# Measuring reliable change in cognition using the Edinburgh Cognitive and Behavioural ALS Screen (ECAS)

Christopher Crockford, Judith Newton, Katie Lonergan, Caoifa Madden, Iain Mays, Meabhdh O'Sullivan, Emmet Costello, Marta Pinto-Grau, Alice Vajda, Mark Heverin, Niall Pender, Ammar Al-Chalabi, Orla Hardiman & Sharon Abrahams

Taylor & Francis
Taylor & Francis Group

Check for updates

RESEARCH ARTICLE

# Measuring reliable change in cognition using the Edinburgh Cognitive and Behavioural ALS Screen (ECAS)

CHRISTOPHER CROCKFORD[1,2] (iD), JUDITH NEWTON[1,3], KATIE LONERGAN[4,5],
CAOIFA MADDEN[4,5], IAIN MAYS[4,5], MEABHDH O'SULLIVAN[4],
EMMET COSTELLO[4,5], MARTA PINTO-GRAU[4,5], ALICE VAJDA[4], MARK HEVERIN[4],
NIALL PENDER[5], AMMAR AL-CHALABI[6] (iD), ORLA HARDIMAN[4,7] (iD)
& SHARON ABRAHAMS[1,2,3]

[1]Department of Psychology, Human Cognitive Neuroscience, University of Edinburgh, Edinburgh, UK, [2]Euan MacDonald Centre for Motor Neurone Disease Research, Royal Infirmary of Edinburgh, Edinburgh, UK, [3]Anne Rowling Regenerative Neurology Clinic, Royal Infirmary of Edinburgh, Edinburgh, UK, [4]Academic Unit of Neurology, Trinity College Dublin, Dublin, Ireland, [5]Department of Psychology, Beaumont Hospital, Dublin, Ireland, [6]Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, King's College London, London, UK, and [7]Department of Neurology, Beaumont Hospital, Dublin, Ireland

## Abstract

*Background*: Cognitive impairment affects approximately 50% of people with amyotrophic lateral sclerosis (ALS). Research has indicated that impairment may worsen with disease progression. The Edinburgh Cognitive and Behavioural ALS Screen (ECAS) was designed to measure neuropsychological functioning in ALS, with its alternate forms (ECAS-A, B, and C) allowing for serial assessment over time. *Objective*: The aim of the present study was to establish reliable change scores for the alternate forms of the ECAS, and to explore practice effects and test-retest reliability of the ECAS's alternate forms. *Method*: Eighty healthy participants were recruited, with 57 completing two and 51 completing three assessments. Participants were administered alternate versions of the ECAS serially (A-B-C) at four-month intervals. Intra-class correlation analysis was employed to explore test-retest reliability, while analysis of variance was used to examine the presence of practice effects. Reliable change indices (RCI) and regression-based methods were utilized to establish change scores for the ECAS alternate forms. *Results*: Test-retest reliability was excellent for ALS Specific, ALS Non-Specific, and ECAS Total scores of the combined ECAS A, B, and C (all > .90). No significant practice effects were observed over the three testing sessions. RCI and regression-based methods produced similar change scores. *Conclusion*: The alternate forms of the ECAS possess excellent test-retest reliability in a healthy control sample, with no significant practice effects. The use of conservative RCI scores is recommended. Therefore, a change of $\geq 8$, $\geq 4$, and $\geq 9$ for ALS Specific, ALS Non-Specific, and ECAS Total score is required for reliable change.

**Keywords:** *Cognition, ECAS, alternate forms, test reliability, practice*

## Introduction

Cognitive and behavioural symptoms affect approximately 50% of patients with amyotrophic lateral sclerosis (ALS), of whom 15% develop FTD (frontotemporal dementia) and the two form a spectrum disease. Executive dysfunction, language dysfunction, and social cognitive deficits are commonly reported (1–3). The presence of cognitive and behavioural symptoms in ALS can precede motor symptoms (4), have been associated with reduced survival (5,6), disengagement with life prolonging interventions (7), and increased caregiver burden (8,9). Neuropsychological status has additional implications for end-of-life care planning, capacity to consent, and powers of attorney (2). Thus, timely and accurate knowledge of patients' cognitive and behavioural status is vital for providing person-centred care.

The Edinburgh Cognitive and Behavioural ALS Screen (ECAS) was developed to offer a comprehensive screening tool to assess the cognitive and behavioural status of patients with ALS (10). The ECAS has been validated against full neuropsychological test batteries (11–16) and was designed specifically for patients with ALS in that it accommodates motor disability.

Cognitive decline over the course of the disease (17), or in response to specific disease factors such as respiratory insufficiency (18), has been suggested and it is consequently important to monitor progression of cognitive symptoms. Howevert, repeated assessment using neuropsychological tests may result in improvement due to practice effects, i.e. improvement in performance due to learning test content, or test-taking strategies. For clinicians, it can be difficult to interpret whether an observed difference in test performance is due to true change in the patient's situation (recovery or decline) or extraneous factors. Measurement error, regression to the mean, and practice effects can produce or exaggerate changes in performance between testing sessions. Furthermore, demographic factors such as age, education level, and baseline performance can influence change in scores between testing sessions, and therefore, obscure a patient's true performance variation (19). With regard to the ECAS, practice effects have been demonstrated with healthy controls in the executive, language, and memory domains over six months, when using version A for repeat assessments (20).

