Digital Biomarkers

Validation – Research Article

Digit Biomark 2022;6:107–116
DOI: 10.1159/000526471

Received: June 1, 2022
Accepted: August 6, 2022
Published online: September 30, 2022

Validation of the Remote Automated ki:e Speech Biomarker for Cognition in Mild Cognitive Impairment: Verification and Validation following DiME V3 Framework

Johannes Tröger[a] Ebru Baykara[a] Jian Zhao[a] Daphne ter Huurne[b] Nina Possemis[b] Elisa Mallick[a] Simona Schäfer[a] Louisa Schwed[a] Mario Mina[a] Nicklas Linz[a] Inez Ramakers[b] Craig Ritchie[c]

[a]ki elements, Saarbrücken, Germany; [b]Alzheimer Center Limburg, School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands; [c]Edinburgh Dementia Prevention, Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK

Keywords
Mild cognitive impairment · Digital biomarker · Speech biomarker · Dementia · Speech analysis · Clinical trials

Abstract
Introduction: Progressive cognitive decline is the cardinal behavioral symptom in most dementia-causing diseases such as Alzheimer's disease. While most well-established measures for cognition might not fit tomorrow's decentralized remote clinical trials, digital cognitive assessments will gain importance. We present the evaluation of a novel digital speech biomarker for cognition (SB-C) following the Digital Medicine Society's V3 framework: verification, analytical validation, and clinical validation. ***Methods:*** Evaluation was done in two independent clinical samples: the Dutch DeepSpA (N = 69 subjective cognitive impairment [SCI], N = 52 mild cognitive impairment [MCI], and N = 13 dementia) and the Scottish SPeAk datasets (N = 25, healthy controls). For validation, two anchor scores were used: the Mini-Mental State Examination (MMSE) and the Clinical Dementia Rating (CDR) scale. ***Results:*** *Verification*: The SB-C could be reliably extracted for both languages using an automatic speech processing pipeline. *Analytical Validation*: In both languages, the SB-C was strongly correlated with MMSE scores. *Clinical Validation:* The SB-C significantly differed between clinical groups (including MCI and dementia), was strongly correlated with the CDR, and could track the clinically meaningful decline. ***Conclusion:*** Our results suggest that the ki:e SB-C is an objective, scalable, and reliable indicator of cognitive decline, fit for purpose as a remote assessment in clinical early dementia trials.

Introduction

Progressive cognitive decline is the cardinal behavioral symptom in most dementia-causing diseases such as Alzheimer's disease [1]. Whereas classic neuropsychological tests often have excellent psychometric properties to measure cognitive decline in dementia, there are scenarios in which they are less suitable or not applicable at all. Since most traditional assessments are not perfectly suitable for decentralized remote clinical trials, as they require physical presence of clinicians, digital cognitive assessments will gain in importance [2]. Digital cognitive assessments are better suited for automated patient-ad-

