Diagnostic criteria in amyotrophic lateral sclerosis
A multicenter prospective study

ABSTRACT

Objective: To assess the sensitivity and specificity of the Awaji and revised El Escorial diagnostic criteria (rEEC) in amyotrophic lateral sclerosis (ALS).

Methods: We conducted a large prospective multicenter study, recruiting 416 patients (253 male, 163 female) between January 1, 2012, and August 31, 2015, to compare the diagnostic accuracy of Awaji and rEEC in accordance with standards of reporting of diagnostic accuracy criteria.

Results: The sensitivity of the Awaji criteria (57%, 50.0%–63.3%) was higher when compared to rEEC (45%, 38.7%–51.7%, \( p < 0.001 \)), translating to a 12% gain in sensitivity. The specificity of both criteria were identical, 99.5%, indicating the number needed to test in order to diagnose one extra case of ALS was 1.8 (1.5–2) for Awaji criteria and 2.4 (2–2.6) for rEEC. The Awaji criteria exhibited a higher sensitivity across subgroups, including bulbar (\( p < 0.001 \)) and limb-onset (\( p < 0.001 \)) patients. The inclusion of the possible diagnostic category as a positive finding enhanced sensitivity of the Awaji criteria and rEEC, particularly in early stages of ALS, while maintaining specificity.

Conclusion: The present study established a higher sensitivity of Awaji criteria when compared to rEEC, with diagnostic benefits evident in bulbar and limb-onset disease. Inclusion of possible as a positive finding enhanced sensitivity of both criteria, while maintaining specificity, and should be considered in clinical practice and future therapeutic trials.

Classification of evidence: This study provides Class IV evidence that the Awaji criteria have a higher sensitivity and the same specificity as the rEEC in identifying patients with ALS. Neurology® 2016;87:1–7

GLOSSARY

ALS = amyotrophic lateral sclerosis; ALSFRS-R = revised Amyotrophic Lateral Sclerosis Functional Rating Score; CI = confidence interval; EEC = El Escorial criteria; LMN = lower motor neuron; MRC = Medical Research Council; NNT = number needed to test; rEEC = revised El Escorial criteria; STARD = Standards of Reporting of Diagnostic Accuracy; TMS = transcranial magnetic stimulation; UMN = upper motor neuron.

In the absence of a pathognomonic test, the diagnosis of amyotrophic lateral sclerosis (ALS) remains phenotypically based, relying on identification of upper motor neuron (UMN) and lower motor neuron (LMN) signs.1–4 The clinically based El Escorial criteria (EEC)1 mandated 4 levels of diagnostic certainty, which were dependent on the extent of UMN and LMN dysfunction. In order to improve diagnostic sensitivity, the EEC were revised (rEEC), incorporating a clinically probable–laboratory-supported diagnostic category.2 Although specific, the sensitivity remained an issue, particularly in the early stages of ALS, resulting in diagnostic delays and recruitment into therapeutic trials.5–8

In an attempt to further reduce diagnostic delays, neurophysiologically based Awaji criteria were developed.3 These criteria proposed that neurophysiologic features of LMN dysfunction were equivalent to LMN signs. The assessment of UMN dysfunction, however, remained clinically based. When compared to rEEC, the Awaji criteria exhibited increased sensitivity,9–17 although lower sensitivities were reported by some studies,18,19 a finding attributed to omission...
of the probable–laboratory-supported diagnostic category. Importantly, the diagnostic benefit appeared to be most prominent in bulbar-onset ALS.\textsuperscript{13,19} In addition, while ALS diagnostic criteria are regarded to be specific, there is a paucity of specificity data,\textsuperscript{13,19} preventing objective conclusions about the reliability of the ALS diagnostic criteria. Consequently, the aim of the current prospective multicenter study was to assess the diagnostic utility of the ALS diagnostic criteria in a clinical setting, undertaken according to the Standards of Reporting of Diagnostic Accuracy (STARD) criteria.

