Supplementary Online Content


eAppendix. Statistical analyses and data synthesis (online supplementary material).
eFigure. QUADAS-list results.

This supplementary material has been provided by the authors to give readers additional information about their work.
eAppendix. Statistical analyses and data synthesis (online supplementary material).

We undertook three statistical analyses to assess the diagnostic performance of rEEC and Awaji criteria.

First, we compared both sets of criteria by determining the absolute risk difference and 95% confidence intervals (CI) in the proportion of patients that would be classified as probable ALS (including probable laboratory-supported by rEEC) or definite ALS. Quantitative data synthesis was accomplished through random-effects meta-analysis to incorporate variation among studies (DerSimonian and Laird method)\(^1\) using Cochrane Revman 5.1 software. Heterogeneity was assessed with the I\(^2\) test, which measures the percentage of total variation across studies due to heterogeneity.\(^2\)

Second, we extracted or derived indices of diagnostic performance (true and false positive cases and true and false negative cases) from data presented in each primary study for each set of diagnostic criteria. The data in the 2x2 contingency tables were used to calculate sensitivity and specificity for each study. We present the results graphically for both individual studies and weighted pooled results (random-effect method with CI estimates corrected for overdispersion) by plotting sensitivity and specificity estimated and 95% CI’s in both forest plots and the receiver operating characteristic (ROC) space.

We determined the AUC (Area Under the Curve) and the Q* index of the summary ROC curves. AUC is an estimate of the probability that the test correctly ranks two individuals, of which one has the disease and one does not have the disease. It provides a single number for global diagnostic performance (best possible result for a diagnostic test is 1.0). The Q* index is another statistic of diagnostic performance and represents the point in ROC space where sensitivity and specificity are equal (ideally closest to the top-left corner).

We explored heterogeneity amongst studies by examining both the forest plots and the ROC plot. We did not include study-level covariates in the analyses to assess factors that might have contributed to heterogeneity, such as threshold effect related to the type of neurophysiological evaluation protocol, because this information was incomplete and non-standardized. In addition we were concerned that in a relatively small meta-analysis this was likely to produce unreliable estimates. As expected, in most of the studies there were no reports of true negative cases. We handled this zero cell issue by adding \(\frac{1}{2}\) to the cells, which creates a “ceiling effect” in the estimate of specificity without heterogeneity. Therefore, an informal comparison between both sets of criteria was made by analyzing the results of a meta-analysis of each test separately.

Third, we compared both sets of criteria by calculating the individual and weighted pooled diagnostic odds ratio (DOR) through random-effects meta-analysis, as well as by determining the relative DOR between the two sets of criteria through regression analysis, using a weighted least squares method where weights were the inverse of variance of the log of the DOR. The DOR is a single indicator of diagnostic performance that expresses how much greater the odds of having the disease are for the people with a positive test result than for the people with a negative test result, and combines both likelihood ratios.\(^3\) It is a convenient statistic in meta-analyses of diagnostic studies and is particularly useful for comparing tests.
whenever the balance between false negative and false positive rates is not of immediate importance. This is clearly the case in our analysis since all except one study reported 0% of false positive rates with both criteria. Moreover, we found significant heterogeneity among sensitivity results in the studies analyzed (see Results section), which may, at least partially, be due to a diagnostic threshold effect related to the number of anatomical regions and muscles evaluated. However, the Spearman correlation coefficient relating sensitivity and specificity did not suggest the existence of a threshold effect ($p \geq 0.247$) for both sets of diagnostic criteria. Nevertheless, simple pooling of sensitivity and specificity data ignores threshold differences and, in the case of heterogeneity, may lead to an underestimate of a test’s performance, while the DOR is often reasonably constant regardless of the diagnostic threshold.\textsuperscript{4}

For all parameters, subgroup analyses were done according to the region of disease onset; bulbar or limb. All indices of diagnostic performance were estimated using the statistical software Meta-DiSc 1.4.\textsuperscript{5}

**References**


eFigure 1. QUADAS-list results.

QUADAS: Quality Assessment of Diagnostic Accuracy Studies