Karger@karger.com
www.karger.com/dib

Correspondence to:
Johannes Tröger, johannes.troeger@ki-elements.de

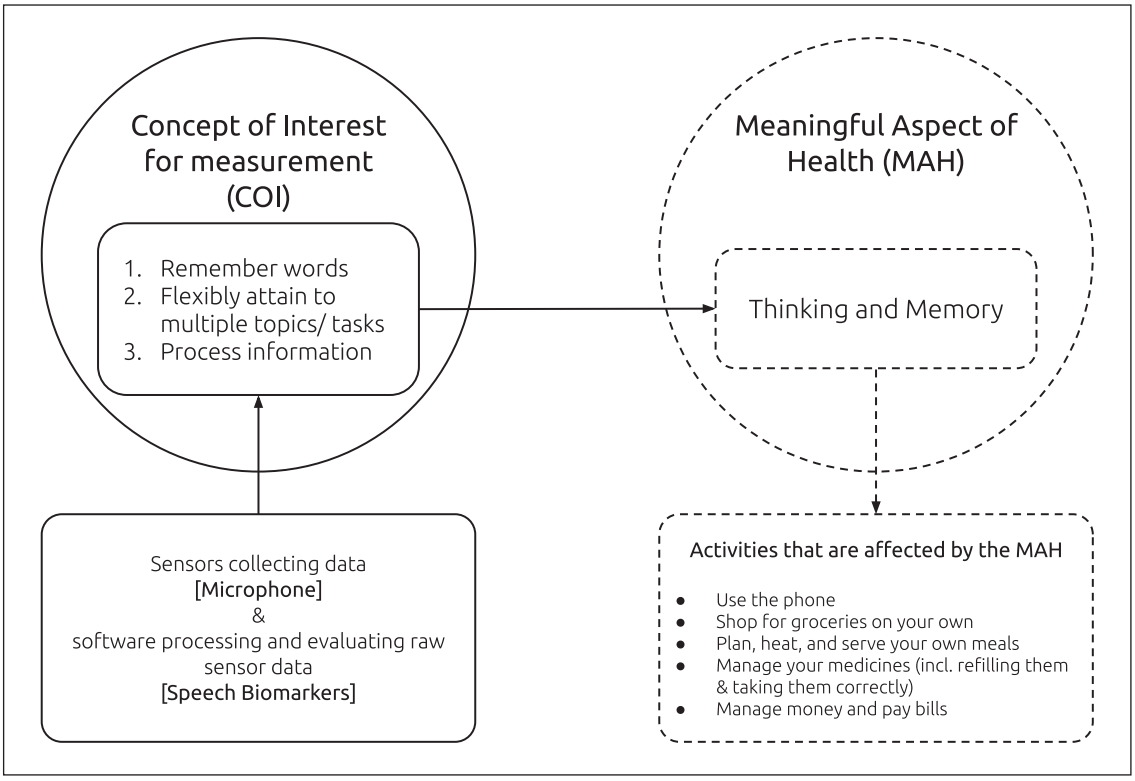

Fig. 1. Conceptual framework connecting the meaningful aspect of health in dementia patients with the SB ki:e SB-C.

ministered screening or prescreening at low cost to accelerate trial inclusion [3]. Furthermore, a high level of automation can easily scale-up outreach to draw unbiased and representative trial populations beyond established clinical site and hospital networks. Also within clinical trials, digital markers deliver objective, high-frequency data to guide interventional clinical trial decision-making and make evaluation more efficient [4]. Speech-based digital biomarkers are especially promising solutions, as they can be deployed remotely and extracted in a highly automated fashion, allowing monitoring and diagnostic solutions to scale both for clinical trials as well as health care [5].

Just as traditional "wet" biomarkers, digital speech biomarkers (SBs) measure indicators of normal biological processes, pathogenic processes, or responses to an exposure or intervention [6]. If this biomarker is collected using a digital sensing product, it is called a digital biomarker. If this sensing product captures speech (e.g., using a microphone), it qualifies as a speech digital biomarker or just SB (sometimes also called voice biomarker [7]).

When referring to SBs, it is important to carefully consider the technical framework that a specific SB is embedded into, as given by its intended use. Depending on this framework, use-case-specific sensor setups need to be evaluated as part of the SB evaluation (e.g., lower audio sampling rate for a telephone-based deployment compared to an app-based usage [8]).

From a conceptual/clinical point of view, it is important to define how speech readouts connect to a concept of interest (e.g., a symptom or syndrome) and eventually a meaningful aspect of health. This conceptual framework eventually also drives the validation efforts. Conceptually, an SB often does not measure the capability of a human being to talk and use language but rather uses speech as a medium to measure language-distant pathogenic processes such as breathiness in spoken language due to asthma or other respiratory diseases.

