Title

Comparison of cognitive performance measures in individuals with systemic lupus erythematosus.

Permalink

Journal

ISSN

Authors

Bowling, C
Dunlop-Thomas, Charmayne
Hoge, Courtney
et al.

Publication Date

DOI

Comparison of cognitive performance measures in individuals with systemic lupus erythematosus

Laura Plantinga ![ORCID],[1] Jinoos Yazdany ![ORCID],[1] C Barrett Bowling,[2,3] Charmayne Dunlop-Thomas,[4] Courtney Hoge,[4] Brad D Pearce,[5] S Sam Lim ![ORCID] ,[4,5] Patricia Katz[1]

[1]Department of Medicine, University of California San Francisco, San Francisco, California, USA
[2]Durham Veterans Affairs Medical Center, Durham, North Carolina, USA
[3]Department of Medicine, Duke University, Durham, North Carolina, USA
[4]Department of Medicine, Emory University, Atlanta, Georgia, USA
[5]Department of Epidemiology, Emory University, Atlanta, Georgia, USA

Correspondence to
Dr Laura Plantinga; laura. plantinga@ucsf.edu

ABSTRACT

Objective Cognitive impairment is a common complaint in SLE, but approaches to measuring cognitive performance objectively vary. Leveraging data collected in a population-based cohort of individuals with validated SLE, we compared performance and potential impairment across multiple measures of cognition.

Methods During a single study visit (October 2019–May 2022), times to complete the Trail Making Test B (TMTB; N=423) were recorded; potential impairment was defined as an age-corrected and education-corrected T-score <35 (>1.5 SD longer than the normative time). A clock drawing assessment (CLOX; N=435) with two parts (free clock draw (CLOX1) and copy (CLOX2)) was also performed (score range: 0–15; higher scores=better performance); potential impairment was defined as CLOX1 <10 or CLOX2 <12. Fluid cognition (N=199; in-person visits only) was measured via the National Institutes of Health (NIH) Toolbox Fluid Cognition Battery and expressed as age-corrected standard scores; potential impairment was defined by a score <77.5 (>1.5 SD lower the normative score).

Results Participants (mean age 46 years; 92% female; 82% black) had a median (IQR) TMTB time of 96 (76–130) s; median (IQR) CLOX1 and CLOX2 scores of 12 (10–13) and 14 (13–15); and a mean (SD) fluid cognition standard score of 87.2 (15.6). TMTB time and fluid cognition score (ρ=−0.53, p<0.001) were the most highly intercorrelated measures. Overall, 65%, 55% and 28% were potentially impaired by the TMTB test, CLOX task and NIH Toolbox Fluid Cognition Battery, respectively. While there was overlap in potential impairment between TMTB and CLOX, more than half (58%) had impairment by only one of these assessments. Few (2%) had impairment in fluid cognition only.

Conclusion The TMTB, CLOX and NIH Fluid Cognition Battery each provided unique and potentially important information about cognitive performance in our SLE cohort. Future studies are needed to validate these measures in SLE and explore interventions that maintain or improve cognitive performance in this population.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Individuals with SLE often complain of cognitive issues, but cognition is rarely measured clinically due to the time burden of many assessments.

WHAT THIS STUDY ADDS

⇒ In this study, we compared performance among individuals with SLE on a longer, multidomain assessment of fluid cognition (National Institutes of Health Toolbox Fluid Cognition Battery) to their performance on two shorter assessments (Trail Making Test B and a clock drawing task) frequently used in geriatric settings to detect cognitive impairment or dementia. We found that the two rapid measurements together identify most with impairment by the longer measurement, but the different measurements may provide important and unique information to guide treatment and support.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study provides more information on shorter assessments that could potentially be used as initial screening tests in clinical settings for patients who complain of cognitive issues, with further neuropsychiatric testing and support considered for those patients with poor performance. Additionally, our study suggests further research to validate these measures in SLE and explore interventions that maintain or improve cognitive performance in this population.

INTRODUCTION

Cognitive dysfunction is a common complaint among individuals with SLE.[1–4] One reason why it is not often addressed or documented clinically[5] may be that cognitive testing is perceived to be burdensome in the clinical setting, where there are frequently competing medical and social issues. Rapid tools to screen for cognitive impairment in the clinical setting may be more acceptable to SLE providers and patients and provide important snapshots of current cognitive functioning that could help with shared decision-making and goal setting. However, these screening measures are generally designed to be most sensitive to severe cognitive impairment or dementia.[6] Furthermore, screening measures

often assess performance in single domains; previous studies have shown that dysfunction is widespread and heterogeneous across multiple cognitive function domains among individuals with SLE, with various levels of impairment in, for example, working and episodic memory, attention, processing speed, executive function and verbal fluency, among other domains.[7 8]

Assessments to detect more subtle deficits in cognitive performance can take considerably more time. For example, the American College of Rheumatology Neuropsychological Battery (ACR-NB), while time-efficient relative to the number of domains assessed, takes 1 hour to administer and is not consistently used, even in the research setting.[9 10] However, it is important to consider that the milder impairments detectable in these longer assessments may have a substantial impact on daily life, including work, school, relationships, social activity, self-management of SLE, and general physical and mental health.[11 12] Additionally, identification of milder impairments offers the opportunity to develop strategies to circumvent impairments and provide support to help prevent further decline.

As part of a recent ancillary study, the Approaches to Positive, Patient-centered Experiences of Aging in Lupus (APPEAL), we administered multiple cognitive performance assessments in a population-based, primarily black cohort of individuals with SLE. Here, we leveraged these data to compare the potential value of brief screening tests and longer, multidomain assessments in SLE care and research. Specifically, we sought to estimate and compare performance and potential impairment across two screening tests (the Trail Making Test B (TMTB) and CLOX, a clock drawing task[13]) and the multidomain National Institutes of Health (NIH) Fluid Cognition Battery.[14–16]

METHODS
Study population and data sources
We recruited participants for a one-time study visit from the population-based Georgians Organized Against Lupus (GOAL) cohort of adults (≥18 years) with SLE (defined as ≥4 revised American College of Rheumatology (ACR) criteria,[17] or 3 ACR criteria plus a diagnosis of SLE by a board-certified rheumatologist) in metropolitan Atlanta.[18 19] For APPEAL, we excluded individuals who were not actively participating in GOAL, were unable to speak English, did not have sufficient vision and hearing to undergo study testing, were unable to consent or were living outside of Georgia at the time of recruitment.

