate method when using the clock drawing test for a dementia screening test because of its high level of sensitivity.
- The Shulman and Sunderland scoring systems had the highest predictive values, making them valuable when using the clock drawing test for clinical purposes.

How should the findings be used to influence policy/practice/research/education?

- For early diagnosis of dementia, nurses, who are at the front lines of medical assessment, should conduct cognitive screening tests with complete understanding of their methods and accuracy.
- The clock drawing test is simple and easy for patients to perform, takes a short amount of time, and has a simple assessment method. It is a valuable cognitive screening tool for nurses to perform. Therefore, it is important for clinicians to have a thorough understanding of the diagnostic accuracy of the clock drawing test by scoring system.

Nurses are at the front lines of medical assessment and spend the most time observing and interacting with patients. Therefore, a cognitive impairment test using CDT, which nurses can easily evaluate, is important and should be widely used.

Several methods have been proposed for the CDT scoring system (Supporting Information Appendix S1); however, there are no universally accepted standards (Mainland, Amodeo, & Shulman, 2014; Ricci et al., 2016; Souillard-Mandar et al., 2016). Each scoring system uses a different methodology and instructions for clock drawing. When comparing the scoring systems, there was no consistently superior system in terms of predictive validity (Mainland et al., 2014). Although the scoring systems have something in common, each CDT scoring system can reflect different brain-damaged areas (Matsuoka et al., 2011). Previous studies on the scoring systems showed significant differences in diagnostic accuracy, such as reliability, sensitivity, and specificity (Jouk & Tuokko, 2012; Juby, Tench, & Baker, 2002; Mainland et al., 2014; Tuokko, Hadjistavropoulos, Rae, & O'Rourke, 2000). To use each scoring system adequately and effectively, clinicians suggest using as much information as possible in clinical settings since there are many different CDT scoring systems (Jouk & Tuokko, 2012; Tuokko et al., 2000). Although the CDT has been widely used as a tool for diagnosing cognitive impairment, there has been no systematic review of literature or meta-analysis on the diagnostic accuracy of the tool and the accuracy of the CDT remains a subject for debate due to its scoring system.

2 | THE REVIEW

2.1 | Aims

The aims of this study were to systematically review and conduct meta-analysis to evaluate the accuracy of the CDT cognitive screening test and to compare the scoring methods.

2.2 | Design

We conducted a systematic review and meta-analysis following Cochrane's methods and the guidelines of the Agency for Healthcare and Research Quality (Rector, Taylor, & Wilt, 2012). This systematic review method has been reported according to the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) (Liberati et al., 2009).

2.3 | Search methods

We adopted the recommendation of the Cochrane Collaboration (Kim, Lee, Choi, Huh, & Park, 2015) and the AHRQ for the search terms and the search strategy (Rector et al., 2012). To obtain highly sensitive results, searches were performed using only CDT-related terms, except for the terms related to patients, reference tests, and outcomes (Supporting Information Appendix S2). A literature search was conducted in Ovid-MEDLINE (1946 to 3rd week of October 2015), Ovid-EMBASE(1988 to 43rd week of 2015), and Ovid-PsycINFO (1987 to 3rd week of October 2015). References of the retrieved studies were also reviewed for additional studies.

All of the following inclusion criteria needed to be met: (a) the study was designed as an observational, cross-sectional, or case series study; (b) the CDT was used to screen the cognitive impairment of patients with dementia and Alzheimer's disease; (c) diagnostic accuracy estimates were reported in the paper, including sensitivity, specificity, true positive (TP), false positive (FP), true negative (TN), and false negative (FN) or had sufficient detail to derive these numbers; (d) the reference test was the Diagnostic and Statistical Manual of Mental Disorders (DSM) III, IV, or neuropsychologist assessment; (e) The CDT scoring system was reported in two or more publications; and (f) the study was written in English. We excluded abstracts presented at congresses, reviews, letters, editorials, and unpublished data.

2.4 | Search outcomes

The systematic search yielded a total of 427 studies. From this total, 343 studies were excluded by reviewing the title and abstract to eliminate duplicates. Titles and abstracts were reviewed independently by two authors using predefined criteria for inclusion and exclusion. We analysed a total of 84 full-text studies. Disagreements were resolved by consensus. When searching, the scoring system was not limited. However, we only included scoring systems with two or more literature references on diagnostic accuracy while selecting or eliminating literature. Figure 1 shows the literature selection process.

2.5 | Quality appraisal

The quality assessment of diagnostic accuracy studies (QUADAS-2) tool (Liberati et al., 2009) was adopted for this systematic review to assess the quality of included studies. In this systematic review, we determined that a low risk of bias was supposed that all answers were “yes”. If an answer was either “no” or “unclear,” there was a high risk of bias or unclear bias.

The risk of bias of the included studies was assessed by the two authors independently, discrepancies were discussed, and a consensus was reached for all domains. A third reviewer who was a methodologist for a systematic review was consulted if discrepancies remained.

2.6 | Data extraction

Two authors independently extracted the relevant data from each study. Disagreements were resolved through discussion. For each study, we recorded the first author, year of publication, country of origin, eligible population size, study population characteristics, name of the scoring system, and diagnostic cut-off point. Variables sought were TP, FP, FN, and TN. The summary estimates were sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratio, and diagnostic odds ratio (DOR).