Recently, alternate forms of the ECAS have been developed to accommodate the repeated assessment of patients with ALS (21). The alternate forms of the ECAS (the ECAS-B and ECAS-C) were designed to retain the construct characteristics and level of difficulty of the original ECAS-A while reducing potential practice effects. The ECAS-B and C have shown to be equivalent to the original ECAS-A, and resistant to practice effects from repeated administration (21). It has yet to be determined what can be regarded as a meaningful change on a case-by-case basis. Numerous methods have been proposed to support the interpretation of change scores on neuropsychological tests. Most notably, clinicians have utilized reliable change indices (RCI) and regression-based methods.

With regard to RCI methods, change in scores for an individual patient is interpreted in the context of normal healthy variation, such that an observed change in a patient's score needs to fall outside of the standard error of healthy controls' test-retest variability (22). Numerous variations of the RCI have been developed which adjust for factors such as practice effects, and regression to the mean (23). Conversely, regression-based methods employ regression models to predict performance at follow-up from initial test performance. Again, significant differences between a patient's predicted and actual score are used to determine reliable change. The regression-based method additionally allows the inclusion of moderating variables such as age and education, and controls for practice effects and regression to the mean (24). RCI and regression-based methods allow clinicians to interpret patients' change scores or can provide a meaningful interpretation or endpoint for clinical trials.

The aims of this study are: (1) to examine whether practice effects are observed using the ECAS alternate forms over clinically meaningful test-retest intervals; (2) to determine test-retest reliability of the ECAS-A-B-C over clinically meaningful intervals; and (3) to compare common methods for measuring reliable change in a patient's ECAS score across serially administered alternate versions.

## Method

### Participants

Eighty Irish and Scottish healthy adults were recruited representative of the demographic characteristics of ALS patients. Only those participants who completed the ECAS at two or more time-points were included in the present study. Fifty-seven participants completed one follow-up assessment, and 51 participants completed two follow-up assessments. Exclusion criteria included: a history of dyslexia or marked premorbid reading or writing difficulties or a learning disability; non-fluent English reading and writing abilities; history of neurological conditions that could affect cognition such as major hemispheric stroke, traumatic brain injury, and severe active epilepsy; alcohol and drug dependencies; and having a known blood relative with ALS. Participants were recruited through a research volunteer panel and through local community noticeboards. Non-blood relatives of ALS patients were also recruited as control participants. All participants provided informed written consent and this research was approved by the University of Edinburgh Psychology Research Ethics Committee and the Beaumont Research Ethics Committee. Participants' travel costs associated with participation were reimbursed.

### Procedure

Participants were assessed every four months for three occasions. The ECAS is an ALS-designed measure of cognitive and behavioural functioning. For the purposes of this study, the ECAS behavioural interview was not included. The ECAS consists of three versions (A-B-C) which were designed to be administered serially. Each version of the ECAS consists of 15 parallel tests, categorized into five cognitive domains. Executive, language, and verbal fluency domains are described as ALS Specific functions, while the memory and

visuospatial domains are described as ALS Non-Specific. The ALS-Specific and ALS Non-Specific domains combine to generate a measure of global cognitive functioning, namely, the ECAS Total score. At each assessment point, an alternate version of the ECAS was administered such that the ECAS-A was given at Time 1, the ECAS-B at Time 2, and the ECAS-C at Time 3.

### Statistical analysis

Statistical analyses were conducted using R 3.3.2. Change scores were calculated for each time comparison by subtracting the baseline score from the follow-up score, i.e. (Time 2 – Time 1, Time 3 – Time 2, and Time 3 – Time 1). Welch *t*-tests were used to compare change scores between centres to ensure comparability. When data did not meet assumptions of normality, Mann-Whitney *U*-tests were employed. In all cases, Time 1 is synonymous with ECAS-A, Time 2 with ECAS-B, and Time 3 with ECAS-C.

*Test-retest reliability*: of the alternate forms of the ECAS (A-B-C) was examined using intraclass correlation coefficients (ICC) with mean-rating absolute agreement two-way random effects models. ICC coefficients were calculated for the component (language, fluency, executive, memory, visuospatial) and composite (ALS Specific, ALS Non-Specific, ECAS Total) domains of the ECAS.

*Practice effects:* were explored using one-way repeated measures analysis of variance models (ANOVA) to examine the presence of a main effect of Time. Repeated measures ANOVAs are limited to balanced designs and only those participants who completed all three time-points were included in this analysis.

*Change indices:* were calculated using four types of model: two regression-based methods and two reliable change index (RCI) methods. Each method corrects for slightly different moderating effects. The RCI JT method accounts for measurement error, while the Chelune method additionally accounts for practice effects. While significant practice effects in using alternate versions of the ECAS have not previously been found (21), small improvements may be present which might not reach statistical significance. The simple regression method incorporates corrections for regression to the mean, whereby individuals who perform in the extremes tend to converge on the group mean at follow-up. Finally, the multiple regression method allows for the incorporation of potential moderating variables such as age and education that may influence change over time. Given the higher sensitivity of the ALS Specific, ALS Non-Specific, and ECAS Total scores to cognitive impairment against a full neuropsychological battery (13), change score analysis was conducted for these composite domains.