**METHODS** The study was designed to provide Class IV evidence that the Awaji criteria have a higher sensitivity and the same specificity as the rEEC in identifying patients with ALS. Between January 1, 2012, and August 31, 2015, we prospectively and consecutively enrolled 416 patients with suspected ALS from 4 large ALS centers on the east coast of Australia in keeping with the STARD criteria. The patients were referred from a number of sources including general practitioners, neurologists, and non-neurologists. Inclusion criteria were as follows: (1) suspected diagnosis of ALS by the referring physician; or (2) neuromuscular disorder defined as muscle weakness and wasting for at least 6 months in at least one body region, irrespective of whether sensory symptoms were present. At time of assessment, the assessor was blinded to the eventual diagnosis, whether the diagnosis of ALS or a mimic disorder. It should be stressed that at the time of application of the diagnostic criteria other disorders that could mimic ALS were not excluded.

Exclusion criteria included (1) pure UMN syndrome in which laboratory and neuroimaging studies suggested a diagnosis other than ALS, such as hereditary spastic paraplegia or progressive forms of multiple sclerosis; (2) diagnosis of a non-neuromuscular neurologic disorder to explain the patient’s symptoms, such as cerebellar or extrapyramidal syndromes; (3) non-neurologic disorders causing the symptoms; (4) inability or refusal to provide informed consent. All patients provided written informed consent to the procedures approved by the Western Sydney Local Health District and South East Sydney Area Health Service Human Research Ethics Committees.

Once recruited, all patients underwent a clinical assessment and detailed investigations to diagnose mimic disorders, with the laboratory investigations tailored for the clinical presentation. The following investigations were performed on all patients: routine biochemistry, hematology, vasculitic screen (antinuclear antibodies, extractable nuclear antigen, antineutrophil cytoplasmic antibodies), immunoelectrophoresis, angiotensin converting enzyme levels, metabolic screen (vitamin B\textsubscript{12}, folate, B\textsubscript{6}, thyroid function tests), infective screen (HIV, syphilis serology, human T-cell lymphotropic virus I and II), and antiganglioside antibodies (GM1, anti-MAG). Genetic testing for Kennedy disease was performed in all male patients, while oculocephalomegal muscular dystrophy and spinal muscular atrophy genetic testing was performed on selected patients. In addition, voltage-gated K\textsuperscript{+} channels, acetylcholine receptor, and muscle-specific tyrosine kinase antibodies were performed on patients with a suspicious clinical presentation for the respective disorders. In addition, MRI of brain and spinal cord was undertaken to exclude structural lesion in all participants.

Patients were clinically staged using the revised Amyotrophic Lateral Sclerosis Functional Rating Scale score (ALSFRS-R).\textsuperscript{20} The disease duration (months) from time of symptom onset and site of disease onset were recorded. Muscle strength was assessed by the Medical Research Council (MRC) score, with the following muscle groups assessed: shoulder abduction, elbow flexion and extension, wrist dorsiflexion, finger abduction and thumb abduction, hip flexion, knee extension, and ankle dorsiflexion bilaterally, yielding a maximal score of 90. UMN function was assessed and graded by a dedicated UMN score, incorporating the following parameters: jaw jerk, facial reflex, upper and lower limb deep tendon reflexes and plantar responses with the score ranging from 0 (no UMN dysfunction) to 16 (severe UMN dysfunction).\textsuperscript{21}

Nerve conduction study and needle EMG was performed according to established techniques by experienced neurophysiologists.\textsuperscript{22} Needle EMG assessments were standardized across the 4 sites to ensure that enough regions were investigated in every patient. Specifically, at least 3 body regions were assessed in every patient (bulbar, cervical, thoracic paraspinals, and lumbosacral), and importantly, the assessment of the bulbar region was mandated in every patient. A minimum of one muscle was assessed in the bulbar region and included either the trapezius or genioglossus muscles. If the first bulbar muscle was normal (always the trapezius muscle) then the genioglossus muscle was assessed. The median number of muscles studied per patient was 12 (interquartile range 7–14). Evidence of ongoing activity (positive sharp waves and fibrillation potentials), fasciculations, and chronic neurogenic changes were recorded. The assessors, namely physicians performing EMG testing at each site, were as follows: Westmead Hospital (C.Y.), Brain and Mind Center (M.C.K.), Royal North Shore Hospital (C.Y.), Royal Brisbane and Women’s Hospital (R.D.H.). Two physicians (N.G. and P.M.) graded the sensitivity of the rEEC\textsuperscript{2} and Awaji criteria\textsuperscript{3} independently, and disagreements were resolved by consensus. At time of initial assessment, the authors were blinded to the eventual diagnosis. Statistical analysis was performed by a separate rater (S.V.). The reference standard for ALS diagnosis related to good clinical practice that incorporated disease progression deemed consistent with ALS and exclusion of mimicking disorders on neurophysiologic, laboratory, and neuroimaging techniques.