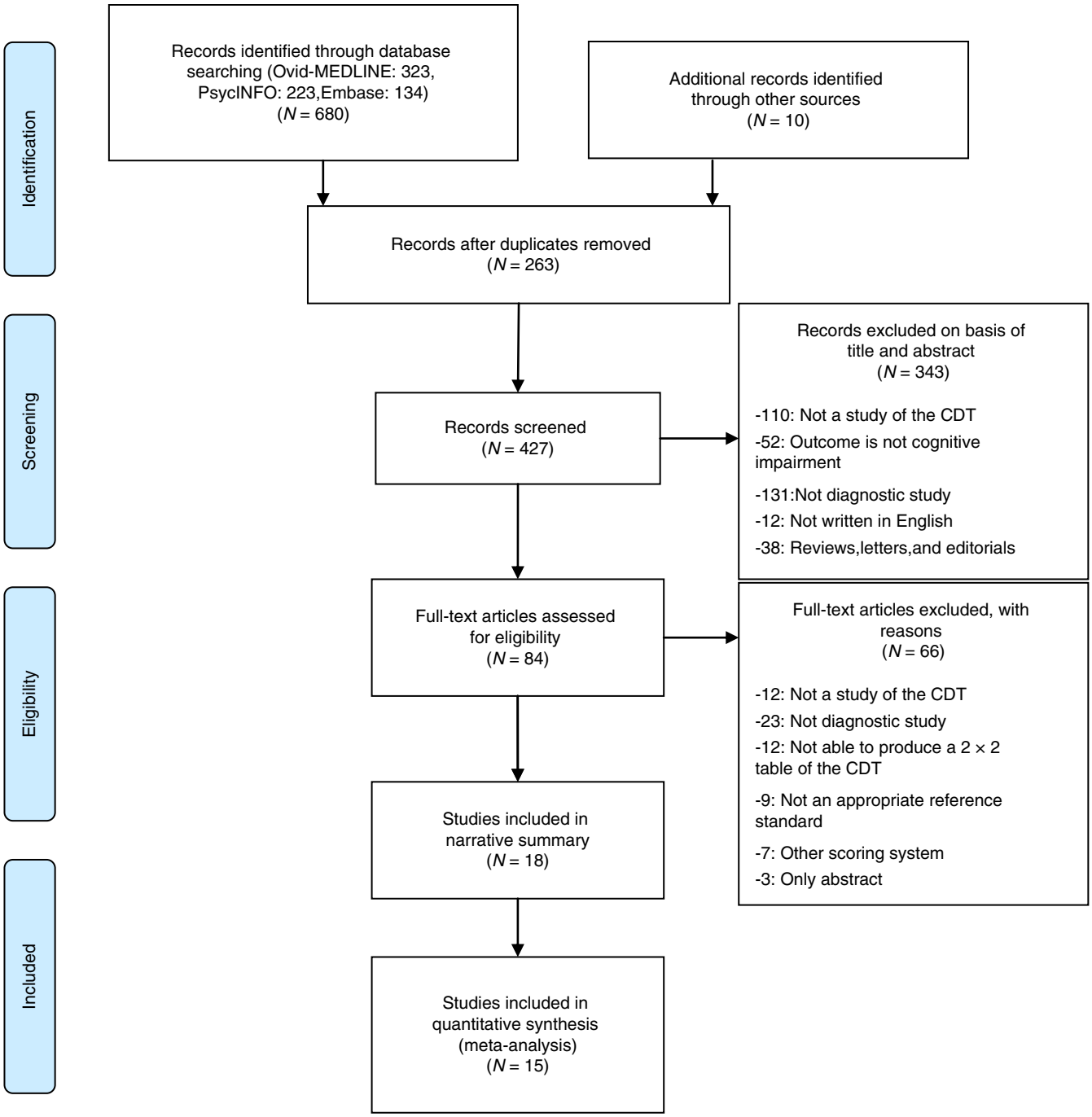

FIGURE 1 The flow chart of the search for eligible studies (PRISMA flow diagram) [Colour figure can be viewed at wileyonlinelibrary.com]

2.7 | Synthesis

We used the hierarchical regression models to pool the values for TP, FP, TN, and FN in a random effects model (Levey et al., 2006; Shulman et al., 2006; Spreen & Strauss, 1998). This bivariate approach described for the potential between-study heterogeneity and included the correlation between sensitivity and specificity (Kim et al., 2015; Manos & Wu, 1994; Powell, Towers, & Milne, 2008). Unlike the traditional method that uses straight lines for pooled estimates, the bivariate approaches generate an ellipse with a pooled mean sensitivity and specificity and a 95% confidence interval. This model is focused on estimating an average sensitivity and specificity with an estimated unexplained variation of the parameters and the correlation between them. Hierarchical summary receiver operating characteristic (HSROC) curves further provide 95% confidence and 95% prediction regions. As summary measures, a diagnostic odds ratio (DOR), positive likelihood ratio (PLR), and negative likelihood ratio (NLR) were calculated using the pooled sensitivity and specificity (Kim et al., 2015; Lee et al., 2015; Powell, et al., 2008). All statistical analyses of this study were carried out in R version 3.2.2 (Lee et al., 2015).

3 | RESULTS

3.1 | Study characteristics

A total of 18 studies (Aprahamian, Martinelli, Neri, & Yassuda, 2010; Berger, Frolich, Weber, & Pantel, 2008; Cacho et al., 2010; Chen et al., 2011; Chiu, Li, Lin, Chiu, & Liu, 2008; Ehreke, Luppa, Konig, Villringer, & Riedel-Heller, 2011; Henderson, Scott, & Hotopf, 2007; Jouk & Tuokko, 2012; Kirby, Denihan, Bruce, Coakley, & Lawlor, 2001; Lee, Swanwick, Coen, & Lawlor, 1996; Lee et al., 2008; Milian et al., 2012; Nunes et al., 2008; Ramlall, Chipps, Bhigjee, & Pillay, 2013; Riedel, Klotsche, Forstl, & Wittchen, 2013; Russo et al., 2014; Yamamoto et al., 2004; Zhou & Jia, 2008) were identified for this systematic review (Table 1). Fifteen of these studies were included for meta-analysis on the diagnostic accuracy of the CDT for evaluating cognitive impairment (Figure 1). The included studies were published between 1996 and 2014 and were conducted throughout the world, including Europe, Africa, Asia, and America.

Seven of these studies used the Shulman scoring system, five studies used the Rouleau scoring system, and five used the Sunderland scoring system. There were also studies using the Watson, Manos, Freedmen, and Wolf-Klein scoring systems.

3.2 | Risk of bias assessment

We employed four domains of assessment from the QUADAS-2 assessment tool to assess the risk of bias for 18 studies. Each domain had 2–3 (signalling) questions, where the risk of bias was

TABLE 1 Characteristics of included studies

Author	Year	Place of study	Study design	*N*	Age (mean, range)	Cognitive impairment disease	Reference standard	CDT scoring system	TP	FP	FN	TN
Lee	1996	Ireland	Case–control	60	70.0	Alzheimer's disease	Neuropsychologist assessment	Sunderland	20	1	10	29
Kirby	2001	Ireland	Cross-sectional	564	75.2	Dementia	DSM III	Sunderland	31	99	10	424
Yamamoto	2004	Japan	Cross-sectional	219	75.1	Mild cognitive impairment	DSM III	Sunderland	109	2	69	39
								Rouleau	101	3	77	38
Henderson	2007	United Kingdom	Cross-sectional	82	59–81	Dementia	DSM IV	Manos	23	15	2	42
Lee	2008	Korea	Cross-sectional	465	70.9	Mild cognitive impairment	Neuropsychologist assessment	Rouleau	126	68	98	173
								Freedman	91	41	133	200
Berger	2008	Germany	Cross-sectional	462	71.5	Dementia	DSM IV	Shulman	300	56	34	72
								Watson	240	46	94	82
								Manos	271	51	63	77
								Wolf-Klein	194	24	140	104
Chiu	2008	Taiwan	Cross-sectional	74	75.6	Dementia	DSM IV	Rouleau	25	14	9	26
Nunes	2008	Brazil	Case–control	92	69.6	Dementia	DSM IV	Sunderland	25	21	5	41
Zhou	2008	China	Cross-sectional	160	66.4	Cognitive impairment	DSM IV	Rouleau	55	17	25	63
Cacho	2010	Spain	Case–control	132	72.9	Alzheimer's disease	DSM IV	Shulman	48	2	18	64
Aprahamian	2010	Brazil	Case–control	220	77.6	Alzheimer's disease	DSM IV	Shulman	104	33	17	66
								Sunderland	90	10	31	89
Ehreke	2011	Germany	Cross-sectional	384	83.2	Dementia	DSM IV	Shulman	19	125	9	231
Chen	2011	Taiwan	Case–control	188	77.0	Dementia	Neuropsychologist assessment	Watson	30	3	48	107
Jouk	2012	Canada	Case–control	356	78.1	Dementia	DSM III-R	Shulman	257	42	19	38
								Watson	163	26	113	54
								Wolf-Klein	204	22	72	58
Milian	2012	Germany	Cross-sectional	502	74.8	Dementia	DSM IV	Shulman	342	2	96	62
Riedel	2013	Germany	Cross-sectional	1383	70.5	Dementia	DSM IV	Shulman	417	247	173	546
Ramlall	2013	South Africa	Case–control	102	75.2	Dementia	DSM IV	Rouleau	5	10	6	81
Russo	2014	Argentina	Cross-sectional	86	73.8	Dementia	DSM IV	Freedman	36	1	20	29

Note. CDT: clock drawing test; N: number of population; TP: true positive; FP: false positive; FN: false negative; TN: true negative.

judged "low" when all the questions were answered "yes." In the patient domain, 10 of 18 studies showed "high risk" because the participants of five of the studies were not (simple) random samples but participants in a case–control study.