### Method 1: RCI (JT method)

The first RCI model calculated was the Jacobson and Truax method (JT method) (25), which accounts for measurement error. The JT method is calculated as the difference between Time 2 and Time 1 divided by the standard error of difference ($SE_{diff}$) between these two time-points. The standard error of the difference is derived from the standard error of the measurement ($SE_m$) such that $SE_{diff} = \sqrt{2(SE_m)^2}$. The standard error of the measurement ($SE_m$) is calculated with the equation $SE_m = s_1\sqrt{1 - r_{xx}}$, where $s_1$ is the standard deviation of the Time 1 ECAS (i.e. for ECAS-B, the preceding version is ECAS-A, and for ECAS-C the preceding version is ECAS-B) and $r_{xx}$ is the test-retest reliability coefficient between these two ECAS forms. Therefore, the RCI equation using the JT method is calculated with the formula:

$$RCI = \frac{x_2 - x_1}{SE_{diff}}$$

Reliable change is defined by values larger than $\pm 1.645$ (two-tailed 90% confidence interval). The formula was then restructured to calculate the upper and lower 'thresholds' of reliable change ($X$) i.e. the number of points increase/decrease required between two testing sessions, which constitutes a reliable change. Therefore, the equation was restructured as:

$$\pm\Delta X = 1.645(SE_{diff})$$

### Method 2: RCI (chelune method)

The second RCI method employed is the Chelune method (26), which corrects for measurement error and practice effects. While the alternate versions of the ECAS were developed to account for practice effects, it is possible that small non-significant improvements continue to exist. Additionally, the accounting for practice effects here may help to account for any small but non-significant differences in difficulty present in the alternate versions. The Chelune method is similar to the JT method, taking the form of:

$$RCI = \frac{(x_2 - x_1) - (\bar{X}_2 - \bar{X}_1)}{SE_{diff}}$$

Here, the denominator is again the standard error of the difference $SE_{diff}$ (i.e. $\sqrt{2(SE_m)^2}$). However, the Chelune method adds a constant as to the numerator to account for systematic changes in performance such as practice effects. This is achieved by calculating the mean difference in performance between Time 2 and Time 1 $(\bar{X}_2 - \bar{X}_1)$ and subtracting this from an individual's

change score. As before, this equation was restructured to solve for $X$ resulting in:

$$\Delta X = (\bar{X}_2 - \bar{X}_1) \pm 1.645(SE_{diff})$$

### Method 3: Simple linear regression

The first regression-based method employs a simple linear regression model that predicts follow-up performance based on the preceding performance. First, a patient's predicted Time 2 score is calculated using the basic regression equation:

$$\hat{X} = \beta X + C$$

Where $\hat{X}$ is an individual's predicted Time 2 score, $\beta$ is the beta coefficient for the predictor in the model, $X$ is the Time 1 score, and $C$ is the intercept estimate of the model. Next, the discrepancy between the observed Time 2 score and the predicted Time 2 score is calculated and referred to as the residual (i.e. Time 2 – predicted Time 2). To extract a change index, this residual is then divided by the residual standard error (or standard error of the estimate; $SEE$). When values of the residual divided by the SEE are greater than $\pm 1.645$ (two-tailed 90% confidence interval) a reliable change can be determined. To determine reliable change 'thresholds', the equation is restructured to solve for the residual such that:

$$\pm(X - \hat{X}) = 1.645(SEE)$$

Where $(X - \hat{X})$ is the difference between the observed score and the estimated score. The same procedure is used for predicting Time 3 performance from Time 2.

### Method 4: Multiple linear regression

The second regression-based method is a multiple linear regression model to explore whether age, education, sex, preceding performance, or testing interval affects the model's prediction. Potential predictors were selected based on their correlation with the respective outcome variable (i.e. ALS Specific, ALS Non-Specific, or ECAS Total scores). A relationship with sex was explored using Mann-Whitney $U$-tests. Variables of interest were entered into each model in a single block and only

retained if their individual contribution to the model was significant (i.e. backward elimination). Significantly influential cases were removed based on diagnostic plots. Multiple regression equations, similar to simple linear regression equations, take the form of:

$$\hat{X} = \beta_1 X_1 + \beta_2 X_2 \ldots + \ldots \beta_j X_j + C$$

Where $\beta_1$ is the coefficient of the first variable of interest, $X_1$ is the observed score for the first variable, $\beta_2$ is the coefficient for the second variable, and so on to the $\mathcal{J}$th variable. Again, this equation is restructured to solve for $(X - \hat{X})$ such that:

$$\pm(X - \hat{X}) = 1.645(SEE)$$

## Results

Fifty-seven participants completed two or more time-points and were included in the present study. Participants were 61.40% male ($n = 35$) with a mean age of $62.32 \pm 13.36$ years and $14.87 \pm 3.19$ years of education. Mean test-retest intervals were $4.30 \pm 0.66$ months and $3.99 \pm 0.54$ months for T1 to T2, and for T2 to T3, respectively. Mean change scores were calculated for Time 1 to Time 2, Time 2 to Time 3, and Time 1 to Time 3. No significant differences were observed in mean change scores, nor in age, gender, or education between Irish and Scottish participants (all $p > 0.05$) indicating comparability.