Overall, N=451 (90.0%) completed a study visit between October 2019 and May 2022. Study visits were conducted either in person at an Emory study site (n=206) or remotely via Health Insurance Portability and Accountability Act-compliant Zoom software (San Jose, California, USA) (n=245). For these analyses, data for participants who had potentially invalid data (n=11), technical issues that resulted in a loss of cognitive data (n=2) or incomplete
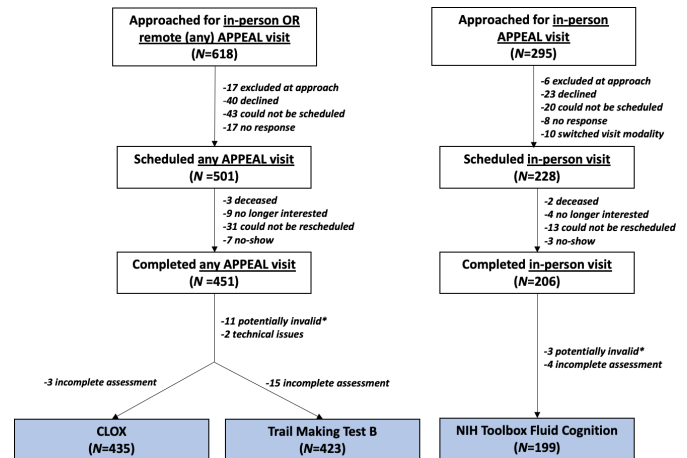

Figure 1 Flow of participants for Trail Making Test B and CLOX assessments (all APPEAL participants) and for NIH Toolbox Fluid Cognition Battery assessments (participants completing in-person visits only). *Potentially invalid=participants who likely had assistance during cognitive testing or who had survey response patterns suggesting invalid responses. APPEAL, Approaches to Positive, Patient-centered Experiences of Aging in Lupus; CLOX, clock drawing assessment; NIH, National Institutes of Health.

assessments were further excluded, leaving N=435, 425 and 199 total visits for analysis of CLOX, TMTB and fluid cognition (in-person visits only), respectively (figure 1). Data on cognitive performance were either derived from NIH Toolbox[14–16] or entered into REDCap[20]; manual data entry was checked by an independent reviewer and errors were corrected. Self-reported data on functioning and other domains were obtained from self-administered surveys via REDCap during APPEAL visits[20] or linked from the closest GOAL survey.

Patient and public involvement
Patients were involved in the design and conduct of this research. We used feedback from our pilot study of patients[21] to create our initial protocol; patient participant feedback was also used to modify the protocol as needed throughout the course of this study. Patient burden was carefully considered in the number and order of measures assessed; patients were able to skip assessments and take breaks as needed. GOAL participants are informed of study results through regular study newsletters suitable for a non-specialist audience.

Variables
Cognitive performance measures
The NIH Fluid Cognition Battery[14–16] was chosen as the primary measurement of cognitive function in APPEAL, due to its use in our pilot and the ability to use normalised scores to compare across studies and populations. The TMTB and CLOX[13] were added to APPEAL because they represented short assessments that are frequently used in geriatric clinics to rapidly assess cognition. These

available measures were then compared in the study reported here.

Trail Making Test B

The TMTB is an often-used tool to screen for impairment of executive function, specifically visual attention and task switching or cognitive flexibility.[22] In this test, the participant is asked to connect numbers and letters in alternating numerical and alphabetical order (total time: <5 min). The participant was instructed to complete the task as quickly as possible without errors, and interviewers told participants to correct errors as they performed the task. For remote visits, participants were shown a brief video demonstrating the task (using the same example used during in-person visits) and results were screen captured (see online supplemental methods). The time to complete the test was recorded. Errors were not penalised, beyond the time taken to correct errors. However, if the participant was still working on the test at 5 min, they were asked to stop. Potential impairment was primarily defined as a time that was >1.5 SD greater than the normative value for the participant's age group and education.[23]

Clock drawing

The CLOX instrument[13] (total time: ~5 min) is a version of clock drawing tests, which measure various aspects of executive functioning and are often used as a screening for dementia[24] and potential driving issues[25] in older adults. Participants were first asked to draw a clock showing 1:45, without further instructions (CLOX1). The interviewer then either drew a correct clock in real time (in-person visits) or via a recorded video played for the participant (remote visits; see online supplemental methods), which the participant was then asked to copy (CLOX2). For both in-person and remote visits, both clocks were scored on various aspects such as size, numbers/order, correct hand size/position, etc (range, 0–15, with lower scores indicating more impairment). Clocks were scored by the interviewer and rescored by another researcher not involved in study visits; differences were resolved between the research manager (CH) and principal investigator (LP).[13] Potential impairment was defined as a CLOX1 score ≤10 (possibly indicating the presence of executive function impairment) or a CLOX2 score ≤12 (possibly indicating both executive function and posterior cortical impairment).[26]

Fluid cognition

Fluid cognition is the ability to think and reason abstractly and solve problems, independent of learning, experience and education. Five individual assessments comprised the composite measure of fluid cognition (episodic memory, working memory, attention and inhibitory control, processing speed and cognitive flexibility; total time: ~20–30 min) and were administered in person via the NIH Toolbox application.[14–16] An age-corrected standard score is provided for each assessment and the composite measure. A score of 100 on these standard scores represents the average performance for the test taker's age in the normative sample (SD=15; higher scores indicate better cognitive functioning). Potential impairment was primarily defined as a score >1.5 SD below the mean (score=77.5). Because not all assessments in the NIH Fluid Cognition Battery could be administered remotely, we have complete data on overall fluid cognition on 199 participants only.