In the index test domain, seven studies that did not explain whether the assessor was blinded to the reference standard result were assigned as "high risk" for bias. In the reference standard domain, studies where the neuropsychologist diagnosed dementia with DSM III, IV, or other validated criteria scored "low risk" of bias. Additionally, three studies were assessed as "unclear" risk of bias because while the results were not clearly delineated, they could be viewed as a blinded index test result. In the flow and timing domain, all papers except for one included all participants in the analysis and had the same reference standard test; thus, it was assigned a "low risk" of bias. As presented in Table 2, most of the included studies were assessed to be of high applicability in the patient selection or risk of bias domains according to the QUADAS-2 assessment tool. In terms of risk of bias assessment, there were uncertain risks in the index test domain. In four studies, the results of the reference standard during the performance and reading of the index test were not presented blindly. In almost all studies, there was a low risk in the domain of flow and timing, but there was a high risk in one study because the time interval of the index test and the reference standard were not presented.

3.3 | Diagnostic accuracy of the CDT for each scoring system

The number of studies on the diagnostic accuracy of the CDT test by each scoring system is as follows: seven studies with 3,439 participants used the Shulman scoring system; five studies with 1,020 participants used the Rouleau scoring system; five studies with 1,155 participants used the Sunderland scoring system; and three studies with 1,006 participants used the Watson scoring system. The Manos, Freedman, and Wolf-Klein scoring systems had two studies each, which was an insufficient number for meta-analysis to be conducted.

As shown in Table 3, the sensitivity of CDT using the Manos scoring system was 92.1% and 81.1% and the specificity was 73.7% and 60.2%, respectively. The sensitivity of CDT using the Freedman scoring system was 57.3% and 40.6% and the specificity was 96.7% and 83.0%, respectively. The sensitivity of CDT using the Wolf-Klein scoring system was 73.9% and 58.1% and the specificity was 81.2% and 72.5%, respectively.

3.3.1 | Diagnostic accuracy of the CDT using the Shulman scoring system

When considering all seven studies, the pooled estimates of the CDT using the Shulman scoring system are placed at a sensitivity of 82% (95% confidence interval [CI]: 73.1–88.4%) and a specificity of 75.7% (95% CI: 54.3–89.1%). The results showed that the HSROC curve was positioned in the desirable upper left corner, the confidence region was small, and the overall weighted area under the curve was 0.86 (Table 3).

The HSROC curve presents a global summary of the test performance and shows the trade-off between sensitivity and specificity. The HSROC curve of the CDT using the Shulman scoring system for the diagnosis of dementia is shown in Figure 2.

TABLE 2 The results of risk of bias assessment for each included study

		Risk of bias				Applicability concerns		
Author	Year	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Lee	1996	High	High	Low	Low	Low	Low	Low
Kirby	2001	High	Low	Unclear	Low	Low	Low	Low
Yamamoto	2004	High	Low	Low	Low	Low	Low	Low
Henderson	2007	High	Low	Unclear	Low	Low	Low	Low
Lee	2008	Low	Low	Low	Low	Low	Low	Low
Berger	2008	Low	Low	Unclear	Low	Low	Low	Low
Chiu	2008	Low	Low	Low	Low	Low	Low	Low
Nunes	2008	High	High	Low	Low	Low	Low	Low
Zhou	2008	High	Low	Low	Low	Low	Low	Low
Cacho	2010	High	High	Low	Low	Low	Low	Low
Aprahamian	2010	High	High	Low	Low	Low	Low	Low
Ehreke	2011	Low	Low	Low	Low	Low	Low	Low
Chen	2011	Low	High	Low	Low	Low	Low	Low
Jouk	2012	High	High	Low	High	Low	Low	Low
Milian	2012	High	Low	Low	Low	Low	Low	Low
Riedel	2013	Low	Low	Low	Low	Low	Low	Low
Ramlall	2013	Low	High	Low	Low	Low	Low	Low
Russo	2014	Low	Low	Low	Low	Low	Low	Low