### Practice effects

Mean test performance for the subdomains of the ECAS and its alternate versions are displayed in Table 1. Mean performance for all cognitive sub-domains (language, executive, fluency, memory, and visuospatial) are analogous across time-points, such that no mean difference exceeded 1 point. The resulting ANOVA models produced no significant main effect for executive functions ($F(2,100) = 1.43$, $p = 0.24$), fluency ($F(2,100) = .25$, $p = 0.780$), or memory ($F(2,100) = 1.04$, $p = 0.358$). Additionally, no main effect was observed for the composite ALS Specific ($F(2,100) = .852$, $p = 0.43$), ALS Non-Specific ($F(2,100) = .838$, $p = 0.435$), nor ECAS Total ($F(2,100) = .428$, $p = 0.653$). Due to ceiling

Table 1. Mean scores and standard deviations for the ECAS-A, ECAS-B, and ECAS-C.

|  | ECAS-A | ECAS-B | ECAS-C |
|---|---|---|---|
| ALS Specific (max 100) | 85.65 ± 8.56 | 86.39 ± 8.55 | 86.20 ± 9.49 |
| Language (max 28) | 27.14 ± 1.87 | 26.74 ± 2.09 | 26.76 ± 1.89 |
| Fluency (max 24) | 19.61 ± 2.71 | 19.82 ± 2.43 | 19.69 ± 4.22 |
| Executive (max 48) | 38.89 ± 5.59 | 39.82 ± 5.37 | 39.75 ± 5.68 |
| ALS Non-Specific (max 36) | 30.14 ± 3.73 | 29.89 ± 3.66 | 30.41 ± 3.11 |
| Memory (max 24) | 18.46 ± 3.43 | 18.51 ± 3.13 | 18.88 ± 2.89 |
| Visuospatial (max 12) | 11.68 ± 0.74 | 11.39 ± 1.00 | 11.53 ± 0.70 |
| ECAS Total (max 136) | 115.79 ± 11.26 | 116.28 ± 11.38 | 116.61 ± 11.87 |

effects and thus the presence of ties in the language and visuospatial subtests, formal analysis was not conducted. However, given the similarity in mean scores and the lack of significant differences for their composite domains, a lack of observable practice effects may be assumed. Therefore, no evidence of practice effects was found for the repeated assessment using the ECAS alternate versions (A-B-C) over clinically relevant test-retest intervals of three to four months.

### Test-retest reliability

Two-way absolute agreement mixed effects ICC models were generated for each subdomain of the ECAS (see Supplementary materials). Test-retest reliability was good for the majority of subdomains (i.e. >.70). The only subtest to fall below .70 was the visuospatial domain. However, a lack of variability due to participants reaching ceiling (i.e. over 50% scoring 12 points for each version of the ECAS) will exacerbate small differences in performance rendering this result unreliable (27). With regard to the ALS Specific, ALS Non-Specific, and ECAS Total scores, test-retest reliability was excellent.

### Change indices

#### RCI method: Methods 1 and 2

Reliable change indices (RCI) were calculated for ALS Specific, ALS Non-Specific, and ECAS Total score using the Jacobson and Truax method (JT method) (25) and the Chelune method (26). Thresholds for reliable change are displayed in Table 2, in addition to recommended clinical thresholds. Data used to calculate RCI thresholds are available in Supplementary Table 1. These thresholds provide the number of points increase required to determine a reliable improvement in cognitive performance and the number of points decrease required to determine a reliable decline in functioning. For example, using the JT method, a drop of 8.23 points or greater in ECAS Total score between the ECAS-A and ECAS-B is required for a reliable decline, while an increase of 8.23 points constitutes a reliable improvement. By comparison, using the Chelune method a drop of 7.74 points is

required for a reliable decline, or an increase of 8.72 points for a reliable improvement due to the incorporation of practice effects.

### Method 3: Simple linear regression

Simple linear regression models were built to predict follow-up scores based on previous performance. Table 3 provides data for calculating predicted ECAS-B performance from ECAS-A, and for predicting ECAS-C from ECAS-B using the equation $\hat{X} = \beta X + C$. This predicted score can then be converted into a change index with $\pm 1.645$ constituting a reliable deviation for predicted performance. Alternatively, the column $X - \hat{X}$ in Table 3 provides upper and lower thresholds calculated as $\pm (X - \hat{X}) = 1.645(SEE)$.

### Method 4: Multiple linear regression

Variables of interest were explored as potential moderating factors in multiple regression models using correlational analysis. For the ECAS-B, education level significantly correlated with ALS Specific ($r = .42$, $p = 0.002$), ALS Non-Specific ($r = .49$, $p < 0.001$), and ECAS Total ($r = .47$, $p < 0.001$) scores. Age at testing and test-retest interval did not significantly correlate with ECAS-

Table 3. Simple linear regression equations for predicting ECAS-B and ECAS-C performance.