As far as SBs-C are concerned, there is a special case in which a SB measures language as a cognitive function, and therefore, the medium (speech) overlaps directly with the concept of interest (speech/language capabilities; for an overview see [7]). However, this paper focuses on

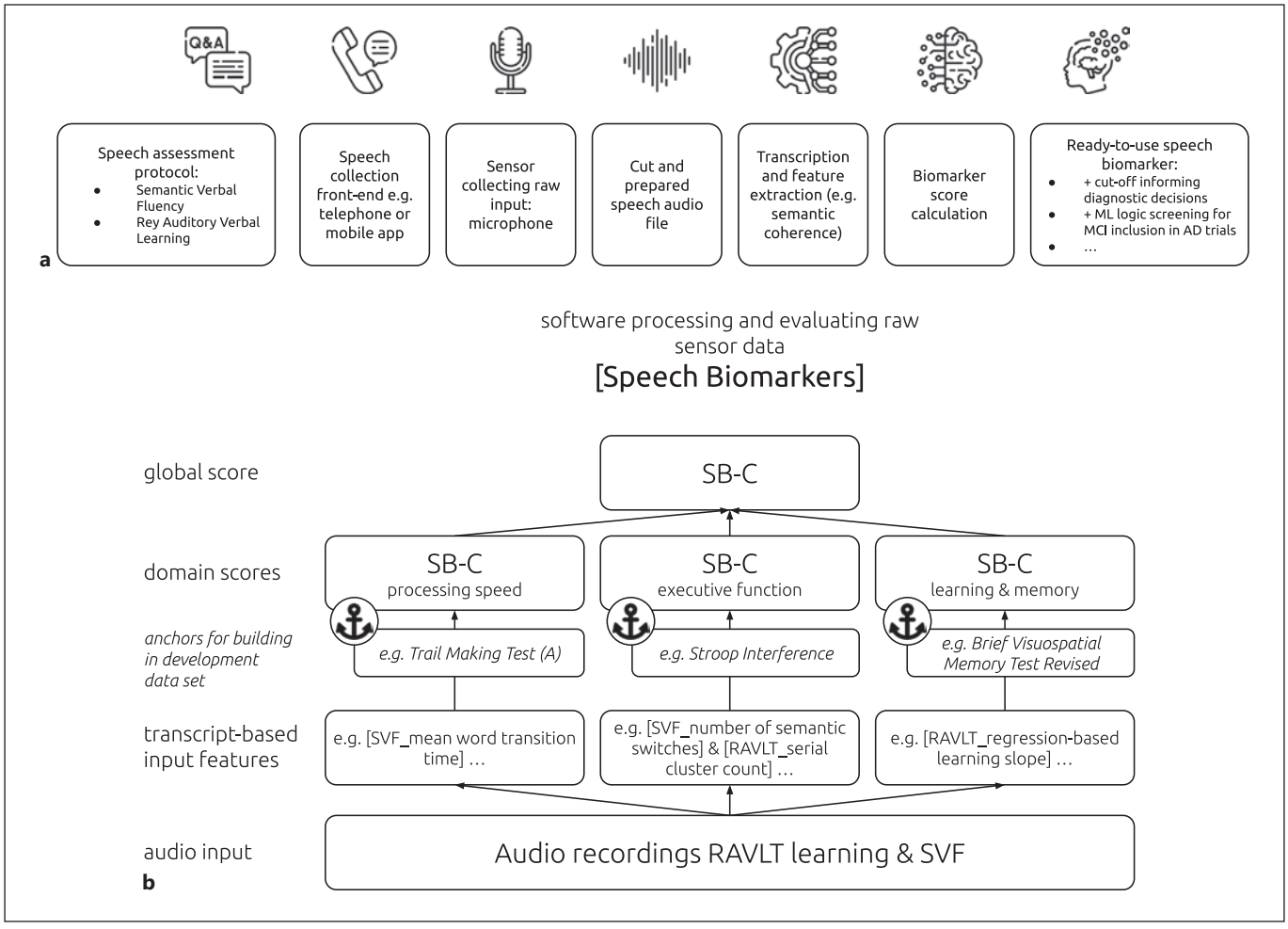

Fig. 2. a Schematic representation of the processing steps involved in collecting and calculating the ki:e SB-C. **b** Schematic representation of the structure and development of the ki:e SB-C and its domain scores.

SBs that also measure other aspects of cognition than language such as learning and memory, executive function, or processing speed (for a breakdown of this conceptual framework of ki:e SB-C, see Fig. 1).

Over the last decade, substantial work has been done on algorithms using artificial intelligence and speech analysis to predict cognitive decline in an elderly population. However, the main limitations of the field are poor standardization as well as limited comparability of results [9], which prevents such innovative solutions from being integrated into clinical practice or trials.