Other variables

Sociodemographics included age, sex, race, ethnicity, education and employment. Age was self-reported on the NIH Toolbox and examined both as a continuous variable and in categories (18–34, 35–49 and ≥50 years). Sex (at birth), race, ethnicity and education were self-reported (from a fixed set of categories) by the participant via the NIH Toolbox. Race was categorised as black (as a single or multiple race), white and other. Education was the highest level attained and categorised as high school graduate/equivalency or lower, some college/associates degree, and college graduate or higher. Current employment was assessed using the Work Productivity and Activity Impairment Questionnaire: General Health V.2.0.[26–29] Clinical variables included disease duration, which was self-reported during the nearest GOAL assessment and adjusted for the date of the APPEAL assessment. Current SLE activity was assessed via the Systemic Lupus Activity Questionnaire (SLAQ) (range 0–44; higher scores indicating greater SLE-related disease activity).[30] Moderate to severe (vs none or mild) self-reported forgetfulness was taken from a single item on the SLAQ.[30] The Brief Index of Lupus Damage (BILD) score (range, 0–46; higher scores indicate greater cumulative SLE-related organ damage)[31 32] closest to the APPEAL visit was obtained from linked GOAL data; dichotomous measures of neuropsychiatric damage present versus absent by system were also obtained from BILD items. Height and weight were measured (in-person visits) or self-reported (remote visits); body mass index (BMI) was calculated as (weight in kg)/(height in m).[2] Current steroid use was self-reported by the participant at the study visit. Depressive symptoms were assessed via the nine-item Patient Reported Outcomes Measurement Information System (PROMIS) Depression Short Form-8a, which has been validated in diverse populations[33] and in other rheumatological conditions[34]; raw scores were converted to T-scores (where 50=mean score and 10=1 SD). Participants' perceived stress was assessed using the 10-item Perceived Stress Scale,[35 36] which measured the degree to which participants found life situations stressful over the past month (range 0–40; higher scores indicate greater perceived stress, with scores of ≥20 considered high levels of stress). Finally, we were interested in physical activity and performance. Physical activity was assessed with the International Physical Activity Questionnaire–Short Form.[37] Scaled T-scores from the PROMIS Physical Functioning-Short Form-12a[38] were used to assess participants' perceptions of their physical functioning (with

higher scores representing better self-reported physical functioning). We assessed physical performance (in both in-person and remote visits)[39 40] using the Short Physical Performance Battery (SPPB; range 0–12, with higher scores representing better performance).[41]

Statistical analysis

Characteristics of participants were described within subpopulations for each cognitive performance measure. Cognitive performance scores were summarised and compared via distributions, histograms, scatter plots and correlation coefficients. Potential impairment was compared across measurements using percentage agreement, kappa values, Venn diagrams and UpSet plots. In secondary analyses, associations between characteristics and potential cognitive functioning impairment were also assessed via logistic regression (adjusting for visit type as appropriate) and estimates were compared across various measurements and definitions. In sensitivity analyses, we examined secondary definitions of potential impairment in TMTB (crude cut-off of >273 s to complete the task[42]), CLOX (CLOX1 and CLOX2 separately) and fluid cognition (using percentiles derived from age-adjusted, sex-adjusted, race-adjusted, ethnicity-adjusted and education-adjusted T-scores (≥4, ≥3 or ≥2 scores below the 25th, 16th and 9th percentiles, respectively)[43] and also examining episodic and working memory, the only individual NIH Toolbox Fluid Cognition assessments administered in all visits[44] (see online supplemental methods), separately). Complete case analysis was used. The statistical significance threshold was set at 0.05. All analyses were conducted using Stata V.18.0 (College Station, Texas, USA).

RESULTS
Characteristics of study participants

Participants had a mean age of 46 years, with most (79%) being 35 years or older, female (92%), black (82%) and non-Hispanic (94%); most had at least some college education (77%) and 48% were working at the time of the visit (table 1). The median duration of SLE was nearly 15 years and median SLAQ and BILD scores were 11 and 2, respectively. About one-quarter (27%) reported moderate to severe symptoms of forgetfulness. Participants' mean BMI was around 30; 74% reported low physical activity, the mean T-score for self-reported physical functioning was 44 and the mean SPPB score was 9. Those who completed NIH Toolbox Fluid Cognition Battery assessments (in-person visits only) were similar to those who completed the other assessments (all visits; table 1).

Cognitive performance among individuals with SLE
Summary of cognitive performance

The median TMTB time was slightly right-skewed (online supplemental figure 1A); the mean and median times to complete the task (including n=8 with a truncated time of 300 s) were 110 and 96 s (table 2). CLOX1 (free draw) and CLOX2 (copy task) were also skewed to the right (online

supplemental figure 1B,C); the median scores were 12 and 14 out of a possible 15, respectively (table 2). TMTB times and CLOX scores were similar within the subset with and without fluid cognition assessments (table 2). Age-corrected standard scores for fluid cognition were normally distributed (online supplemental figure 1D), and the mean score was 87 (table 2), nearly 1 SD below the population mean. Episodic and working memory assessments were also normally distributed (online supplemental figure 1E,F) and had similar mean scores across visit type (table 2).

Comparisons of cognitive performance across measurements

The TMTB time was weakly negatively correlated with CLOX1 and CLOX2 scores (ρ=−0.20 and −0.33) and moderately negatively correlated with age-corrected standard scores for fluid cognition (ρ=−0.53; online supplemental figure 2) and, specifically, cognitive flexibility scores (ρ=−0.44, p<0.001). Scores for CLOX1 and CLOX2 were moderately positively correlated with each other (ρ=0.45) and weakly positively correlated with fluid cognition scores (ρ=0.16 and 0.21; online supplemental figure 2).