TABLE 3 Diagnostic accuracy of the CDT according to scoring system

Author	Year	Sample size	Sn (95% CI)	Sp (95% CI)	DOR (95% CI)	PLR (95% CI)	NLR (95% CI)
Shulman scoring system							
Berger	2008	462	89.8 (86.1–92.6)	56.2 (47.6–64.5)	11.345 (6.896–18.662)	2.053 ((1.681–2.507)	0.181 (0.127–0.258)
Cacho	2010	132	72.7 (61.0–82.0)	97.0 (89.6–99.2)	85.333 (18.888–385.526)	24.000 (6.082–94.705)	0.281 (0.189–0.418)
Aprahamian	2010	220	86.0 (78.6–91.0)	66.7 (56.9–75.2)	12.235 (6.315–23.707)	2.579 (1.934–3.438)	0.211 (0.133–0.335)
Ehreke	2011	384	67.9 (49.3–82.1)	64.9 (59.8–69.7)	3.901 ((1.714–8.879)	1.933 (1.444–2.586)	0.495 (0.288–0.853)
Jouk	2012	356	93.1 (89.5–95.5)	47.5 (36.9–58.3)	12.238 (6.453–23.211)	1.774 (1.436–2.190)	0.145 (0.089–0.237)
Riedel	2013	1383	70.7 (66.9–74.2)	68.9 (65.5–72.0)	5.328 (4.223–6.722)	2.269 (2.021–2.548)	0.426 (0.373–0.487)
Milian	2013	502	78.1 (74.0–81.7)	96.9 (89.3–99.1)	110.438 (26.530–459.717)	24.986 (6.381–97.837)	0.226 (0.189–0.271)
Pooled estimates[a]		3439	82.0 (73.1–88.4)	75.7 (54.3–89.1)	AUC: 0.857		
Rouleau scoring system							
Yamamoto	2004	219	56.7 (49.4–63.8)	92.7 (80.6–97.5)	16.615 (4.943–55.842)	7.755 (2.589–23.225)	0.467 (0.386–0.564)
Chiu	2008	74	73.5 (56.9–85.4)	65.0 (49.5–77.9)	5.159 (1.895–14.041)	2.101 (1.316–3.355)	0.407 (0.222–0.745)
Lee	2008	465	56.2 (49.7–62.6)	71.8 (65.8–77.1)	3.271 (2.225–4.808)	1.994 (1.581–2.514)	0.609 (0.515–0.721)
Zhou	2008	160	68.8 ((57.9–77.8)	78.8 (68.6–86.3)	8.153 (3.991–16.657)	3.235 (2.069–5.059)	0.397 (0.281–0.560)
Ramlall	2008	102	45.5 (21.3–72.0)	89.0 (80.9–93.9)	6.750 (1.739–26.208)	4.136 (1.729–9.896)	0.613 (0.356–1.056)
Pooled estimates[a]		1020	60.5 (52.4–68.1)	80.1 (68.0–88.4)	AUC: 0.709		
Sunderland scoring system							
Lee	1996	60	66.7 (48.8–80.8)	96.7 (83.3–99.4)	58.000 (6.871–489.581)	20.000 (2.864–139.672)	0.345 (0.207–0.574)
Kirby	2001	564	75.6 (60.7–86.2)	81.1 (77.5–84.2)	13.277 (6.299–27.984)	3.994 (3.116–5.120)	0.301 (0.175–0.517)
Yamamoto	2004	219	61.2 (53.9–68.1)	95.1 (83.9–98.7)	30.804 (7.207–131.669)	12.553 (3.233–48.750)	0.408 (0.335–0.496)
Numes	2008	92	83.3 (66.4–92.7)	66.1 (53.7–76.7)	9.762 (3.266–29.174)	2.460 (1.678–3.608)	0.252 (0.111–0.572)
Aprahamian	2010	220	74.4 (65.9–81.3)	89.9 (82.4–94.4)	25.839 (11.956–55.842)	7.364 (4.054–13.376)	0.285 (0.209–0.389)
Pooled estimates[a]		1155	72.6 (62.9–80.6)	87.9 (74.0–94.8)	AUC: 0.835		
Watson scoring system							
Berger	2008	462	71.9 (66.8–76.4)	64.1 (55.5–71.9)	4.551 (2.953–7.016)	1.999 (1.572–2.544)	0.439 (0.354–0.545)
Chen	2011	188	38.5 (28.4–49.6)	97.3 (92.3–99.1)	22.292 (6.485–76.624)	14.103 (4.462–44.575)	0.633 (0.529–0.756)
Jouk	2012	356	59.1 (53.2–64.7)	67.5 (56.6–76.8)	2.996 (1.771–5.068)	1.817 (1.305–2.529)	0.607 (0.493–0.747)
Pooled estimates[a]		1006	56.9 (36.8–74.9)	82.7 (43.9–96.7)	AUC: 0.7		
Manos scoring system							
Henderson	2007	82	92.0 (75.0–97.8)	73.7 (61.0–83.4)	32.200 (6.763–153.318)	3.496 (2.230–5.480)	0.109 (0.028–0.414)
Berger	2008	462	81.1 (76.6–85.0)	60.2 (51.5–68.2)	6.495 (4.151–10.161)	2.036 (1.636–2.535)	0.314 (0.241–0.408)
Freedman scoring system							
Lee	2008	465	40.6 (34.4–47.2)	83.0 (77.7–87.2)	3.338 (2.173–5.125)	2.388 (1.733–3.291)	0.715 (0.633–0.809)
Russo	2014	147	57.3 (48.2–65.9)	96.7 (83.3–99.4)	38.860 (5.120–294.950)	17.179 (2.485–118.758)	0.442 (0.355–0.551)
Wolf-Klein scoring system							
Berger	2008	462	58.1 (52.7–63.3)	81.2 (73.6–87.1)	6.005 (3.663–9.843)	3.098 (2.136–4.494)	0.516 (0.443–0.600)
Jouk	2012	356	73.9 (68.4–78.7)	72.5 (61.9–81.1)	7.470 (4.269–13.070)	2.688 (1.870–3.863)	0.360 (0.283–0.457)

Note. CI: confidence interval; Sn: sensitivity; Sp: specificity; DOR: diagnostic odds ratio; PLR: positive likelihood ratio; NLR: negative likelihood ratio; AUC: area under curve.
[a]Bivariate random effect.

3.3.2 | Diagnostic accuracy of the CDT using the Rouleau scoring system

When considering all five studies, the pooled estimates of the CDT using the Rouleau scoring system are placed at a sensitivity of 60.5% (95% CI: 52.4–68.1%) and a specificity of 80.1% (95% CI: 68.0–88.4%) (Table 3). The HSROC curve of the CDT using the Rouleau scoring system is shown in Figure 2. The results showed that the HSROC curve was positioned near the desirable upper left corner and that the overall weighted area under the curve was 0.709.

3.3.3 | Diagnostic accuracy of the CDT using the Sunderland scoring system

When considering all five studies, the pooled estimates of the CDT using the Sunderland scoring system are placed at a sensitivity of 72.6% (95% CI: 62.9–80.6%) and a specificity of 87.9% (95% CI: 74.0–94.8%) (Table 3). The HSROC curve of the CDT using the Sunderland scoring system is shown in Figure 2. The results showed that the HSROC curve was positioned in the desirable upper left corner, the confidence region was small, and the overall weighted area under the curve was 0.84.

3.3.4 | Diagnostic accuracy of the CDT using the Watson scoring system

When considering all three studies, the pooled estimates of the CDT using the Watson scoring system are placed at a sensitivity of 56.9% (95% CI: 36.8–74.9%) and a specificity of 82.7% (95%