**Statistics.** The primary outcome measure was the diagnostic utility (sensitivity and specificity) of the consensus criteria (rEEC vs Awaji) in differentiating ALS from non-ALS mimic disorders. The secondary outcome measures included the diagnostic utility of the criteria in ALS subgroups, defined by site of disease onset (bulbar vs limb), functional disability (ALSFRS-R ≥ 38 less severe; <38 more severe disease), and disease duration at time of assessment (<12 months early; ≥12 months longer disease). The cutoff values were generated by consensus and reflected early stages of the disease. The sensitivity, specificity, and number needed to test (NNT) were determined for each criteria set (rEEC and Awaji criteria). Pearson χ\textsuperscript{2} test or McNemar test were utilized to assess differences between the criteria. A probability value <0.05 was considered statistically significant. Results were expressed as median (interquartile range).

**RESULTS** Clinical features. Between January 1, 2012, and August 31, 2015, we enrolled 416 patients (253 men, 163 women, median age 61 years [range 49–69]) with suspected ALS from 4 neuromuscular clinics on the east coast of Australia who met the inclusion criteria (figure 1). The breakdown of
Patient recruitment at each center was as follows: Westmead Hospital, 143; Brain and Mind Center, 112; Royal North Shore Hospital, 90; Royal Brisbane and Women’s Hospital, 71. After a detailed clinical, laboratory, neurophysiologic, and radiologic assessment, as well as follow-up for at least 6 months, 183 patients were eventually diagnosed with ALS (142 men, 91 women, median age 62 years [range 52–69]), while 212 patients were diagnosed with an ALS mimic disorder (table 1).

The degree of functional impairment in the patients with ALS, as indicated by the ALSFRS-R score, was mild (table 1). Muscle strength, as measured by the MRC score, was comparable between the groups, as was the median age at time of assessment (table 1). In contrast, UMN signs were more prominent in patients with ALS, while the disease duration was significantly shorter in ALS (table 1). During the course of the study, 42 (18%) patients with ALS died, with median survival from symptom onset being 28 months (range 20–39). At last follow-up, the median disease duration in the surviving patients was 35 months (range 26–53.5).

Comparison of diagnostic criteria. The sensitivity of the Awaji criteria, as defined by the proportion of patients categorized as definite or probable ALS, at time of initial assessment was 57% (95% confidence interval [CI] 50.0%–63.3%) and was higher when compared to rEEC (table 2). The specificity for the Awaji and rEEC were 99.5% (95% CI 96%–100%), with the NNT in order to diagnose one extra case of ALS, in a population composed of neuromuscular diseases,
The Awaji criteria exhibited a significantly higher sensitivity when compared to the revised El Escorial criteria (rEEC) in the entire cohort of amyotrophic lateral sclerosis (ALS) as well as patients with bulbar and limb-onset disease, with the number needed to test (NNT) being lower. The specificity of both criteria were identical at 99.5%. Inclusion of the ALS possible diagnostic category as a positive finding (sensitivity [def/prob]) significantly enhanced the sensitivity of both the Awaji and rEEC in ALS and subgroups as follows.