Potential cognitive impairment among individuals with SLE
Summary of potential cognitive impairment

The percentage of individuals with potential impairment in TMTB, CLOX and fluid cognition by our primary definitions was 65%, 55% and 28%, respectively (figure 2). For TMTB, only 3% had impairment by the crude cut-off of 273 s (online supplemental figure 3). For CLOX, potential impairment in CLOX1 (score ≤10) was common (54%) but potential impairment in CLOX2 (score ≤12) was uncommon (4%) (online supplemental figure 3). Finally, impairment in fluid cognition was the same (28%) when defined by ≥2 NIH Toolbox Fluid Cognition individual domain scores <9th percentile, ≥3 domain scores <16th percentile or ≥4 domain scores <25th percentile (based on fully adjusted T-scores). Potential impairments in episodic and working memory (age-corrected standard scores <77.5) were seen in 6% and 18%, respectively (online supplemental figure 3).

Overlap of potential cognitive impairment

Percentage agreement between potential impairment across measurements by our primary definitions ranged from 52% to 61% (figure 2 and online supplemental table 1). Values for percentage agreement between measures and definitions for potential impairment were high (as high as 94% for CLOX2 vs TMTB time <273 s), but between-measurement kappa values were low (most ≤0.1; online supplemental table 1).

Figure 3A shows that about 42% of the population with potential TMTB or CLOX impairment had both, and most (93%) of those with potential fluid cognition impairment have either TMTB or CLOX impairment. Figure 3B displays the same data as percentages of individuals with various patterns: for example, 20% had

Table 1 Selected characteristics of included study participants with SLE

Characteristic	Participants with complete assessments for:		
	CLOX (N=435)	Trail Making Test B (N=423)	NIH Toolbox Fluid Cognition (N=199)
Sociodemographics			
Mean (SD) age, years	46.1 (11.9)	45.8 (11.4)	46.1 (12.4)
Age category, n (%)			
18–34	90 (20.7)	90 (21.3)	44 (22.1)
35–49	167 (38.4)	165 (39.0)	74 (37.2)
≥50	178 (40.9)	168 (39.7)	81 (40.7)
Sex,* n (%)			
Female	398 (91.5)	387 (91.5)	174 (87.4)
Male	37 (8.5)	36 (8.5)	25 (12.6)
Race, n (%)			
Black	358 (82.3)	346 (81.8)	172 (86.4)
Other	27 (6.2)	27 (6.4)	12 (6.0)
White	50 (11.5)	50 (11.8)	15 (7.5)
Hispanic ethnicity, n (%)			
Hispanic	25 (5.8)	25 (5.9)	10 (5.0)
Not Hispanic	410 (94.3)	398 (94.1)	189 (95.0)
Level of education completed, n (%)			
High school degree or less	100 (23.0)	96 (22.7)	54 (27.1)
Some college/associates degree	166 (38.2)	159 (37.6)	70 (35.2)
College graduate or higher	169 (38.9)	168 (39.7)	75 (37.7)
Currently working, n (%)			
Yes	202 (47.8)	202 (49.0)	84 (44.0)
No	221 (52.3)	210 (51.0)	107 (56.0)
Clinical			
Median (IQR) disease duration, years	14.7 (9.2–22.3)	14.6 (9.1–22.3)	14.7 (8.3–23.0)
Median (IQR) SLAQ score	11 (6–16)	11 (6–16)	11 (7–16)
Median (IQR) BILD score†	2 (1–4)	2 (1–4)	2 (1–4)
Self-reported moderate-severe forgetfulness, n (%)			
Yes	117 (27.2)	114 (27.1)	52 (26.4)
No	314 (72.9)	306 (72.9)	145 (73.6)
Neuropsychiatric damage, n (%)			
Yes	75 (17.2)	70 (16.6)	34 (17.1)
No	360 (82.8)	353 (83.5)	165 (82.9)
Currently taking steroids, n (%)			
Yes	182 (41.9)	181 (42.9)	76 (38.4)
No	252 (58.1)	241 (57.1)	122 (61.6)
Mean (SD) BMI, kg/m^2	30.2 (8.3)	30.3 (8.3)	30.8 (8.9)
Physical activity‡			
Low	317 (74.1%)	307 (73.8%)	144 (73.5%)
Moderate	65 (15.2%)	64 (15.4%)	29 (14.8%)
High	46 (10.8%)	45 (10.8%)	23 (11.7%)
Mean (SD) depression T-score§	48.3 (9.1)	48.3 (9.1)	48.5 (9.1)
Mean (SD) perceived stress score¶	15.2 (7.1)	15.3 (7.1)	15.2 (7.3)

Table 1 Continued

| Characteristic | Participants with complete assessments for: | | |
	CLOX (N=435)	Trail Making Test B (N=423)	NIH Toolbox Fluid Cognition (N=199)
Mean (SD) physical functioning T-score**	43.5 (8.9)	43.7 (8.8)	42.6 (8.3)
Mean (SD) SPPB score	9.1 (2.1)	9.1 (2.1)	9.0 (2.0)

*Represents sex assigned at birth.
†From the closest GOAL assessment.
‡From the International Physical Activity Questionnaire–Short Form.
§From the Patient Reported Outcomes Measurement Information System (PROMIS) Depression Short Form-8a.
¶From Cohen's 10-item Perceived Stress Scale (range 0–40; higher scores representing greater perceived stress).
**From the PROMIS Physical Functioning-Short Form 12a (higher T-scores representing better self-reported physical functioning).
BILD, Brief Index of Lupus Damage (range 0–46; 46 is maximum damage); BMI, body mass index; CLOX, clock drawing assessment; GOAL, Georgians Organized Against Lupus (parent study); NIH, National Institutes of Health; SLAQ, Systemic Lupus Activity Questionnaire (range 0–47; 47 is maximum activity); SPPB, Short Physical Performance Battery (range 0–12; higher scores=better performance).

potential impairment in CLOX only, 16% had impairment in TMTB only and 18% had no impairments. When fluid impairment was defined by domain percentiles, patterns were similar (online supplemental figure 4A). Potential episodic and working memory impairment were less common than overall fluid cognition impairment but overlap with either TMTB or CLOX impairment remained high, at 100% (online supplemental figure 4B) and 91% (online supplemental figure 4C). When potential TMTB impairment was alternately defined as a time <273 s, 34% of the cohort had no impairments and 0% had potential impairment in TMTB only (online supplemental figure 4D).