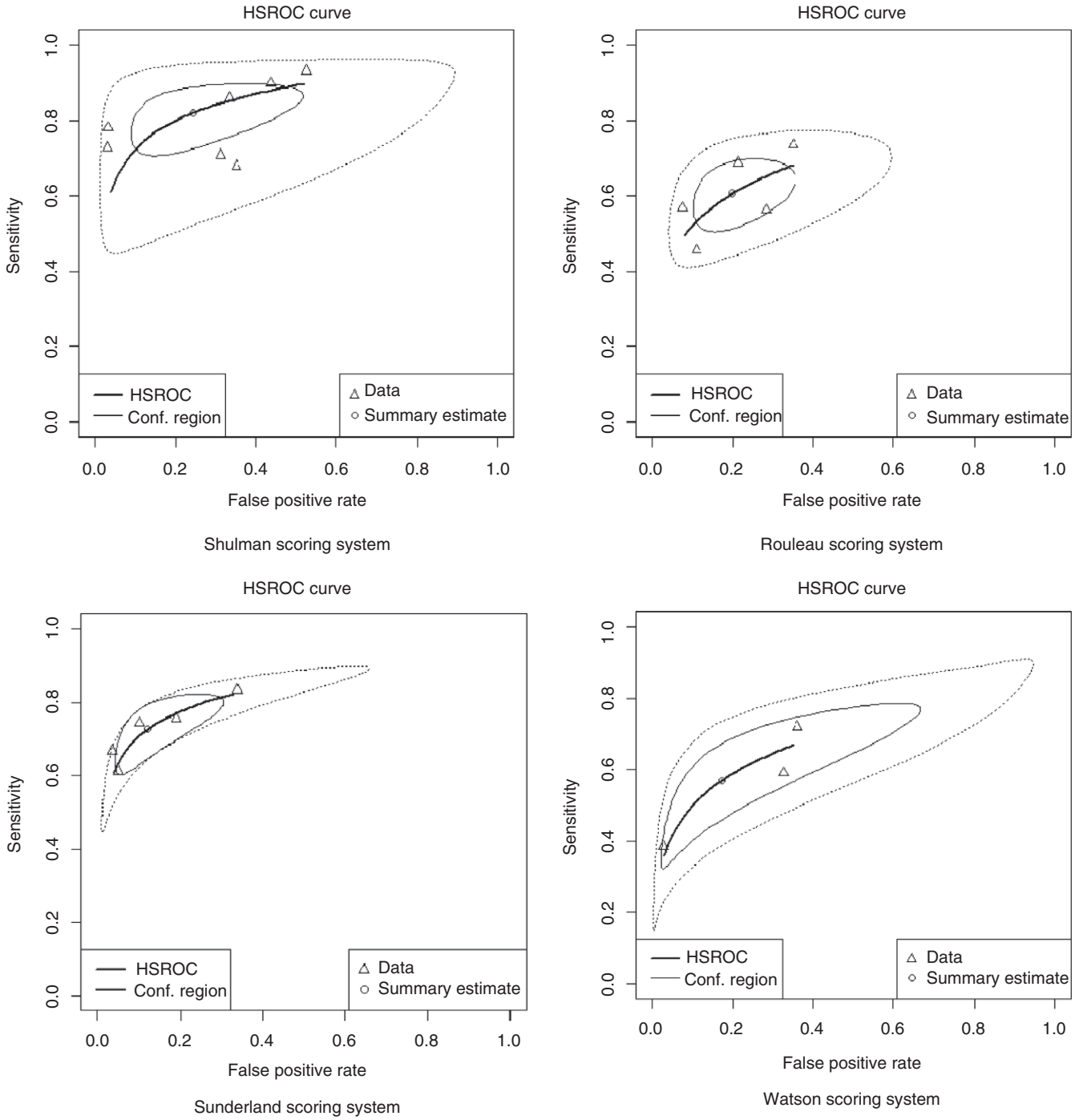

FIGURE 2 The hierarchical summary receiver operating characteristics curve of clock drawing test according to scoring system

CI: 43.9–96.7%) (Table 3). The HSROC curve of the CDT using the Watson scoring system is shown in Figure 2. The results showed that the HSROC curve was positioned near the desirable upper left corner and that the overall weighted area under the curve was 0.70.

4 | DISCUSSION

The CDT is the one of the most used tools in evaluating cognitive impairment. It has the advantage of easy application and a short lead time without being affected by language, education level, or cultural background. Many studies have been conducted on the diagnostic accuracy of using the CDT to assess the degree of cognitive impairment caused by dementia, using the various CDT scoring systems that have been developed. To the best of our knowledge, however, there are no studies that draw on the integrated results of the accuracy of the CDT. This research systematically reviewed the diagnostic accuracy of the CDT. We carried out meta-analysis for each scoring system and compared their diagnostic accuracy. We identified 18 papers that studied the accuracy of the CDT in 5,531 participants. These studies were published between 1996-2014 and were conducted throughout the world, including Europe, Africa, Asia, and America.

This systematic review extracted pooled estimates of diagnostic accuracy. Based on random effects models, this bivariate approach described for potential between study heterogeneity and included the correlation between the sensitivity and the specificity. The overall sensitivity and specificity and their 95% CIs were calculated found on the binominal distributions of the TPs and TNs. The findings show that the Shulman scoring system had a sensitivity, specificity, and area under the curve of 82%, 75.7%, and 0.857, respectively, which was the most accurate among the scoring systems.

The Shulman scoring system also showed the highest sensitivity at 82%, which suggests that a negative CDT rules out the possibility of dementia. Other scoring systems, such as Sunderland, Rouleau, and Watson, showed comparatively low sensitivity at 72.0%, 60.5%, and 56.9%, respectively.

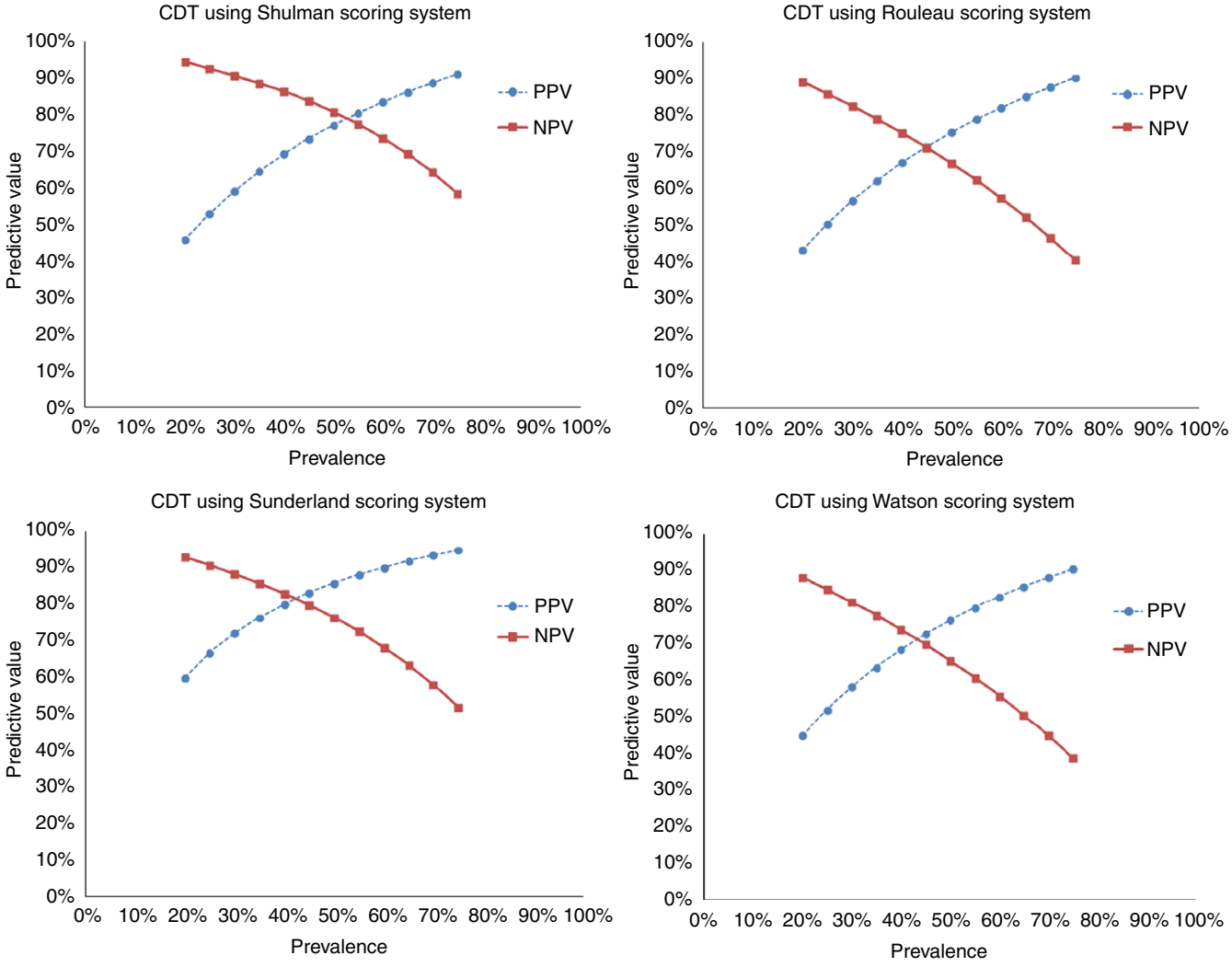

FIGURE 3 The predictive value of the clock drawing test according to scoring systems

Therefore, the Shulman scoring system is recommended for use with the CDT as a screening test for dementia because it is the most studied and highly sensitive system. However, the Sunderland scoring system showed the highest specificity at 87.9%, meaning that it demonstrates highest accuracy for detecting dementia.