Table 2 Diagnostic accuracy of ALS criteria

<table>
<thead>
<tr>
<th>Category</th>
<th>Awaji criteria</th>
<th>Revised El Escorial criteria</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(95% confidence interval)</td>
<td>(95% confidence interval)</td>
<td></td>
</tr>
<tr>
<td>Total ALS population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (def/prob)</td>
<td>57 (50.0–63.3)</td>
<td>45 (38.7–51.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>NNT</td>
<td>1.8 (1.5–2)</td>
<td>2.4 (2–2.6)</td>
<td></td>
</tr>
<tr>
<td>Sensitivity (+ possible)</td>
<td>83.4 (78.0–87.9)</td>
<td>79.6 (73.9–84.5)</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>Bulbar-onset ALS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (def/prob)</td>
<td>52.2 (39.8–64.4)</td>
<td>42 (30.2–54.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>NNT</td>
<td>1.9 (1.6–2.5)</td>
<td>2.5 (1.9–3.4)</td>
<td></td>
</tr>
<tr>
<td>Sensitivity (+ possible)</td>
<td>75.8 (63.6–85.5)</td>
<td>67.7 (55.2–78.5)</td>
<td></td>
</tr>
<tr>
<td><strong>ALSFRS-R</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (def/prob) (&gt;38)</td>
<td>49.1 (41.1–57.1)</td>
<td>36.8 (29.4–44.7)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Sensitivity (def/prob) (&lt;38)</td>
<td>77.3 (65.3–86.7)</td>
<td>67.6 (55–78.8)</td>
<td>0.15</td>
</tr>
<tr>
<td><strong>ALSFRS-R</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (+ possible) (&gt;38)</td>
<td>78.9 (71.8–84.9)</td>
<td>71.6 (64–78.4)</td>
<td>0.13</td>
</tr>
<tr>
<td>Sensitivity (+ possible) (&lt;38)</td>
<td>90.9 (81.3–96.6)</td>
<td>86.2 (75.3–93.5)</td>
<td>0.39</td>
</tr>
<tr>
<td><strong>Disease duration, mo</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (def/prob) (&lt;12 mo)</td>
<td>62.4 (53.6–70.7)</td>
<td>47 (38.2–55.9)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Sensitivity (def/prob) (&gt;12 mo)</td>
<td>48.9 (38.1–59.8)</td>
<td>41.4 (30.9–52.4)</td>
<td>0.32</td>
</tr>
<tr>
<td><strong>Disease duration, mo</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (+ possible) (&gt;12 mo)</td>
<td>84.2 (76.9–90)</td>
<td>79.7 (71.7–79)</td>
<td>0.34</td>
</tr>
<tr>
<td>Sensitivity (+ possible) (&gt;12 mo)</td>
<td>78.9 (68.7–86.6)</td>
<td>75 (64.6–83.8)</td>
<td>0.66</td>
</tr>
</tbody>
</table>

ALSFRS-R = Amyotrophic Lateral Functional Rating Scale-Revised.

The Awaji criteria exhibited a significantly higher sensitivity when compared to the revised El Escorial criteria (rEEC) in the entire cohort of amyotrophic lateral sclerosis (ALS) as well as patients with bulbar and limb-onset disease, with the number needed to test (NNT) being lower. The specificity of both criteria were identical at 99.5%. Inclusion of the ALS possible diagnostic category as a positive finding (sensitivity [+] possible) significantly enhanced the sensitivity of both the Awaji and rEEC in ALS and subgroups as follows.

* p < 0.001 (Awaji [definite/probable] vs Awaji [add possible]).
* p < 0.001 (rEEC [definite/probable] vs rEEC [add possible]).
* p < 0.001 (Bulbar and limb-onset ALS Awaji [definite/probable] vs Awaji [add possible]).
* p < 0.01 (Bulbar and limb-onset ALS, rEEC [definite/probable] vs rEEC [add possible]).
* p < 0.05 (Awaji and rEEC sensitivity higher in patients with worse functional deficits).
* p < 0.001 (Addition of possible category increased sensitivity for Awaji and rEEC in patients with shorter and longer disease duration).