Finally, higher educational attainment, working status, higher self-reported physical functioning score and higher physical performance score were associated with lower odds of potential cognitive impairment across all measurements and definitions, although these associations were not all statistically significant (online supplemental table 2). Younger age, black race, higher disease activity and damage, and higher perceived stress were uniquely associated with higher odds of TMTB but not CLOX or fluid cognition impairment; neuropsychiatric damage was associated with higher odds TMTB and CLOX impairment, but only the former was statistically significant (online supplemental table 2). This damage was more associated with impairment in CLOX2 (OR=2.07, 95% CI 0.70 to 6.07) than CLOX1 (OR=1.53, 95% CI 0.92 to 2.55), although these associations were not statistically significant. Self-reported moderate to severe forgetfulness was not associated with any potential impairment.

DISCUSSION

In this comparison of assessments of cognitive performance among a population-based cohort with SLE, we found that 65%, 55% and 28% were potentially impaired by the TMTB test, CLOX task and NIH Fluid Cognition Battery, respectively. There was incomplete overlap in potential impairment between TMTB and CLOX, with 42% of those with potential impairment in either TMTB or CLOX having potential impairment in both. Few (2%) had impairment in fluid cognition only, and 18% had no potential impairments in any of the assessments. Although many people with SLE report cognitive impairment, quantifying and following this symptom over time have posed challenges. These results suggest that the two brief assessments (TMTB and CLOX) combined would capture most of those with impairment in fluid cognition. We also found that the TMTB and CLOX assessment provided non-overlapping information about cognitive performance in this population that may be important in the absence of other impairment.

For the TMTB task, using the crude cut-off of >273 s[42] to complete the task as a marker of impairment, only 3% of cohort was considered potentially impaired. However, when individuals' scores were instead compared with norms across multiple age and education groups, potential impairment (a time >1.5 SD than the norm for age and education)[23] was much higher (65%), suggesting that regardless of impairment, those with SLE do not perform as well as their peers on tasks that involve executive functioning tasks, including visual attention and task switching (equivalent to cognitive flexibility). While the NIH Toolbox Fluid Cognition Battery also measures many aspects of executive function, including cognitive flexibility, we found only a moderate negative correlation between TMTB time and fluid cognition or cognitive flexibility scores; overlap in potential impairment was modest. Additionally, TMTB impairment was associated with more patient characteristics than the fluid cognition score, including perceived stress and disease activity and damage. This suggests that, while the TMTB (originally designed to detect brain damage in the military)[45] and the NIH Toolbox Fluid Cognition Battery both measure high-level cognitive functioning involving the frontotemporal regions of the brain, they provide different information. However, the TMTB may be useful as an initial screener for cognitive impairment in SLE, particularly in patients who are unwilling or unable to do longer assessments, which can be distressing and fatiguing.

Table 2 Summary of continuous cognitive performance measures

Measure	All study visits			In-person visits only			Remote study visits		
	N	Mean (SD)	Median (IQR)	N	Mean (SD)	Median (IQR)	N	Mean (SD)	Median (IQR)
Trail Making Test B time, in seconds	423	110.1 (52.4)	96 (76–130)	202	109.4 (58.2)	91 (72–125)	221	110.8 (46.5)	98 (81–134)
CLOX scores*									
CLOX1	436	11.6 (2.5)	12 (10–13)	203	11.6 (2.4)	12 (10–13)	233	11.6 (2.4)	12 (10–13)
CLOX2	435	13.7 (1.4)	14 (13–15)	202	14.0 (1.1)	14 (13–15)	233	13.5 (1.7)	14 (13–15)
NIH Toolbox Fluid Cognition†									
Overall score	–	–	–	199	87.2 (15.6)	88 (76–97)	–	–	–
Episodic memory	435	97.1 (13.9)	96 (87–105)	202	96.2 (12.5)	95 (88–104)	233	97.8 (15.0)	97 (87–105)
Working memory	435	91.4 (15.5)	89 (80–103)	202	91.2 (17.0)	88 (79–105)	233	91.6 (14.2)	90 (81–100)

*Score range 0–15, with 15 representing a perfect clock. CLOX1 is the score for an unaided clock draw, while CLOX2 represents a clock copying task.
†Age-corrected standard scores, with a score of 100 representing the population mean and a score of 85 representing a score 1 SD below the mean.
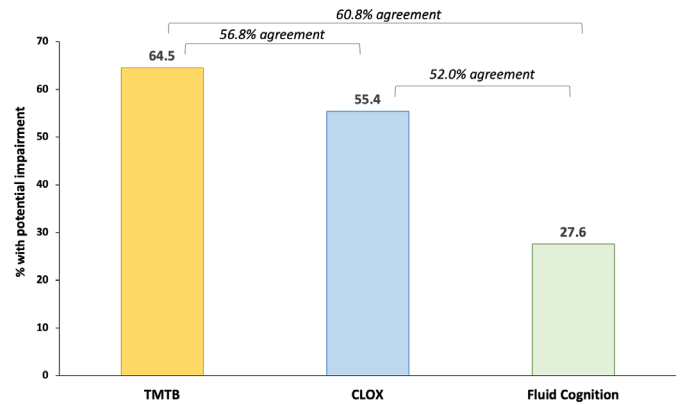CLOX, clock drawing assessment; NIH, National Institutes of Health.