|  | $R^2$ | $SEE$ | C | $\beta_{(ECAS-A)}$ | $X - \hat{X}$ |
|---|---|---|---|---|---|
| **Predicting ECAS-B from ECAS-A** |  |  |  |  |  |
| ALS Specific | 0.398 | 5.29 | 34.18 | 0.613 | ±8.70 |
| ALS Non-Specific | 0.297 | 2.25 | 15.31 | 0.490 | ±3.70 |
| ECAS Total | 0.448 | 6.23 | 41.62 | 0.648 | ±10.24 |
|  | $R^2$ | $SEE$ | C | $\beta_{(ECAS-B)}$ | $X - \hat{X}$ |
| **Predicting ECAS-C from ECAS-B** |  |  |  |  |  |
| ALS Specific | 0.700 | 5.30 | 34.18 | 0.938 | ±8.72 |
| ALS Non-Specific | 0.597 | 2.00 | 11.60 | 0.628 | ±3.29 |
| ECAS Total | 0.741 | 6.12 | 14.78 | 0.872 | ±10.07 |

$R^2$ is the multiple $R^2$. $SEE$ is the residual standard error, C is the intercept, $\beta$ is the beta coefficient associated with the subscript ECAS, $X - \hat{X}$ is the residual (i.e. the difference between the model predicted score and the observed score). The $X - \hat{X}$ column indicates the number of points difference required between observed and estimated score to determine reliable difference – this is calculated as 1.645*($SEE$).

Table 2. RCI thresholds for the ECAS-A, ECAS-B, and ECAS-C.

|  | JT method | | | Chelune method | | |
|---|---|---|---|---|---|---|
|  | ALS Specific ΔX | ALS Non-Specific ΔX | ECAS Total ΔX | ALS Specific ΔX | ALS Non-Specific ΔX | ECAS Total ΔX |
| ECAS-A to ECAS-B | ± 7.23 | ± 3.35 | ± 8.23 | −6.50 <> 7.97 | −3.60 <> 3.11 | −7.74 <> 8.72 |
| ECAS-B to ECAS-C | ± 6.44 | ± 3.19 | ±7.18 | −6.63 <> 6.25 | −2.68 <> 3.70 | −6.86 <> 7.51 |
| ECAS-A to ECAS-C | ± 6.02 | ± 2.97 | ± 6.64 | −5.47 <> 6.56 | −2.70 <> 3.24 | −5.82 <> 7.46 |
| Recommended for clinical use | ≥ 8 | ≥ 4 | ≥ 9 |  |  |  |

ΔX is the change in score required to be considered significant. The Chelune method results in different upper and lower thresholds due to its subtraction of a constant.

Table 4. Multiple regression models to predict ECAS performance.

| | $R^2$ | SEE | C | $\beta_{(ECAS-A)}$ | $\beta_{(Education)}$ | $X - \hat{X}$ | |
|---|---|---|---|---|---|---|---|
| **Predicting ECAS-B from ECAS-A** | | | | | | | |
| ALS Specific | .451 | 5.01 | 28.82 | .573 | .587 | ± 8.24 | |
| ALS Non-Specific | .297 | 2.25 | 15.31 | .490 | – | ± 3.70 | |
| ECAS Total | .483 | 5.97 | 37.94 | .599 | .628 | ± 9.82 | |
| | $R^2$ | SEE | C | $\beta_{(ECAS-B)}$ | $\beta_{(Education)}$ | $X - \hat{X}$ | |
| **Predicting ECAS-C from ECAS-B** | | | | | | | |
| ALS Specific | .700 | 5.30 | 34.18 | .938 | – | ± 8.72 | |
| ALS Non-Specific | .431 | 1.87 | 12.69 | .472 | .243 | ± 3.08 | |
| ECAS Total | .741 | 6.12 | 14.78 | .872 | – | ± 10.07 | |
| | $R^2$ | SEE | C | $\beta_{(ECAS-A)}$ | $\beta_{(ECAS-B)}$ | $X - \hat{X}$ | |
| **Predicting ECAS-C from ECAS-A AND ECAS-B** | | | | | | | |
| ALS Specific | .744 | 4.81 | −1.28 | .583 | .432 | ± 7.91 | |
| ALS Non-Specific | .762 | 1.51 | 7.37 | .473 | .289 | ± 2.48 | |
| ECAS Total | .815 | 5.11 | 2.74 | .563 | .416 | ± 8.41 | |
| | $R^2$ | SEE | C | $\beta_{(ECAS-A)}$ | $\beta_{(ECAS-B)}$ | $\beta_{(Age)}$ | $X - \hat{X}$ |
| **ECAS-C performance from ECAS-A, ECAS-B, and age** | | | | | | | |
| ALS Non-Specific | .665 | 1.42 | 10.91 | .473 | .269 | −.044 | ± 2.34 |

$R^2$ is the multiple $R^2$ when model contains one predictor and adjusted $R^2$ when model contains more than one predictor. *SEE* is the residual standard error, C is the intercept, $\beta$ is the beta coefficient associated with the subscript, $X - \hat{X}$ is the residual (i.e. the difference between the model predicted score and the observed score). The $X - \hat{X}$ column indicates the number of points difference required between observed and estimated score to determine reliable difference – this is calculated as $1.645^*(SEE)$.

B performance, and no significant effect of sex was observed. For the ECAS-C, education level again significantly correlated with ALS Specific ($r = .40$, $p = 0.002$), ALS Non- Specific ($r = .31$, $p = 0.02$), and ECAS Total ($r = .40$, $p = 0.002$) scores. The ALS Non-Specific functions of ECAS-C significantly correlated with age ($r = -.32$, $p = 0.02$) and test-retest interval ($r = -.30$, $p = 0.03$). Additionally, a marginally significant Mann-Whitney $U$-test was observed for sex and ALS Specific functions of ECAS-C ($W = 207$, $p = 0.047$). While these variables are retained for the regression models, the significant correlations with age, sex, and test-retest interval do not survive Holm correction for multiple comparisons (all $p > 0.05$).