Comparison of diagnostic utility in ALS subgroups. Subgroup analysis, based on site of disease onset, disclosed that the sensitivity of the Awaji criteria was higher for bulbar onset disease when compared to the rEEC (table 2). Importantly, the number of patients with bulbar onset ALS needed to be tested in order to diagnose one extra case of ALS was 1.9 (1.6–2.5) by the Awaji criteria and 2.5 (1.9–3.4) by the rEEC. In addition, the sensitivity of the Awaji criteria were higher in limb-onset patients when compared to the rEEC (table 2), with the NNT for Awaji criteria being 1.7 (1.5–2) and for rEEC 2.3 (2–3.3).

being 1.8 (1.5–2) for the Awaji criteria and 2.4 (2–2.6) for the rEEC.

The diagnostic utility of the Awaji and rEEC were further assessed by considering the possible diagnostic category as a positive finding. The sensitivity of the Awaji and rEEC was increased, although the sensitivity remained higher for the Awaji criteria (table 2). Importantly, the specificity for both the Awaji and rEEC remained unchanged (99.5% [95% CI 96%–100%]). As expected, the NNT in order to diagnose an extra ALS case was reduced to 1.2 (1.1–1.3) for the Awaji criteria and 1.3 (1.2–1.4) for the rEEC.
The inclusion of the possible diagnostic category as a positive finding significantly enhanced the sensitivity of both criteria while the specificity remained unchanged. Specifically, in bulbar-onset ALS, the sensitivity of both the Awaji criteria and rEEC were enhanced (table 2). In limb-onset ALS, a more impressive increase in sensitivity was evident for both the Awaji criteria and rEEC (table 2).

The degree of functional disability, as indicated by ALSFRS-R score, also influenced the sensitivity of both the Awaji criteria and rEEC. Specifically, the sensitivity of the Awaji criteria in patients with greater functional disability (ALSFRS-R <38) was 77.3% (65.3%–86.7%) and was higher when compared to patients with less severe disease (table 2). In addition, the rEEC criteria also exhibited a higher sensitivity in patients with more severe disease (67.6%, 55%–78.8%) compared to patients with less functional impairment (36.8%, 29.4%–44.7%, p < 0.05). The Awaji criteria, however, were more sensitive when compared to rEEC in less severe disease (p < 0.05), while comparable sensitivities between the criteria were evident in the more severe disease subgroup (p = 0.15). The diagnostic utility of the criteria did not change when the median ALSFRS-R value of 42 was utilized as a cutoff. Taken together, these findings suggest that while both sets of criteria are less efficient at diagnosing patients in the earlier stages of the disease process, the Awaji criteria may exhibit a greater sensitivity at the early stages. Separately, the inclusion of the possible diagnostic category as a positive finding enhanced the sensitivity of both criteria (table 2), while maintaining the specificity.

The disease duration at time of assessment influenced the sensitivity of the Awaji criteria. Namely, the sensitivity of the Awaji criteria was higher than rEEC in patients with shorter disease duration (<12 months, table 2), while the sensitivity was comparable in patients with longer disease duration at time of assessment. The sensitivity of both criteria was increased by incorporating the possible diagnostic category, and this benefit was evident in patients with shorter and longer disease duration (table 2 and figure 2). Importantly, diagnostic rate between Awaji probable/definite and possible was similar, suggesting that possible disease on Awaji is a reliable early diagnostic marker.

**DISCUSSION** The current prospective multicenter study established a higher sensitivity of the Awaji criteria when compared to rEEC, although it should be acknowledged that use of centers from a single country could have potentially limited the evaluation of the variability in the methods of EMG investigation. Importantly, the increase in sensitivity was evident in patients with bulbar and limb-onset ALS, as well as those with milder functional impairment and shorter disease duration, suggesting a greater role for the Awaji criteria in recruitment of patients with ALS into treatment trials at an early stage in the disease process.

