

Optimizing Diagnostic Efficiency: Trade-Off Analysis of RAVLT Trials in Discriminating Mild and Subjective Cognitive Impairment

Felix Dörr¹, Louisa Schwed¹, Nicklas Linz¹, Felix Menne¹, Alexandra König¹, Nina Possemis², Daphne ter Huurne², Inez Ramakers², Johannes Tröger¹

¹ ki elements GmbH, Saarbrücken, Germany; ² Maastricht University, Maastricht, Netherlands

Introduction

Screening for or assessment of cognitive impairment in healthcare and clinical trials always represents a trade-off between diagnostic accuracy and patient-burden in terms of assessment duration. Shorter testing procedures both decrease patient burden and cost. Time-efficient application of cognitive testing also accelerates clinical trial procedures. In the context of MCI/AD, domains of cognitive performance, such as memory, are commonly assessed by Verbal Learning Tests, such as the RAVLT. These Verbal Learning tests differ, amongst other things, in their number of trials. For the reasons mentioned above, there is a demand to find the number of trials needed for an optimal trade-off between accuracy and assessment duration. We present an evaluation of the RAVLT regarding its performance in discriminating between MCI and SCI as a function of the number of trials.

Methods

Evaluation is based on data from a clinical study administered by the Maastricht University Medical Center as part of the DeepSpA study using RAVLT recordings of 100 participants (63 female; 50 patients with MCI in total). MCI was assessed based on the regular DSM-V criteria of minor neurocognitive disorder and everyday functioning. Diagnoses were established in the panel including neuropsychologists, caregivers, and neurologists within the standard memory clinic routine; see Table 1 for demographics.

The RAVLT consists of the repetitive presentation of the same list of 15 common nouns, read out loud one-by-one by an examiner to a participant. In its most common form, it consists of five trials. Participants are asked to immediately recall those words from their memory (i.e., free recall) after each presentation of the list. Then, after a 20 minutes break, the participant is asked to recall the list again without further presentation of the list (i.e., delayed recall).

RAVLT performance was evaluated on a per-trial basis resulting in the traditional performance measure of number of correctly remembered words. Subsequently we used a differing number of RAVLT trials to differentiate between the MCI and the SCI group and thereby simulate and evaluate the performance in discriminating between these diagnostic groups.

In order to assess the diagnostic potential, we used an optimal cutoff procedure where we tried to find the optimal trade-off between sensitivity and specificity in discriminating between the diagnostic groups. With the respective cut-offs for using a different number of trials (i.e., 1 - 5 + delayed recall), participants have been classified in the two groups, SCI and MCI. By comparing the cut-off based referral decision with the initial grouping based on the clinical assessment, we derived a confusion matrix. Performance is reported as F1 Score, accuracy, AUC, sensitivity and specificity.