Figure 2 Percentage of individuals with SLE with potentially impaired cognition and percentage agreement between measurements. For TMTB, impairment was defined as time >1.5 SD greater than the normative value for the participant's age group and education. CLOX impairment defined as a free draw (CLOX1) score ≤10 or copy draw (CLOX2) ≤12. For the NIH Toolbox Fluid Cognition scores, impairment was defined as an age-corrected standard score >1.5 SD lower than the general population mean. CLOX, clock drawing assessment; NIH, National Institutes of Health; TMTB, Trail Making Test B.

The CLOX also showed wide variation in potential impairment: 54% had potential impairment in the free clock draw but only 4% had impairment in the clock copy task. This pattern may reflect both the types of impairment in SLE and the purpose of the CLOX test.[26] Impairment in the free clock draw reflects executive function or frontotemporal impairment, which can occur early in both vascular and Alzheimer's-related dementia, as well as in major depressive disorder, hypothyroidism and polypharmacy.[13] Copy task impairment reflects posterior cortical impairment, suggesting early Alzheimer's-type impairment. We found that CLOX impairment, and particularly impairment was associated in the copy task, was non-statistically significantly associated with higher odds of neuropsychiatric damage, which may point to pathways by which this type of impairment might occur in SLE. Importantly, CLOX scores were only weakly associated with TMTB times and fluid cognition score, suggesting that this task is measuring domains that differ from both. For example, the free draw clock requires planning (eg, spacing numbers on the clock, including a short and long hand, etc), unlike the other tasks, and there is also no timed element, which is present in the TMTB and three of the five fluid cognition tasks. Thus, like the TMTB, the CLOX may be useful as a screening tool, but neither the CLOX or TMTB alone would be likely to capture multiple domains of performance and potential impairment. Further, the CLOX (particularly, the free clock draw portion) may identify cognitive impairment in this population, but the aetiology of impairment would require further exploration.

In our study, cut-offs for TMTB (age and education adjusted) and CLOX impairment combined captured most of the individuals identified as potentially impaired by the NIH Fluid Cognition Battery, which supports the
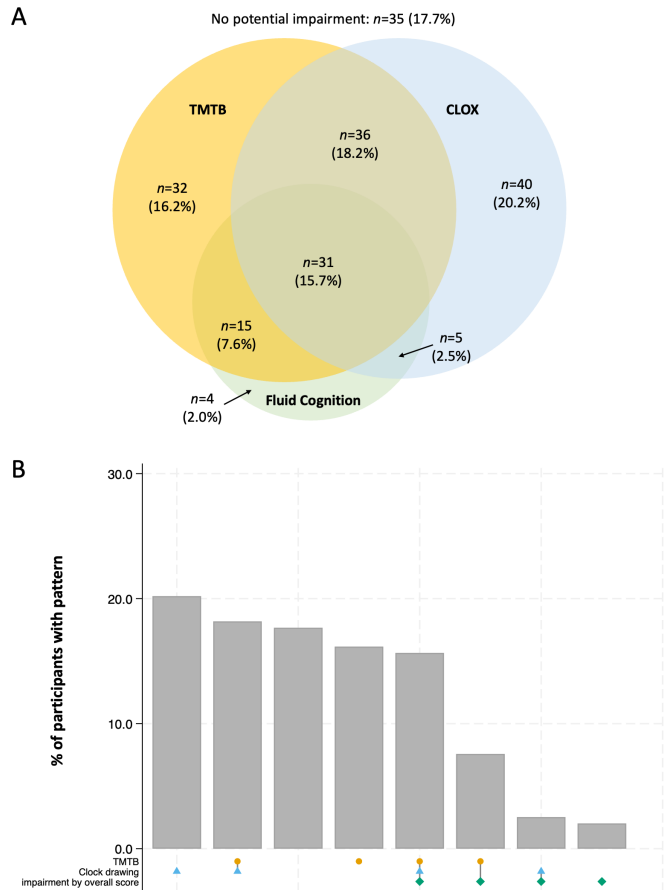
Figure 3 Overlap of potential impairment in domains visualised by Venn diagram (A) and UpSet plot (B). For TMTB, impairment was defined as time >1.5 SD greater than the normative value for the participant's age group and education. For clock drawing, impairment defined as a free draw (CLOX1) score ≤10 or copy draw (CLOX2) ≤12. For the fluid cognition scores, impairment was defined as an age-corrected standard score >1.5 SD lower than the general population mean. CLOX, clock drawing assessment; TMTB, Trail Making Test B.

use of the shorter assessments for screening purposes. However, this method is likely to capture a large proportion of patients with SLE, which undermines the utility if there are inadequate resources to do additional testing, evaluation and management (eg, neuropsychology consults, memory clinics) for large numbers of patients with SLE. Additionally, while the NIH Fluid Cognition Test is longer (20–30 vs 5–10 min), it provides unique information that the TMTB and CLOX cannot, including overall and domain-specific scores that could be used to guide treatment and support.[46] For example, we found that self-reported forgetfulness was not associated with overall fluid cognition impairment. If we consider a patient who reports memory issues and has normal episodic and working memory scores but potentially impaired attention and inhibitory control, the approach to improving perceived memory may shift to strategies to improve focus rather than memory. Another key advantage of the NIH Fluid Cognition Test is that it also allows

for comparison with the general population (via provided norms) and across studies, including studies of other populations.

Although our measurements are not widely used in SLE, making it difficult to compare our results with other studies of populations with SLE, they were chosen to provide multidomain information on how patients with SLE are doing relative to other populations with the same measurements (NIH Toolbox Fluid Cognition Battery) and information on how patients with SLE fare on rapid screening tests primarily used in geriatric settings (TMTB and CLOX). The ACR-NB[47] might be considered a gold standard against which we could compare our results, estimating measures of validity like sensitivity and specificity; it is used inconsistently,[10] likely due to its length (around 1 hour). Previous studies in SLE have shown that the Automated Neuropsychological Assessment Metrics, a multidomain test, did well relative to ACR-NB[48]; but the Montreal Cognitive Assessment, a rapid assessment for the clinical setting, had limited sensitivity and specificity relative to the ACR-NB.[6] However, we cannot assess whether our measurements follow this pattern since we did not include the ACR-NB. Additionally, cut-offs for impairment do not necessarily provide a complete picture of cognitive functioning, particularly on the individual level and among those who are high-functioning, which is highly correlated with education.[49] Finally, changes over time in scores within an individual patient, regardless of whether the cut-off for impairment is reached, could have substantial impacts on daily life. Future studies to track individual trajectories of cognitive performance and their impact on patient-reported outcomes, including quality of life and disability, are warranted.