The improvement in sensitivity of the Awaji criteria was largely due to the more comprehensive incorporation of the neurophysiologic findings, including the presence of fasciculations, which appear to be an early feature in the disease process. Of relevance, application of the Awaji criteria resulted in a 12% gain in sensitivity, which was lower than previously reported. Given the large sample size and a multidisciplinary prospective design, it seems plausible to conclude that the present findings reflect more accurately the diagnostic accuracy of the criteria, although it should be acknowledged that the criteria were not developed for the purposes of clinical diagnosis of ALS, but rather as a diagnostic algorithm for research purposes. The use of additional techniques, such as more demanding EMG protocols, encompassing a longer time for needle EMG recording from each muscle, and muscle ultrasonography could further increase the sensitivity of Awaji criteria.

The Awaji criteria have been critiqued for the omission of the probable–laboratory-supported diagnostic category, which has been suggested as a likely explanation for the lower sensitivity of Awaji criteria in limb-onset disease. Specifically, it has been reported that the Awaji criteria downgraded approximately 20% of ALS cases classified as probable–laboratory-supported...
on the rEEC. In the present study, 10 (4.3%) patients were classified as probable–laboratory-supported, of which only 2 were downgraded to the Awaji possible category. Importantly, the sensitivity of the Awaji criteria was significantly higher when compared to rEEC in limb-onset ALS, thereby arguing against an effect of omitting the probable–laboratory-supported diagnostic category from the Awaji criteria.

In an attempt to increase sensitivity of the diagnostic criteria, a number of trials have permitted the inclusion of patients with ALS classified within the possible diagnostic category.24,25 The latest revision proposed a further liberalization of the EEC, whereby possible ALS was considered a positive finding.26 While such a revision would ultimately increase the sensitivity, the issue of whether specificity is maintained remains to be determined. The present study confirms a significant increase in sensitivity by incorporating the possible diagnostic category for both Awaji criteria and rEEC, evident across the ALS subgroups. Importantly, the specificity remained unchanged, thereby suggesting that liberalization of the criteria could significantly enhance the diagnosis of ALS, which could be of particular relevance in the early stages of the disease, ultimately enabling earlier recruitment of patients into clinical trials where neuroprotective therapies are likely to be most effective.

A potential limitation of the present study pertains to inclusion of neurophysiologic testing in both the index test and reference standard. This could potentially introduce the risk of incorporation bias, since the reference standard may not have been applied independently of the index tests. Consequently, the present study would provide Class IV evidence that the Awaji criteria exhibit a higher sensitivity and comparable specificity as the rEEC in identifying ALS.

The utility of diagnostic criteria may be limited in the restricted or atypical ALS phenotypes; specifically, in clinically pure lower motor neuron phenotypes, such as flail arm syndrome or progressive muscular atrophy.26 Identification of subclinical UMN dysfunction with transcranial magnetic stimulation (TMS) techniques has been documented to significantly increase the sensitivity of the Awaji criteria and rEEC, reliably differentiating ALS from potential mimic disorders and reducing the diagnostic delay by months.29,30 Consequently, incorporation of the threshold-tracking TMS technique into future revisions of diagnostic criteria could further enhance the diagnosis of ALS.

AUTHOR CONTRIBUTIONS
Nimeshan Geesinga: conducted the study, data collection, analysis, writing of manuscript, Parvathi Menon: data collection, data analysis, editing of manuscript, Daniel B. Scherman: data collection, data analysis, editing of manuscript. Neil Simon: conceptualization of study, editing of manuscript. Con Yiannikas: conceptualization of study, editing of manuscript. Robert D. Henderson: conceptualization of study, editing of manuscript. Matthew C. Kiernan: conceptualization of study, editing of manuscript. Steve Vucic: conceptualization of study, data analysis, editing of manuscript.

STUDY FUNDING
Supported by a research grant from the Motor Neuron Disease Research Institute of Australia, National Health and Medical Research Council of Australia (project grant numbers 510233, 1024915, 1055778), and by funding to Forefront, a collaborative research group dedicated to the study of frontotemporal dementia and motor neuron disease, from a National Health and Medical Research Council of Australia Program Grant (1037746).

DISCLOSURE
The authors report no disclosures relevant to the manuscript. Go to Neurology.org for full disclosures.

Received February 3, 2016. Accepted in final form May 9, 2016.

REFERENCES