Significant variables were entered into regression models in a single block and individual variables were only retained once their contribution to the model remained significant. For the multiple regression models, the variance inflation factor for the predictors did not exceed 2. Table 4 displays the results of these models. For the prediction of ECAS-B, education is retained in the model for ALS Specific and ECAS Total scores. For the ECAS-C, education significantly added to the model for ALS Non-Specific scores. No other variables were retained in the final multiple regression models. Additional models were generated to predict ECAS-C from the combined performance on ECAS-A and ECAS-B. As with the previous multiple regression models, variables of interest were correlated with, and regressed onto, the ECAS-C. However, in this instance, the only variable retained is that of age on ALS Non-Specific.

*Example data*

A 62-years-old male limb-onset ALS patient with 10.5 years of education was assessed at two time-points, with a four-month interval between Time 1 and Time 2. The patient had no behavioural or respiratory symptoms at either time-point. For the ECAS total, the patient scored 108 on the ECAS-A, and 96 on the ECAS-B. This resulted in change scores of −12 which falls below the RCI threshold for significant decline by both the JT method and the Chelune method, and the recommended clinical thresholds (Table 2).

Using the simple regression-based method, this patient's predicted ECAS-B score is calculated using the equation $\hat{X} = \beta X + C$, where in this case $\hat{X} = (.648)(108) + 41.62$. The resulting predicted ECAS-B score is therefore 111.60 with a residual (i.e. ECAS-B minus predicted ECAS-B) of −15.2. These values are the entered into the equation $\frac{(X-\hat{X})}{SEE} = \frac{(96-111.60)}{6.23} = -2.44$ which is less than −1.645). Therefore the patient's score was significantly lower than predicted. Furthermore, the residual of −15.2 falls below the simple regression-based threshold of ±10.24. The multiple regression-based method includes the variable education, here 10.5 years, with the equation

$\hat{X} = \beta_A X_A + \beta_{Ed} X_{Ed} + C$. The patient's predicted score is then $\hat{X} = (.599)(108) + (.628)(10.5) + 37.94$ which results in a predicted ECAS-B score of 109.23. Using the same equation as above, $\frac{(X-\hat{X})}{SEE} = \frac{(96 - 109.23)}{5.97} = -2.22$. Again, this falls outside of $\pm1.645$, and the residual ($-13.23$) is less than the threshold of $\pm9.82$. Therefore, under all measures, the patient presents with a significant and reliable decrease in cognitive functioning.

## Discussion

The monitoring of cognitive and behavioural symptoms longitudinally in ALS is integral to measurement of progression of disease, outcome of clinical trials and in providing person-centred care. While the recent development of alternate forms of the Edinburgh Cognitive and Behavioural ALS Screen (ECAS) provided tools necessary to assess neuropsychological functioning over time, the present paper aimed to provide data necessary for individual-level interpretation. Moreover, thresholds for significant decline or improvement provide viable end-points for clinical trials. An additional goal was to explore the test-retest reliability of the ECAS's alternate forms when testing across clinically relevant intervals. The present results demonstrate that the alternate versions of the ECAS provide a consistent method by which cognitive functioning can be monitored over time in patients with ALS. Building from the study by Crockford et al. (21), the present study aimed to explore whether the alternate forms of the ECAS ameliorate practice effects when administered over clinically-meaningful testing intervals. No significant practice effects were observed for the ECAS-A, ECAS-B, and ECAS-C when administered sequentially. The alternate forms of the ECAS are successful in ameliorating practice, thus confirming the findings of the authors (21).

The test-retest reliability of the alternate forms was additionally explored. Intraclass correlation coefficients (ICCs) were excellent for the composite ALS Specific, ALS Non-Specific, and ECAS Total scores. The individual cognitive subtests of ECAS performed well also, achieving ICC values greater than .70. While the ICC values for the visuospatial task appear quite low, this is in part due to the dependency of the ICC calculation on between-subjects variability. Because there is very little between-subject variability in the visuospatial task, the small differences present are exaggerated suggesting a smaller test-retest reliability than is warranted (27). However, as noted, the composite ALS Specific, ALS Non-Specific, and ECAS Total reliability was excellent. This is particularly pertinent given the sensitivity of these domains in detecting cognitive impairment against a full neuropsychological battery (13). As such, participants who were administered the ECAS forms serially showed good consistency and stability across testing sessions.

The primary aim of this paper was to establish methods of interpreting reliable change for patients with ALS. Four models were utilized, including two reliable change indices (RCI) and two regression based (RB) methods. These methods account for slightly different factors that may influence performance change. The RCI thresholds for reliable change are the minimum increase or decrease in performance necessary to be considered reliable. Both RCI methods produced similar thresholds for all comparisons (i.e. ECAS-A to ECAS-B, ECAS-B to ECAS-C, and ECAS-A to ECAS-C).

