Adjusting for Sex Improves the Screening Performance of a Digital Cognitive Assessment in an AD Memory Clinic Cohort

Submitter Johannes Tröger

Affiliation ki:elements GmbH, Saarbrücken, Germany

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

I have used the poster abstract template to develop my abstract. Yes

Methodological Issue Being Addressed This study investigates whether adjusting for sex improves the screening performance and demographic balance of a digital cognitive assessment in an AD cohort.

Introduction Sex differences in cognitive performance are a common issue across psychometric testing, clinical screening tools, and neuropsychological assessments. These differences can introduce systematic bias, particularly in one-time, cross-sectional screening instruments routinely used for participant selection in Alzheimer's disease (AD) clinical trials and dementia screening. While some cognitive assessments provide sex-specific normative data, most do not, raising the risk of unequal inclusion of men and women and potentially compromising sample representativeness.

Digital cognitive assessments—including speech-based tools such as the ki:elements Speech Biomarker for Cognition (SB-C)—may also be susceptible to sex-related performance variance depending on their task composition and underlying test paradigms. Some research indicates that verbal paradigms, and verbal memory tasks in particular, show sex effects, with women tending to outperform men. Such effects may distort screening thresholds and reduce the accuracy of cognitive classification.

Given these considerations, this study investigates whether adjusting for sex improves the screening performance and balance of a digital cognitive assessment in an AD cohort.

Methods We analyzed data from a memory clinic population enrolled in the German DESCRIBE-AD registry (N=132 (64F); mean age = 72.8 ± 7.6 years). All participants completed the ki:elements Speech Biomarker for Cognition (SB-C) digital cognitive screener in addition to a standard neuropsychological test battery and were clinically staged as either healthy controls (HC; N=86 (44F)) or cognitively impaired (CI; N=46 (20F)).

SB-C scores were derived from recorded speech samples obtained during the Semantic Verbal Fluency (SVF) and Rey Auditory Verbal Learning Test (RAVLT) tasks. Recordings were processed using ki:elements' proprietary speech-analysis pipeline incorporating automatic speech recognition and multi-level feature extraction. The primary outcome, the SB-C global cognition score,

represents an aggregate measure of overall cognitive performance and is used in AD clinical trial (pre-)screening workflows.

In addition to the global score, we generated two standardized scaled scores: (1) a demographics-adjusted score controlling for age and education, and (2) a sex-adjusted scaled score that additionally accounts for biological sex. All compared score variants of the SB-C were developed and normed on an independent cohort as part of the SB-C development process, meaning that no model fitting, threshold calibration, or norm development was performed on the DESCRIBE-AD sample itself.

Each scoring variant was evaluated for its ability to differentiate HC from CI. For each score, we identified the optimal classification cut-off (maximising F1 score), computed the area under the receiver operating characteristic curve (AUC), and derived the full confusion matrix. Analyses focused on assessing whether the sex-adjusted scaled score improved overall screening performance and yielded a more balanced distribution of classification outcomes across sexes.

Results Screening performance improved when demographic adjustments were applied to the SB-C scores. The unscaled SB-C global score achieved an ROC–AUC of 0.73. Adjusting for age and education increased AUC to 0.74, and the inclusion of sex in the scaling model further improved performance to 0.80.

Examination of score distributions showed that, on average, women outperformed men on the SB-C, even after adjusting for age and education (SB-C_w > SB-C_m; $F = 17.0$, $p < 0.001$). Relative to their clinically established cognitive status, this pattern led to systematic misestimation: men's cognitive performance tended to be underestimated, while women's performance tended to be overestimated.

Confusion matrix analyses revealed that the sex-adjusted scaled score helps to account for these imbalances. Specifically, more men were correctly classified as HC and fewer were misclassified as CI following sex adjustment, which appears to drive the observed improvement in overall screening accuracy.

Conclusion Overall, we present evidence showing that accounting for sex can improve the AUC screening performance and confusion matrix balance of a digital speech-based cognitive assessment in an AD memory-clinic cohort screening scenario. Failure to adjust for sex in a verbal neuropsychological test paradigm may lead to systematic misclassification while incorporating sex into the SB-C scaling procedure yielded more balanced classification outcomes and a marked increase in AUC. These findings highlight the importance of considering sex effects when screening with digital cognitive assessment that rely on a verbal test paradigm.

Co-Authors

Johannes Tröger[1], Elisa Mallick[1], Stefanie Köhler[2], Alexandra König[1], Stefan Teipel[2], Nicklas Linz[1]

[1] ki:elements GmbH, Saarbrücken, Germany
[2] German Center for Neurodegenerative Diseases (DZNE), Rostock, Germany

Keywords

Keywords
Alzheimer’s Disease
Screening
Clinical Trials
Digital Cognitive Assessment
Sex-specific Cognitive Norms

Guidelines I have read and understand the Poster Guidelines

Disclosures JT, EM, NL 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.