PPV is the probability of having a disease in the positive outcome and NPV is the probability of no disease in the negative outcome. These values are very useful because they provide clinicians the likelihood of a disease. If the test results show high PPV or NPV, the clinician can explain the condition to patients using the test result. However, PPV and NPV are affected by disease prevalence, which may be inconsistent among studies. Therefore, the values are often calculated and plotted according to the range of prevalence (Spreen & Strauss, 1998). We have calculated from the pooled sensitivity and specificity for a range of prevalence values in the included studies. As a result, according to a dementia prevalence of 10%–75%, the PPVs and NPVs for the CDT using each scoring system are described in Figure 3.

The PPV and NPV of CDT using the Shulman scoring system was 80% when the prevalence was approximately 55%. The Sunderland scoring system showed 80% for the PPV and NPV of CDT when the prevalence was approximately 45%. The Watson scoring system showed 70% for the PPV and NPV of CDT for a prevalence of 45% and 70% when the prevalence was approximately 55% (Figure 3). Therefore, if the approximate prevalence of a population is known, the practitioners can choose a scoring system with a high predictive value in that prevalence group. Predictive values are influenced by disease prevalence, so the range of prevalence was first calculated to derive the PPV and NPV within that range.

There are many tests for diagnosing dementia. Among these, the MMSE is the most commonly used test and many studies have demonstrated that its sensitivity ranges from 85% to 87% and its specificity from 82% to 90% (Espino et al., 2004). However, the MMSE is based on a language component, so low scores may be obtained when the patient's language level is low. The level of education has also been reported to affect the results (Espino et al., 2004; Kliegel et al., 2004; Shulman, 2000). Therefore, it is important to evaluate the degree of cognitive disorders using the CDT, which is easy to perform without being influenced by the patient's language or education level. Additionally, several studies have reported that the diagnostic accuracy was higher when the MMSE was conducted along with the CDT compared with a single test.

Assessment of "risk of bias" on the included studies was carried out using QUADAS-2, which showed low risk of bias overall. However, 10 of the included studies showed a high risk of bias in the patient selection domain due to the case–control study design. The index domain of included studies that had a case–control study design showed high risk of bias. In all but one of the studies, the participants were confirmed to have been assessed by the same reference standard test, so the risk of bias for the flow and timing domain was low as well.

The first limitation of this study is that meta-analysis could not be performed for one of the scoring systems due to insufficient literature. Continuous studies on the diagnostic accuracy of the CDT are needed for further analysis. Second, like other systematic reviews, this study may be susceptible to publication bias and may exaggerate the estimate of test accuracy. Third, heterogeneity exists in the study due to the characteristics of diagnostic tests. However, the CDT has a clear cut-off score for each system and the meta-analysis was performed on the studies with the same cut-off value so that the threshold effect would not be considered statistically significant. Additionally, because a bivariate model was used to consider the heterogeneity of studies, the confidence region was confirmed through a 95% CI and the HSROC curve.

We suggest that the studies on the diagnostic accuracy of CDT be carried out using various scoring systems. This would result in more precise estimates of the diagnostic accuracy of the CDT, including comparison analysis of scoring systems, in the future.

5 | CONCLUSIONS

This study analysed the diagnostic accuracy of each CDT scoring system through a systematic review and meta-analysis on 18 studies on diagnostic accuracy. The CDT is simple and easy for patients to perform, takes a short amount of time, has a simple assessment method, is solid in psychometric perspectives, and can test/examine the overall cognitive domains. It is a valuable test in primary medical centres, nursing homes, and public health services. It has been recognized for its accuracy as a screening test for dementia in the general older people population in several studies. Because nurses often conduct cognitive impairment tests to screen for dementia in the general older people population in nursing homes and public health services, they need to have knowledge of cognitive impairment test methods and their diagnostic accuracy. The CDT can be easily used by family as well as nurses because it is simple, easy, acceptable, and independent of education/language difference. It will be important to conduct research on the user and timing of the CDT in the future.

Through a systematic review and meta-analysis of the diagnostic accuracy of the CDT, the Shulman scoring system, which was the most studied and highly sensitive, was determined to be the most useful in the cognitive testing for dementia. After gaining a better understanding of how to use the CDT and its accuracy with different scoring systems through this study, we recommend it for widespread use in the diagnosis of dementia.

CONFLICTS OF INTEREST

No conflict of interest has been declared by the authors.

AUTHORS CONTRIBUTIONS

All authors have agreed on the final version and meet at least one of the following criteria [recommended by the ICMJE (http://www.icmje.org/recommendations/)]:

- substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data;
- drafting the article or revising it critically for important intellectual content.

ORCID

JinKyung Park https://orcid.org/0000-0001-8538-0476
EunHye Jeong http://orcid.org/0000-0002-2510-9112
GyeongAe Seomun http://orcid.org/0000-0002-1651-5741

REFERENCES

Amodeo, S., Mainland, B. J., Herrmann, N., & Shulman, K. I. (2015). The times they are a-changin': Clock drawing and prediction of dementia. *Journal of Geriatric Psychiatry and Neurology*, *28*(2), 145–155. https://doi.org/10.1177/0891988714554709

Aprahamian, I., Martinelli, J. E., Neri, A. L., & Yassuda, M. S. (2010). The accuracy of the clock drawing test compared to that of standard screening tests for Alzheimer's disease: Results from a study of Brazilian elderly with heterogeneous educational backgrounds. *International Psychogeriatrics*, *22*(1), 64–71. https://doi.org/10.1017/S1041610209991141

Berger, G., Frolich, L., Weber, B., & Pantel, J. (2008). Diagnostic accuracy of the clock drawing test: The relevance of "time setting'' in screening for dementia. *Journal of Geriatric Psychiatry and Neurology*, *21*(4), 250–260. https://doi.org/10.1177/0891988708324939

Cacho, J., Benito-Leon, J., Garcia-Garcia, R., Fernand ez-Calvo, B., Vicente-Villardon, J. L., & Mitchell, A. J. (2010). Does the combination of the MMSE and clock drawing test (mini-clock) improve the detection of mild Alzheimer's disease and mild cognitive impairment? *Journal of Alzheimer's Disease*, *22*(3), 889–896. https://doi.org/10.3233/JAD-2010-101182

Chen, C. Y., Leung, K. K., & Chen, C. Y. (2011). A quick dementia screening tool for primary care physicians. *Archives of Gerontology and Geriatrics*, *53*(1), 100–103. https://doi.org/10.1016/j.archger.2010.06.008

