Guest Editorial

The clock-drawing test: a critique

Let me begin with a declaration of disinterest. The esteemed editor of International Psychogeriatrics commissioned this article after reading my letter to Another Journal (Philpot, 2003) in which I called for a moratorium on all clock-drawing test (CDT) research. It seemed to me that an awful lot of effort was being put into a fairly trivial issue and that researchers’ time might be better spent. That much effort is being expended cannot be denied. Using the search words ‘clock drawing’, MEDLINE currently lists 177 publications, 91 since the turn of the century. The first reference to this specific form of drawing test comes in 1983 (Goodglass and Kaplan), and the first paper on validation is that of Shulman et al. (1986).

The CDT has emerged as the front runner from a range of drawing tests that have been used to investigate cognition over the last 40 years. When I trained at the Maudsley Hospital in the early 1980s we used the ‘draw-a-cat’ test and used what is now known as a ‘naïve’, informal scoring method to judge ‘normality’ or ‘abnormality’. Cube-drawing used to be popular but is strongly influenced by educational and cultural background (Rosselli and Ardila, 2003). The problem of three-dimensionality and perspective also applies to the house-drawing test (Moore and Wyke, 1984), but is the specific focus of the ‘draw-a-tree-lined-avenue’ test (Rennert, 1971). The House-Person-Tree test can be used to assess intellectual function, as well as having a projective element (Fukunishi et al., 1997).

The properties of any screening test should not only include ease of administration and acceptability but ease of scoring and, if used sensibly as a first stage in disease identification, there should be a high sensitivity and positive predictive value. Specificity and negative predictive value are less important in this respect, unless the screening method is itself lengthy, aversive or expensive, none of which apply to any of the forms of the CDT. It seems to this reviewer that, while the CDT is often proposed as a means of identifying patients with early stages of dementia who might benefit from specific treatment, it has largely been validated in groups of patients with established diagnoses. Or worse still, validated in patients whose ‘dementia’ is only established by scores on other screening tests.

Ken Shulman (2000) has written an excellent and balanced review of the CDT in which he emphasized that it should be used in conjunction with other broader tests of intellectual function, such as the Mini-mental State Examination (MMSE), the verbal fluency test and informant reports. My quibble here would
be that, if the test can’t be used in isolation, then it is not really a screening test in its own right and is merely part of a battery, as can be said of the digit span test for example.

What have we learnt since Shulman’s review? The following represents a selective review of recent publications.

**Drawing and scoring methods**

Shulman (2000) emphasized the virtue of simple scoring methods. He identified twelve different procedures for eliciting and scoring clock drawings that had been published up to 1998. These differ a good deal in complexity. At least three new methods have appeared. Manos (1999) has a 10-point scale, Royall and co-workers (1999) have developed a version (the CLOX) that examines more rigorously the executive function components and Lin and colleagues (2003) employ drawing and copying tasks with both complex and simple scoring methods. Many papers have compared the validity and reliability of scoring methods. Almost without exception, the sensitivity and specificity values are derived from comparisons of patients with well-established Alzheimer’s disease (AD) and healthy controls matched for age and education. However, several studies have demonstrated that the CDT, however scored, is only modestly successful in identifying cases of mild dementia (Storey et al., 2001; Powlishta et al., 2002; Seigerschmidt et al., 2002). Similarly, some doubts have been expressed about its usefulness in the community (Kirby et al., 2001).

Why are there so many different scoring systems? Perhaps it is now the time to concentrate on two or three. The Mendez system (Mendez et al., 1992) was identified as the most accurate discriminator by Storey et al. (2001) and Scanlan et al. (2002) among others. ‘Naïve’ judgment (abnormal or not) has been shown to be quite adequate for quick clinical purposes (Scanlan et al., 2002). The simple method of Shulman and colleagues (1986) provides a happy medium.

Several authors have explored the added value, in terms of diagnostic discriminating power, that might be achieved by combining the CDT with other tasks. The Mini-Cog is a composite of a three-item recall test and the CDT (Borson et al., 2000), and is reported to have high reliability and validity in identifying cases of dementia (Scanlan and Borson, 2001). Other combinations have been suggested with the Syndrom Kurz Test (Schramm et al., 2002), the Short Cognitive Evaluation Battery (Robert et al., 2003) and the MMSE (Heinik et al., 2003). The CDT was originally part of the Boston Aphasia battery (Goodglass and Kaplan, 1983). There is a certain irony in advocating its return to use in combination with other tests, after all that has been written about it in the last twenty years.
Transcultural acceptability

Advocates of the CDT often cite the ease with which the test can be administered to patients from different cultures and languages. Successful validation studies have been carried out in Italy (Ravaglia et al., 2003), non-English US citizens (Borson et al., 1999), Thailand (Jitapunkul et al., 2000), Brazil (Fuzikawa et al., 2003), Mexico (Royall et al., 2003) and Hong Kong (Lam et al., 1998). However, Storey et al. (2002) found only modest sensitivity and low specificity in a culturally diverse group in Australia.

Discrimination between AD and other dementias

Sensitivity and specificity values are more modest in the studies comparing groups of patients with different dementias. These include Cahn-Weiner et al. (2003), comparing AD with Parkinson’s disease dementia, and dementia with Lewy bodies, and Heinik et al. (2002) comparing AD with vascular dementia. Rascovsky et al. (2002) and Moretti et al. (2002) both found that patients with fronto-temporal dementia (FTD) performed better on the CDT than AD patients matched for severity, whereas the FTD patients were worse at verbal fluency tests. Here my concern is whether the CDT score adds anything to the usual diagnostic process, clinical history and supporting investigations. What are presented in these studies are comparisons of group means rather than the predictive value of a particular score.

Neurological localization

Some researchers have argued, largely based on correlational data with other test methods, that the CDT is chiefly a test of executive function (Royall et al., 1999). However, brain imaging studies do not support this view. Using a variety of techniques CDT scores have been associated with ventricular enlargement (Heinik et al., 2000), grey matter volume in the non-dominant temporal lobe (Cahn-Wiener et al., 1999) and increased cerebral blood flow in the dominant posterior temporal regions (Ueda et al., 2002). Ino and colleagues (2003) found that actual performance of the CDT was associated with increased activity on functional magnetic resonance imaging in the posterior parietal cortices.

Calling time

While many authors do sensibly couch their enthusiasm for the CDT in qualifying statements about its limitations and its proper use in conjunction with other tests, many do not. Clearly it is essential for us to have adequate screening methods for dementia that are quick and easy to administer, and that
examine the range of disabilities with which dementia might present. My original call for a moratorium was motivated in part by an altruistic urge to encourage colleagues to focus their minds on something more robust than the CDT (or ‘get a life’ as modern parlance has it). With the exception of the interesting but inconsistent brain imaging studies, I am not impressed that we have learnt much more since Shulman’s review (2000). Perhaps now is the moment to ‘Stop all the clocks, cut off the mobile phone . . . ’ (to misquote W. H. Auden . . . again).

MICHAEL PHILPOT
Consultant Psychiatrist
Maudsley Hospital
London
U.K.
Email: mike.philpot@slam.nhs.uk

References


Critique of the clock-drawing test