The RCI methods proposed by Jacobson and Truax (25) and modified by Chelune et al. (26) were developed on the assumption of repeated assessment using the same version of a test. For instance, the $SE_m$ in these authors' studies is calculated using the standard deviation and test-retest reliability of the same instrument. This does not pose an issue when one considers that the test-retest reliability used in the present study is the intraclass correlations between two ECAS forms. However, RCI calculations traditionally use the standard deviation of the instrument assuming equality of variation for Time 1 and Time 2. Fortunately, the standard deviations for ALS Specific, ALS Non-Specific, and ECAS total across alternate forms were only trivially different (e.g. for ECAS Total scores, the standard deviations were 11.26, 11.38, and 11.87 for the ECAS-A, ECAS-B, and ECAS-C, respectively). It was not deemed necessary to use a measure of shared variance in place of standard deviation.

With regard to the regression-based methods, linear regression models provide predictive scores for patients based on their baseline ECAS performance, or a combination of this and demographic variables. The deviation from a patient's predicted score and their actual score is used to determine whether a deviation constitutes a reliable difference. By dividing the residual by the standard error of the residual, one can determine if an individual's departure from their predicted performance is within normal variation, i.e. variation due to measurement error, practice effects, or regression to the mean. While regression-based methods may be more complicated to calculate, they may also provide more accurate predictions that take account of important moderating variables such as education level. However, some authors have argued that regression-based methods are not necessarily superior to RCI methods (e.g. (19)). Additionally, regression-based methods do have their own limitations. As noted by Crawford and Garthwaite (24), the error associated with predicting follow-up performance from baseline performance using regression based techniques will be larger at the extremes, i.e. the residuals at the extremes are greater. Therefore, caution should be paid to interpreting

change of patients who score poorly at baseline. Because the sample herein is of healthy controls while the target patient population would be expected to score toward the lower extremes, further research would be needed to clarify whether RCI and regression-based thresholds need to be developed based on initial test performance. This may be achieved by exploring these thresholds in a sample of MND patients where cognitive deterioration is not expected or found, for example in patients who possess a slower disease progression.

An important caveat in utilizing these thresholds is that the ECAS is a cognitive screening tool, and not designed to replace full neuropsychological assessment. While a patient's test-retest performance may be reliably described as a decline using the thresholds herein, such findings should be corroborated with specialist neuropsychological input.

In deciding which method to utilize for detecting reliable change, a pragmatic approach is recommended. For research purposes, the multiple regression-based methods may provide more specific indicators of change. However, these regression-based methods are relatively more technical and complex to calculate. The ECAS was designed to be accessible to non-specialist health care professionals, and thus, the recommendation of regression-based methods may compromise the clinical utility of the ECAS. Given the similarity in scores across all four methods and the ease with which RCI methods can be included in a clinical environment, a conservative application of change scores is recommended for clinical purposes. Based on the most conservative JT method, and to reduce the number of false-positives, a change of $\geq 8$, $\geq 4$, or $\geq 9$ points is recommended for a significant change in ALS Specific, ALS Non-Specific, or ECAS Total score, respectively.

## Conclusions

Measuring the progression of cognitive symptoms in ALS has important clinical implications. Cognitive status can play an important role in how patients engage with interventions, in how clinicians engage with patients, and in what services may be appropriate. The alternate forms of the ECAS provide a method by which cognitive symptoms can be monitored over time. The present study built on this by providing a means by which a patient's change over time can be reliably interpreted. Four models of change indices were calculated. The reliable change indices may be the method with the highest clinical utility; however, regression based methods may play a role in more detailed analysis or clinical research. Additionally, the present study demonstrated that the test-retest reliability of the ECAS and its alternate forms is excellent for the ALS Specific, ALS Non-Specific, and ECAS Total scores. This,

along with no evidence of significant practice effects, suggests that the ECAS-A-B-C are stable, consistent, and useful in monitoring ALS patients' cognitive performance over time.

## ORCID

Christopher Crockford    http://orcid.org/0000-0001-8829-467X
Ammar Al-Chalabi    http://orcid.org/0000-0002-4924-7712
Orla Hardiman    http://orcid.org/0000-0003-2610-1291

## References

1. Abrahams S, Leigh PN, Harvey A, Vythelingum GN, Grise D, Goldstein LH. Verbal fluency and executive dysfunction in amyotrophic lateral sclerosis (ALS). Neuropsychologia. 2000;38:734–47.
2. Goldstein LH, Abrahams S. Changes in cognition and behaviour in amyotrophic lateral sclerosis: nature of impairment and implications for assessment. Lancet Neurol. 2013;12:368–80.
3. Taylor LJ, Brown RG, Tsermentseli S, Al-Chalabi A, Shaw CE, Ellis CM, et al. Is language impairment more common

than executive dysfunction in amyotrophic lateral sclerosis? J Neurol Neurosurg Psychiatry. 2013;84:494–8.

4. Mioshi E, Caga J, Lillo P, Hsieh S, Ramsey E, Devenney E, et al. Neuropsychiatric changes precede classic motor symptoms in ALS and do not affect survival. Neurology. 2014;82:149–55.

5. Caga J, Turner MR, Hsieh S, Ahmed RM, Devenney E, Ramsey E, et al. Apathy is associated with poor prognosis in amyotrophic lateral sclerosis. Eur J Neurol. 2016;23:891–7.