Our study has several limitations. While the study compares multiple measures of cognition in the same cohort of individuals with SLE as part of a larger study, it was not designed as a validation study. Thus, measures that may have provided useful information, such as the ACR-NB (as a potential gold standard) and/or other measures of memory or attention, were not included. There is also the possibility of residual confounding by unknown or unmeasured factors (eg, pain, sleep quality, serologies including anti-phospholipid antibodies) or factors that may be inadequately measured (eg, participant-reported activity vs physician-assessed disease activity); future studies could also address the association of cognitive performance with these clinical variables. Additionally, we cannot yet estimate the association between cognitive performance across various measures and outcomes such as mortality, institutionalisation or healthcare utilisation. Because of COVID-19-related changes to our protocol, we cannot rule out misclassification due to visit type[39] or changes in performance due to pandemic-related factors. Finally, not all our measures have norms based on educational attainment, and CLOX normative values were not available except for those aged ≥75 years,[50] which may affect comparisons of measures.

In conclusion, we found that the TMTB, CLOX and NIH Fluid Cognition Battery each provided unique and potentially important information about cognitive performance in our SLE cohort. While the TMTB and CLOX are less burdensome and could be used as screening tests in less time, the NIH Fluid Cognition Battery provides more domain-specific information that may allow more targeted interventions. Future studies are needed to validate these measures in SLE, follow these measures over time within individuals, assess the clinical implications of impairment by these measures, and explore the outcomes of each of these performance measures and their change over time.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. The Emory Institutional Review Board approved the APPEAL (IRB00110977) and GOAL (IRB00003656) study protocols. All APPEAL participants provided informed consent before completing study visits.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. De-identified data will be made available per NIH requirements when funding is complete.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Laura Plantinga http://orcid.org/0000-0003-0809-8981
Jinoos Yazdany http://orcid.org/0000-0002-3508-4094
S Sam Lim http://orcid.org/0000-0003-2361-0787

REFERENCES

1. Hanly JG, Fisk JD, Sherwood G, *et al*. Clinical course of cognitive dysfunction in systemic lupus erythematosus. *J Rheumatol* 1994;21:1825–31.
2. Hanly JG, Fisk JD, Sherwood G, *et al*. Cognitive impairment in patients with systemic lupus erythematosus. *J Rheumatol* 1992;19:562–7.
3. Mizrachi M, Anderson E, Carroll KR, *et al*. Cognitive dysfunction in SLE: an understudied clinical manifestation. *J Autoimmun* 2022;132:102911.
4. Lim XR, Chew POK, Lim GH, *et al*. Montreal cognitive assessment as a screening instrument for cognitive impairment in systemic lupus erythematosus patients without overt neuropsychiatric manifestations. *Lupus* 2022;31:1759–69.
5. Raghunath S, Glikmann-Johnston Y, Golder V, *et al*. Clinical associations of cognitive dysfunction in systemic lupus erythematosus. *Lupus Sci Med* 2023;10:e000835.
6. Tayer-Shifman OE, Yuen K, Green R, *et al*. Assessing the utility of the Montreal cognitive assessment in screening for cognitive impairment in patients with systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* 2023;75:569–77.
7. Raghunath S, Glikmann-Johnston Y, Vincent FB, *et al*. Patterns and prevalence of cognitive dysfunction in systemic lupus erythematosus. *J Int Neuropsychol Soc* 2023;29:421–30.
8. Kozora E, Arciniegas DB, Filley CM, *et al*. Cognitive and neurologic status in patients with systemic lupus erythematosus without major neuropsychiatric syndromes. *Arthritis Rheum* 2008;59:1639–46.
9. Kozora E, Ellison MC, West S. Reliability and validity of the proposed American College of Rheumatology neuropsychological battery for systemic lupus erythematosus. *Arthritis Rheum* 2004;51:810–8.
10. Yuen K, Green R, Bingham K, *et al*. Metrics and definitions used in the assessment of cognitive impairment in systemic lupus erythematosus: a systematic review. *Semin Arthritis Rheum* 2021;51:819–30.
11. Kello N, Anderson E, Diamond B. Cognitive dysfunction in systemic lupus erythematosus: a case for initiating trials. *Arthritis Rheumatol* 2019;71:1413–25.
12. Panopalis P, Julian L, Yazdany J, *et al*. Impact of memory impairment on employment status in persons with systemic lupus erythematosus. *Arthritis Rheum* 2007;57:1453–60.
13. Royall DR, Cordes JA, Polk M. CLOX: an executive clock drawing task. *J Neurol Neurosurg Psychiatry* 1998;64:588–94.
14. Northwestern University. Cognition measures: NIH Toolbox cognition batteries. 2017. Available: http://www.healthmeasures.net/explore-measurement-systems/nih-toolbox/intro-to-nih-toolbox/cognition
15. Gershon RC, Cella D, Fox NA, *et al*. Assessment of neurological and behavioural function: the NIH toolbox. *Lancet Neurol* 2010;9:138–9.
16. Weintraub S, Dikmen SS, Heaton RK, *et al*. The cognition battery of the NIH Toolbox for assessment of neurological and behavioral function: validation in an adult sample. *J Int Neuropsychol Soc* 2014;20:567–78.
17. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725.
18. Drenkard C, Rask KJ, Easley KA, *et al*. Primary preventive services in patients with systemic lupus erythematosus: study from a population-based sample in Southeast U.S. *Semin Arthritis Rheum* 2013;43:209–16.
19. Lim SS, Drenkard C. Understanding lupus disparities through a social determinants of health framework: the Georgians organized against lupus research cohort. *Rheum Dis Clin North Am* 2020;46:613–21.
20. Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research Informatics support. *J Biomed Inform* 2009;42:377–81.
21. Plantinga L, Tift BD, Dunlop-Thomas C, *et al*. Geriatric assessment of physical and cognitive functioning in a diverse cohort of systemic lupus erythematosus patients: a pilot study. *Arthritis Care Res (Hoboken)* 2018;70:1469–77.