Chiu, Y. C., Li, C. L., Lin, K. N., Chiu, Y. F., & Liu, H. C. (2008). Sensitivity and specificity of the clock drawing test, incorporating Rouleau scoring system, as a screening instrument for questionable and mild dementia: Scale development. *International Journal of Nursing Studies*, *45*(1), 75–84. https://doi.org/10.1016/j.ijnurstu.2006.09.005

DESA, U.N.P.D. (2000). United Nations Population Division DESA: Report on world population ageing 1950–2050., pp. xxvii–xxxi. http://www.un.org/esa/population/publications/worldageing19502050/pdf/62executivesummary_english.pdf

Downing, L. J., Caprio, T. V., & Lyness, J. M. (2013). Geriatric psychiatry review: Differential diagnosis and treatment of the 3 D's-delirium, dementia and depression. *Current Psychiatry Reports*, *15*(6), 365. https://doi.org/10.1007/s11920-013-0365-4

Ehreke, L., Luppa, M., Konig, H. H., Villringer, A., & Riedel-Heller, S. G. (2011). Does the clock drawing test predict dementia? Results of the Leipzig longitudinal study of the aged (LEILA 75 +). *Dementia and Geriatric Cognitive Disorders*, *31*(2), 89–97. https://doi.org/10.1159/000323317

Espino, D. V., Lichtenstein, M. J., Palmer, R. F., & Hazuda, H. P. (2004). Evaluation of the mini-mental state examination's internal consistency in a community-based sample of Mexican-American and European-American elders: Results from the San Antonio Longitudinal Study of Aging. *Journal of the American Geriatrics Society*, *52*(5), 822–827. https://doi.org/10.1111/(ISSN)1532-5415

Fiest, K. M., Jette, N., Roberts, J. I., Maxwell, C. J., Smith, E. E., Black, S. E., … Hogan, D. B. (2016). The prevalence and incidence of dementia: A systematic review and meta-analysis. *Canadian Journal of Neurological Sciences*, *43*(Suppl 1), S3–S50. https://doi.org/10.1017/cjn.2016.18

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Gruber, N. P., Varner, R. V., Chen, Y. W., & Lesser, J. M. (1997). A comparison of the clock drawing test and the Pfeiffer Short Portable Mental Status Questionnaire in a geropsychiatry clinic. *International Journal of Geriatric Psychiatry*, *12*(5), 526–532. https://doi.org/10.1002/(ISSN)1099-1166

Henderson, M., Scott, S., & Hotopf, M. (2007). Use of the clock-drawing test in a hospice population. *Palliative Medicine*, *21*(7), 559–565. https://doi.org/10.1177/0269216307081936

Herrmann, N., Kidron, D., Shulman, K. I., Kaplan, E., Binns, M., Leach, L., & Freedman, M. (1998). Clock tests in depression, Alzheimer's disease and elderly controls. *International Journal of Psychiatry in Medicine*, *28*(4), 437–447. https://doi.org/10.2190/5QA5-PHUN-1Q9F-C0PB

Hesson, A. M., & Pichler, H. (2016). Interpreting "I don't know" use by persons living with dementia in Mini-Mental State Examinations. *Patient Education and Counseling*, *99*(9), 1534–1541. https://doi.org/10.1016/j.pec.2016.07.001

Ismail, Z., Rajji, T. K., & Shulman, K. I. (2010). Brief cognitive screening instruments: An update. *International Journal of Geriatric Psychiatry*, *25*(2), 111–120. https://doi.org/10.1002/gps.2306

Johnson, J., Sims, R., & Gottlieb, G. (1994). Differential diagnosis of dementia, delirium and depression. Implications for drug therapy. *Drugs and Aging*, *5*(6), 431–445. https://doi.org/10.2165/00002512-199405060-00005

Jouk, A., & Tuokko, H. (2012). A reduced scoring system for the clock drawing test using a population-based sample. *International Psychogeriatrics*, *24*(11), 1738–1748. https://doi.org/10.1017/S1041610212000804

Juby, A., Tench, S., & Baker, V. (2002). The value of clock drawing in identifying executive cognitive dysfunction in people with a normal Mini-Mental State Examination score. *CMAJ*, *167*(8), 859–864.

Kim, H., & Chey, J. (2010). Effects of education, literacy and dementia on the clock drawing test performance. *Journal of the International Neuropsychological Society*, *16*(6), 1138–1146. https://doi.org/10.1017/S1355617710000731

Kim, K. W., Lee, J., Choi, S. H., Huh, J., & Park, S. H. (2015). Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: A practical review for clinical researchers-part I. general guidance and tips. *Korean J Radiol*, *16*(6), 1175–1187. https://doi.org/10.3348/kjr.2015.16.6.1175

Kirby, M., Denihan, A., Bruce, I., Coakley, D., & Lawlor, B. A. (2001). The clock drawing test in primary care: Sensitivity in dementia detection and specificity against normal and depressed elderly. *International Journal of Geriatric Psychiatry*, *16*(10), 935–940. https://doi.org/10.1002/(ISSN)1099-1166

Kliegel, M., Zimprich, D., & Rott, C. (2004). Life-long intellectual activities mediate the predictive effect of early education on cognitive impairment in centenarians: A retrospective study. *Aging & Mental Health*, *8*(5), 430–437. https://doi.org/10.1080/13607860410001725072

Kronhaus, A., Fuller, S., Zimmerman, S., & Reed, D. (2016). Prevalence and medication management of dementia by a medical practice providing onsite care in assisted living. *J Am Med Dir Assoc*, *17*(7), 673.e9–673.e15. https://doi.org/10.1016/j.jamda.2016.04.028

Laakkonen, M. L., Kautiainen, H., Holtta, E., Savikko, N., Tilvis, R. S., Strand Berg, T. E., & Pitkala, K. H. (2016). Effects of self-management

groups for people with dementia and their spouses-randomized controlled trial. *Journal of the American Geriatrics Society*, *64*(4), 752–760. https://doi.org/10.1111/jgs.14055

Lee, J., Kim, K. W., Choi, S. H., Huh, J., & Park, S. H. (2015). Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: A practical review for clinical researchers-part II. Statistical methods of meta-analysis. *Korean J Radiol*, *16*(6), 1188–1196. https://doi.org/10.3348/kjr.2015.16.6.1188

Lee, K. S., Kim, E. A., Hong, C. H., Lee, D. W., Oh, B. H., & Cheong, H. K. (2008). Clock drawing test in mild cognitive impairment: Quantitative analysis of four scoring methods and qualitative analysis. *Dementia and Geriatric Cognitive Disorders*, *26*(6), 483–489. https://doi.org/10.1159/000167879

Lee, H., Swanwick, G. R., Coen, R. F., & Lawlor, B. A. (1996). Use of the clock drawing task in the diagnosis of mild and very mild Alzheimer's disease. *International Psychogeriatrics*, *8*(3), 469–476. https://doi.org/10.1017/S1041610296002827

Levey, A., Lah, J., Goldstein, F., Steenland, K., & Bliwise, D. (2006). Mild cognitive impairment: An opportunity to identify patients at high risk for progression to Alzheimer's disease. *Clinical Therapeutics*, *28*(7), 991–1001. https://doi.org/10.1016/j.clinthera.2006.07.006

Liberati, A., Altmanx, A. D., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*(7), e1000100.