Addressing this challenge, the Digital Medicine (DiME) Society established the V3 framework to evaluate digital assessments (including SBs) as fit for purpose for use in clinical trials [10]. This paper evaluates the ki:e SB-C, following the DiME's V3 framework, presenting results on verification, analytical, as well as clinical validation from a Dutch elderly clinical cohort.

Methods

ki:e SB-C

The ki:e SB-C [11] is a composite score built up of more than 50 automatically extracted speech features that compose three distinct neurocognitive domain scores (learning and memory, executive function, and processing speed). From the three domain scores, one aggregated global score for cognition is derived. The ki:e SB-C takes as input speech recordings from two standard neuropsychological assessments: the Rey Auditory Verbal Learning Test (RAVLT; [12]) and the Semantic Verbal Fluency task (SVF). Speech from both tests is automatically processed using the pro-

Table 1. Sample description of the data sets used for evaluation of the ki:e SB-C

Language	DeepSpA T0			DeepSpA T12 (N T0 and T12)		DeepSpA T12 (N T12 and T15)		DeepSpA T15 (N T12 and T15)		SPeAk (EPAD) T0	SPeAk (EPAD) T3
	Dutch			Dutch		Dutch		Dutch		English	English
Subgroups	SCI	MCI	Dementia	SCI	MCI	SCI	MCI	SCI	MCI	HC	HC
N	69 (24 F)	52 (19 F)	13 (7 F)	43 (15 F)	23 (6 F)	21 (7 F)	8 (2 F)	*same as T12*	*same as T12*	25 (15 F)	25 (15 F)
Age	62.20 (10.71)	70.29 (9.83)	77.38 (6.50)	63.95 (9.87)	70.35 (9.18)	62.48 (9.48)	68.63 (8.81)	*same as T12*	*same as T12*	–	–
CDR-GS	0.37 (0.28)	0.48 (0.20)	0.73 (0.26)	0.30 (0.27)	0.52 (0.18)	0.31 (0.25)	0.44 (0.18)	*same as T12*	*same as T12*	–	–
CDR-SOB	0.74 (0.86)	1.56 (1.25)	3.62 (2.27)	0.70 (0.87)	1.76 (1.40)	0.64 (0.82)	1.38 (1.16)	*same as T12*	*same as T12*	0.00 (0.00)	0.00 (0.00)
MMSE	28.71 (0.23)	26.85 (2.04)	23.31 (2.66)	28.91 (1.19)	26.87 (1.94)	29.29 (0.72)	26.88 (2.17)	*same as T12*	*same as T12*	–	–
ki:e SB-C	0.49 (0.12)	0.32 (0.12)	0.17 (0.12)	0.50 (0.12)	0.31 (0.09)	0.49 (0.15)	0.36 (0.06)	0.50 (0.16)	0.24 (0.08)	0.61 (0.14)	0.65 (0.15)

Values are expressed as mean (standard deviation), unless otherwise indicated. SCI, subjective cognitive impairment; MCI, mild cognitive impairment; HC, healthy controls; CDR, Clinical Dementia Rating; GS, Global Score; SOB, sum of boxes; MMSE, Mini-Mental State Examination; SB-C, speech biomarker for cognition.

prietary speech analysis pipeline from ki:elements that involves automatic speech recognition to transcribe speech and feature extraction. Subsequently, the domain scores and the global score for cognition are calculated (for the processing steps, see Fig. 2a). The ki:e SB-C can be collected fully automatically both over traditional landline phone infrastructure as well as in face-to-face on-site settings using mobile front ends.

The ki:e SB-C has been developed based on the 2019–2021-recruited H70 Swedish birth cohort leveraging a population-representative sample of 404 non-demented 75-year-old participants that, among other psychometric testing, also produced recordings of the RAVLT and SVF. From those recordings, meaningful linguistic features have been extracted and combined into neurocognitive domain scores validated by established psychometric anchors for each domain available in the H70 dataset. Please see Figure 2b for a schematic overview of how the ki:e SB-C and its domains are built.