6. Elamin M, Phukan J, Bede P, Jordan N, Byrne S, Pender N, et al. Executive dysfunction is a negative prognostic indicator in patients with ALS without dementia. Neurology. 2011;76:1263–9.

7. Chio A, Vignola A, Mastro E, Giudici AD, Iazzolino B, Calvo A, et al. Neurobehavioral symptoms in ALS are negatively related to caregivers' burden and quality of life. Eur J Neurol. 2010;17:1298–303.

8. Olney RK, Murphy J, Forshew DBSN, Garwood E, Miller BL, Langmore S, et al. The effects of executive and behavioral dysfunction on the course of ALS. Neurology. 2005;65:1774–7.

9. Martin NH, Landau S, Janssen A, Lyall R, Higginson I, Burman R, et al. Psychological as well as illness factors influence acceptance of non-invasive ventilation (NIV) and gastrostomy in amyotrophic lateral sclerosis (ALS): a prospective population study. Amyotroph Lateral Scler Frontotemporal Degener. 2014;15:376–87.

10. Abrahams S, Newton J, Niven E, Foley J, Bak TH. Screening for cognition and behaviour changes in ALS. Amyotroph Lateral Scler Frontotemporal Degener. 2014;15:9–14.

11. Loose M, Burkhardt C, Aho-Özhan H, Keller J, Abdulla S, Böhm S, et al. Age and education-matched cut-off scores for the revised German/Swiss-German version of ECAS. Amyotroph Lateral Scler Frontotemporal Degener. 2016;17:374–6.

12. Lulé D, Burkhardt C, Abdulla S, Böhm S, Kollewe K, Uttner I, et al. The Edinburgh cognitive and behavioural amyotrophic lateral sclerosis screen: a cross-sectional comparison of established screening tools in a German-Swiss population. Amyotroph Lateral Scler Frontotemporal Degener. 2015;16:16–23.

13. Niven E, Newton J, Foley J, Colville S, Swingler R, Chandran S, et al. Validation of the Edinburgh cognitive and behavioural amyotrophic lateral sclerosis screen (ECAS): a cognitive tool for motor disorders. Amyotroph Lateral Scler Frontotemporal Degener. 2015;16:172–9.

14. Pinto-Grau M, Burke T, Lonergan K, McHugh C, Mays I, Madden C, et al. Screening for cognitive dysfunction in ALS: validation of the Edinburgh cognitive and behavioural ALS screen (ECAS) using age and education adjusted normative data. Amyotroph Lateral Scler Frontotemporal Degener. 2016;18:99–106.

15. Poletti B, Solca F, Carelli L, Madotto F, Lafronza A, Faini A, et al. The validation of the Italian Edinburgh cognitive and behavioural ALS screen (ECAS). Amyotroph Lateral Scler Frontotemporal Degener. 2016;17:489–98.

16. Ye S, Ji Y, Li C, He J, Liu X, Fan D. The Edinburgh cognitive and behavioural ALS screen in a Chinese amyotrophic lateral sclerosis population. PLoS One. 2016;11:e0155496.

17. Elamin M, Bede P, Byrne S, Jordan N, Gallagher L, Wynne B, et al. Cognitive changes predict functional decline in ALS a population-based longitudinal study. Neurology. 2013;80:1590–7.

18. Newsom-Davis IC, Lyall RA, Leigh PN, Moxham J, Goldstein LH. The effect of non-invasive positive pressure ventilation (NIPPV) on cognitive function in amyotrophic lateral sclerosis (ALS): a prospective study. J Neurol Neurosurg Psychiatry. 2001;71:482–7.

19. Duff K. Evidence-based indicators of neuropsychological change in the individual patient: relevant concepts and methods. Arch Clin Neuropsychol. 2012;27:248–61.

20. Burkhardt C, Neuwirth C, Weber M. Longitudinal assessment of the Edinburgh cognitive and behavioural amyotrophic lateral sclerosis screen (ECAS): lack of practice effect in ALS patients? Amyotroph Lateral Scler Frontotemporal Degener. 2017;18:202–9.

21. Crockford C, Kleynhans M, Wilton E, Radakovic R, Newton J, Niven E, et al. ECAS A-B-C: alternate forms of the Edinburgh cognitive and behavioural ALS screen. Amyotroph Lateral Scler Frontotemporal Degener. 2017. doi: 10.1080/21678421.2017.1407793

22. Temkin NR, Heaton RK, Grant I, Dikmen SS. Detecting significant change in neuropsychological test performance: a comparison of four models. J Inter Neuropsych Soc. 1999;5:357–69.

23. Stein J, Luppa M, Brähler E, König HH, Riedel-Heller SG. The assessment of changes in cognitive functioning: reliable change indices for neuropsychological instruments in the elderly–a systematic review. Dement Geriatr Cogn Disord. 2010;29:275–86.

24. Crawford JR, Garthwaite PH. Comparing patients' predicted test scores from a regression equation with their obtained scores: a significance test and point estimate of abnormality with accompanying confidence limits. Neuropsychology. 2006;20:259.

25. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59:12.

26. Chelune GJ, Naugle RI, Lüders H, Sedlak J, Awad IA. Individual change after epilepsy surgery: practice effects and base-rate information. Neuropsychology. 1993;7:41.

27. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Condition Res. 2005;19:231–40.

**Supplementary material available online**