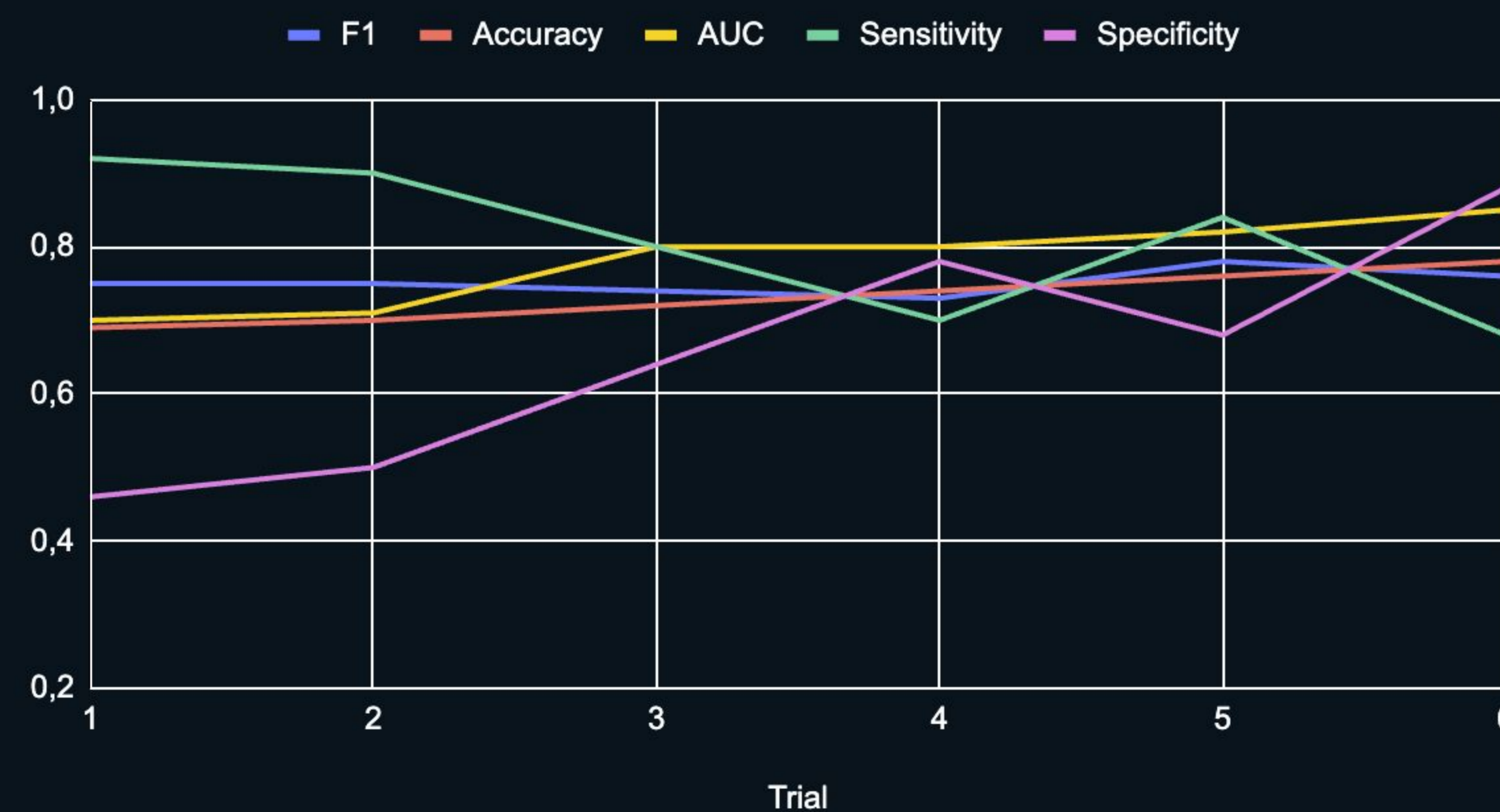


Figure 1: Performance as function of trials; Performance metrics as obtained in cut-off analysis by using the corresponding number of trials of the RAVLT.

Results

The performance metrics as a function of trials can be seen in Table 2 and Figure 1. Overall, we can see that the F1 Score stays relatively stable over time, while Accuracy and AUC keep increasing. Whereas after trial 3, the AUC does not increase substantially, we find the optimal trade-off between sensitivity and specificity in trial 4.

Conclusion

Overall, we present evidence showing that the diagnostic potential of the RAVLT does not increase linearly as a function of trials. Rather, we find that this potential increases substantially over the first four trials and then reaches a plateau. We suggest four trials as the optimal trade-off between diagnostic accuracy and assessment burden. Particularly the delayed recall, which poses a challenge especially in remote assessments due to the 20 min retention interval, does not seem to improve discrimination performance substantially. We show that both cost and patient burden can be reduced substantially by only applying four trials of the RAVLT.

Table 1: Demographics of the sample; MMSE = Mini Mental State Examination, RAVLT total = total number of correctly remembered words over 5 trials

| | SCI (N = 50) | | | MCI (N = 50) | | |
|-------------|--------------|---------------|----------|--------------|---------------|----------|
| | N | M (SD) | Range | N | M (SD) | Range |
| Sex | 50 | 37 Female | | 50 | 36 female | |
| MMSE | 50 | 28.68 (1.27) | [25, 30] | 49 | 26.92 (2.07) | [20, 30] |
| Age | 50 | 65.94 (7.80) | [52, 84] | 50 | 70.66 (8.96) | [52, 86] |
| RAVLT total | 50 | 44.24 (11.11) | [23, 67] | 50 | 31.78 (10.11) | [12, 62] |

Disclosure

FD, LS, JT, NL, FM, and AK are employed by the speech biomarker company ki:elements. JT and NL hold shares in the speech biomarker company ki:elements. The other authors have nothing to disclose.