22 Corrigan JD, Hinkeldey NS. Relationships between parts A and B of the trail making test. *J Clin Psychol* 1987;43:402–9.

23 Tombaugh TN. Trail making test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol* 2004;19:203–14.

24 Aprahamian I, Martinelli JE, Neri AL, *et al*. The clock drawing test: a review of its accuracy in screening for dementia. *Dement Neuropsychol* 2009;3:74–81.

25 Freund B, Gravenstein S, Ferris R, *et al*. Drawing clocks and driving cars. *J Gen Intern Med* 2005;20:240–4.

26 Schillerstrom JE, Sawyer Baker P, Allman RM, *et al*. Clock drawing phenotypes in community-dwelling African Americans and Caucasians: results from the University of Alabama at Birmingham study of aging. *Neuroepidemiology* 2007;28:175–80.

27 Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics* 1993;4:353–65.

28 Tang K, Beaton DE, Boonen A, *et al*. Measures of work disability and productivity: rheumatoid arthritis specific work productivity survey (WPS-RA), workplace activity limitations scale (WALS), work instability scale for rheumatoid arthritis (RA-WIS), work limitations questionnaire (WLQ), and work productivity and activity impairment questionnaire (WPAI). *Arthritis Care Res (Hoboken)* 2011;63 Suppl 11:S337–49.

29 Gordon C, Isenberg D, Lerstrøm K, *et al*. The substantial burden of systemic lupus erythematosus on the productivity and careers of patients: a European patient-driven online survey. *Rheumatology (Oxford)* 2013;52:2292–301.

30 Karlson EW, Daltroy LH, Rivest C, *et al*. Validation of a systemic lupus activity questionnaire (SLAQ) for population studies. *Lupus* 2003;12:280–6.

31 Yazdany J, Trupin L, Gansky SA, *et al*. Brief index of lupus damage: a patient-reported measure of damage in systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* 2011;63:1170–7.

32 Drenkard C, Yazdany J, Trupin L, *et al*. Validity of a self-administered version of the brief index of lupus damage in a predominantly African American systemic lupus erythematosus cohort. *Arthritis Care Res (Hoboken)* 2014;66:888–96.

33 Teresi JA, Ocepek-Welikson K, Kleinman M, *et al*. Psychometric properties and performance of the patient reported outcomes measurement information system(®) (PROMIS(®)) depression short forms in ethnically diverse groups. *Psychol Test Assess Model* 2016;58:141–81.

34 Hitchon CA, Zhang L, Peschken CA, *et al*. The validity and reliability of screening measures for depression and anxiety disorders in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2020;72:1130–9.

35 Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983;24:385–96.

36 Ezzati A, Jiang J, Katz MJ, *et al*. Validation of the perceived stress scale in a community sample of older adults. *Int J Geriatr Psychiatry* 2014;29:645–52.

37 Lee PH, Macfarlane DJ, Lam TH, *et al*. Validity of the International physical activity questionnaire short form (IPAQ-SF): a systematic review. *Int J Behav Nutr Phys Act* 2011;8:115.

38 Northwestern University. Introduction to PROMIS 2017. n.d. Available: http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis

39 Hoge C, Bowling CB, Dunlop-Thomas C, *et al*. Remote administration of physical and cognitive performance assessments in a predominantly black cohort of persons with systemic lupus erythematosus. *ACR Open Rheumatol* 2023;5:499–507.

40 Plantinga LC, Bowling CB, Hoge C, *et al*. Physical performance in a diverse, population-based cohort of individuals with systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* 2024;76:517–25.

41 Guralnik JM, Simonsick EM, Ferrucci L, *et al*. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol* 1994;49:M85–94.

42 Lezak MDH, Loring DW. *Neuropsychological assessment*. 4th edn. New York: Oxford University Press, 2004.

43 Holdnack JA, Tulsky DS, Brooks BL, *et al*. Interpreting patterns of low scores on the NIH Toolbox cognition battery. *Arch Clin Neuropsychol* 2017;32:574–84.

44 NIH Toolbox. Remote administration guidelines for the NIH Toolbox: response to COVID-19. 2020. Available: https://nihtoolbox.zendesk.com/hc/en-us/article_attachments/13434157309332]

45 Reitan RM. Validity of the trail making test as an indicator of organic brain damage. *Percept Mot Skills* 1958;8:271–6.

46 Plantinga L, Bowling CB, Dunlop-Thomas C, *et al*. Cognitive functioning among individuals with systemic lupus erythematosus: a population-based study. *Arthritis Rheumatol* 2023;75.

47 Yuen K, Beaton D, Bingham K, *et al*. Validation of the automated neuropsychological assessment Metrics for assessing cognitive impairment in systemic lupus erythematosus. *Lupus* 2022;31:45–54.

48 Tayer-Shifman OE, Green R, Beaton DE, *et al*. Validity evidence for the use of automated neuropsychologic assessment Metrics as a screening tool for cognitive impairment in systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* 2020;72:1809–19.

49 Karr JE, Iverson GL. Interpreting high scores on the NIH Toolbox cognition battery: potential utility for detecting cognitive decline in high-functioning individuals. *Neuropsychology* 2020.

50 Crowe M, Allman RM, Triebel K, *et al*. Normative performance on an executive clock drawing task (CLOX) in a community-dwelling sample of older adults. *Arch Clin Neuropsychol* 2010;25:610–7.