Lobo, A., Launer, L. J., Fratiglioni, L., Andersen, K., Di Carlo, A., Breteler, M. M., … Hofman, A. (2000). Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic diseases in the elderly research group. *Neurology*, *54*(11 Suppl 5), S4–S9.

Mainland, B. J., Amodeo, S., & Shulman, K. I. (2014). Multiple clock drawing scoring systems: Simpler is better. *International Journal of Geriatric Psychiatry*, *29*(2), 127–136. https://doi.org/10.1002/gps.3992

Manos, P. J., & Wu, R., (1994). The ten point clock test: a quick screen and grading method for cognitive impairment in medical and surgical patients. *International Journal of Psychiatry in Medicine*, *24*(3), 229–244.

Matsuoka, T., Narumoto, J., Shibata, K., Okamura, A., Nakamura, K., Nakamae, T., … Fukui, K. (2011). Neural correlates of performance on the different scoring systems of the clock drawing test. *Neuroscience Letters*, *487*(3), 421–425. https://doi.org/10.1016/j.neulet.2010.10.069

Milian, M., Leiherr, A. M., Straten, G., Muller, S., Leyhe, T., & Eschweiler, G. W. (2012). The mini-cog versus the mini-mental state examination and the clock drawing test in daily clinical practice: Screening value in a German memory clinic. *International Psychogeriatrics*, *24*(5), 766–774. https://doi.org/10.1017/S1041610211002286

Molnar, F. J., Patel, A., Marshall, S. C., Man-Son-Hing, M., & Wilson, K. G. (2006). Clinical utility of office-based cognitive predictors of fitness to drive in persons with dementia: A systematic review. *Journal of the American Geriatrics Society*, *54*(12), 1809–1824. https://doi.org/10.1111/j.1532-5415.2006.00967.x

Nishiwaki, Y., Breeze, E., Smeeth, L., Bulpitt, C. J., Peters, R., & Fletcher, A. E. (2004). Validity of the Clock-Drawing Test as a screening tool for cognitive impairment in the elderly. *American Journal of Epidemiology*, *160*(8), 797–807. https://doi.org/10.1093/aje/kwh288

Nunes, P. V., Diniz, B. S., Radanovic, M., Abreu, I. D., Borelli, D. T., Yassuda, M. S., & Forlenza, O. V. (2008). CAMcog as a screening tool for diagnosis of mild cognitive impairment and dementia in a Brazilian clinical sample of moderate to high education. *International Journal of Geriatric Psychiatry*, *23*(11), 1127–1133. https://doi.org/10.1002/gps.2038

Powell, S., Towers, A., & Milne, P. (2008). The public health view on closing the gap between public health and primary care. *Family Practice*, *25*(Suppl 1), i17–i19.

Ramlall, S., Chipps, J., Bhigjee, A. I., & Pillay, B. J. (2013). The sensitivity and specificity of subjective memory complaints and the subjective memory rating scale, deterioration cognitive observee, mini-mental state examination, six-item screener and clock drawing test in dementia screening. *Dementia and Geriatric Cognitive Disorders*, *36*(1–2), 119–135. https://doi.org/10.1159/000350768

Rector, T. S., Taylor, B. C., & Wilt, T. J. (2012). AHRQ Methods for Effective Health Care Systematic Review of Prognostic Tests. In S. M. Chang, D. B. Matchar, G. W. Smetana, & C. A. Umscheid (Eds.), *Methods guide for medical test reviews*. Rockville (MD): Agency for Healthcare Research and Quality (US).

Ricci, M., Pigliautile, M., D'Ambrosio, V., Ercolani, S., Bianchini, C., Ruggiero, C., … Mecocci, P. (2016). The clock drawing test as a screening tool in mild cognitive impairment and very mild dementia: A new brief method of scoring and normative data in the elderly. *Neurol Sci*, *37*(6), 867–873. https://doi.org/10.1007/s10072-016-2480-6

Riedel, O., Klotsche, J., Forstl, H., & Wittchen, H. U. (2013). Clock drawing test: Is it useful for dementia screening in patients having Parkinson disease with and without depression? *Journal of Geriatric Psychiatry and Neurology*, *26*(3), 151–157. https://doi.org/10.1177/0891988713490994

Royall, D. R. (1996). Comments on the executive control of clock-drawing. *Journal of the American Geriatrics Society*, *44*(2), 218–219. https://doi.org/10.1111/j.1532-5415.1996.tb02452.x

Russo, M. J., Iturry, M., Sraka, M. A., Bartoloni, L., Carnero Pardo, C., & Allegri, R. F. (2014). Diagnostic accuracy of the Phototest for cognitive impairment and dementia in Argentina. *Clin Neuropsychol*, *28*(5), 826–840. https://doi.org/10.1080/13854046.2014.928748

Shulman, K. I. (2000). Clock-drawing: Is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry*, *15*(6), 548–561. https://doi.org/10.1002/(ISSN)1099-1166

Shulman, K. I., Herrmann, N., Brodaty, H., Chiu, H., Lawlor, B., Ritchie, K., & Scanlan, J. M. (2006). IPA survey of brief cognitive screening instruments. *International Psychogeriatrics*, *18*(2), 281–294. https://doi.org/10.1017/S1041610205002693

Souillard-Mandar, W., Davis, R., Rudin, C., Au, R., Libon, D. J., Swenson, R., … Penney, D. L. (2016). Learning classification models of cognitive conditions from subtle behaviors in the digital clock drawing test. *Machine Learning*, *102*(3), 393–441. https://doi.org/10.1007/s10994-015-5529-5

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests. Administration, norms and commentary*. Oxford, UK: Oxford University Press 2 ed.

Sugawara, N., Yasui-Furukori, N., Umeda, T., Sato, Y., Kaneda, A., Tsuchimine, S., … Kaneko, S. (2010). Clock drawing performance in a community-dwelling population: Normative data for Japanese subjects. *Aging & Mental Health*, *14*(5), 587–592. https://doi.org/10.1080/13607860903586086

Tuokko, H., Hadjistavropoulos, T., Rae, S., & O'Rourke, N. (2000). A comparison of alternative approaches to the scoring of clock drawing. *Archives of Clinical Neuropsychology*, *15*(2), 137–148.

Wolf-Klein, G. P., Silverstone, F. A., Levy, A. P., & Brod, M. S. (1989). Screening for Alzheimer's disease by clock drawing. *Journal of the American Geriatrics Society*, *37*(8), 730–734. https://doi.org/10.1111/j.1532-5415.1989.tb02234.x

Yamamoto, S., Mogi, N., Umegaki, H., Suzuki, Y., Ando, F., Shimokata, H., & Iguchi, A. (2004). The clock drawing test as a valid screening method for mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*, *18*(2), 172–179. https://doi.org/10.1159/000079198

Zhou, A., & Jia, J. (2008). The value of the clock drawing test and the mini-mental state examination for identifying vascular cognitive impairment no dementia. *International Journal of Geriatric Psychiatry*, *23*(4), 422–426. https://doi.org/10.1002/(ISSN)1099-1166

How to cite this article: Park JK, Jeong EH, Seomun GA. The clock drawing test: A systematic review and meta-analysis of diagnostic accuracy. *J Adv Nurs*. 2018;74:2742–2754. https://doi.org/10.1111/jan.13810

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.