V3 Framework

In this paper, we evaluated the ki:e SB-C as fit for-purpose for use in clinical trials in an elderly population potentially being candidates for prodromal or preclinical AD trials. The V3 framework established by the DiME Society [10] provides a unified evaluation framework for digital tools such as SBs. V3 includes three distinct phases in sequential order: verification, analytical validation, and clinical validation. For all the three phases, different data have to be collected and statistically analyzed to provide the necessary results.

Data

Evaluation of the ki:e SB-C is based on two clinical studies: the Dutch DeepSpA study [13, 14] and the English SPeAk dataset (based on the EPAD generation Scotland cohort [15]). Both studies collected the ki:e SB-C over telephone or app-based front-end following the same protocol mentioned above. For an overview of the datasets, see Table 1.

DeepSpA

140 participants were recruited at the memory clinic of the Maastricht University Medical Center as part of the BioBank Alzheimer Center Limburg study (BBCL, including participants with subjective cognitive impairment – SCI, mild cognitive impairment – MCI or dementia). For this research, we excluded 6 from the original 140 subjects due to bad audio quality and operational issues. Participants underwent a yearly in-person assessment (T0 - baseline and T12 - at 12th month) at the clinic, and during those assessments, the speech data for ki:e SB-C were collected using a mobile application. For the follow-up assessment (T15 - at 15th month), participants were contacted via the fully automatic telephone front-end performing the assessment. Not all participants concluded all timepoints. This was either due to BBCL study procedures, which prevent patients with dementia from participating in follow-ups or especially for the on-site data collection at T12 (Q4 of 2021) due to the COVID-19 pandemic (multiple on-site in clinic visits had to be canceled due to COVID-19 preventive measures). Overall, there were only 29 participants from the initial 140 that concluded both T12 and T15. Mini-Mental State Examination (MMSE [16]) and Clinical Dementia Rating (CDR) [17]) data are available for the yearly on-site visits T0 and T12. Based on the CDR total score, which has values of 0, 0.5, 1, 2, 3 corresponding to de-

mentia stages [18], participants were classified into decliners (N = 10; change from 0 to 0.5 between T0 and T12) and non-decliners (N = 47, no change between T0 and T12).

As part of their on-site visit, all patients underwent a 2-hour extensive neuropsychological assessment in addition to the standard clinical assessments. A clinical diagnosis of MCI or dementia was made based on the regular DSM-V criteria of minor and major neurocognitive disorder and everyday functioning. All other patients were categorized as having SCI since they had cognitive complaints that were not objectified in cognitive testing. Diagnoses were established in the panel including neuropsychologists, caregivers, and neurologists within the standard memory clinic routine.

SPeAk

Twenty-five healthy participants were recruited from the EPAD Generation Scotland readiness cohort. Participants underwent an in-person baseline assessment as part of their yearly cohort assessments (T0) at the clinic (for study protocol see [19]). The ki:e SB-C was collected during each assessment using a mobile application. For the follow-up assessment (T3 - at 3rd month), participants were contacted via the fully automatic telephone front-end performing the assessment.

Ethics

Both studies that provided data to this research have been conducted in compliance with the ethical principles for medical research involving human subjects, as defined in the Declaration of Helsinki and the European General Data Protection Regulation. For the Dutch DeepSpA study, the local Medical Ethical Committee (METC MUMC/UM) approved the study (MEC 15-4-100). For the English SPeAk study, the study has been approved by the Edinburgh Medical School Research Ethics Committee (REC reference 20-EMREC-007). All participants provided informed consent before completing any study-related procedures; participants had to have capacity to consent to participate in this study. For both studies, each participant gave written informed consent before the assessment.

Statistical Analysis

Across the V3 phases, different statistical tests are performed on the abovementioned datasets to evaluate the ki:e SB-C.

Verification

Verification of SBs entails the systematic evaluation of sample-level sensor outputs against prespecified criteria. The ki:e SB-C uses transcribed speech as input and no acoustic features. Therefore, the most critical part of the sensor output and preprocessing pipeline is the automatic transcription of speech (automatic speech recognition, ASR). The ki:e SB-C uses a proprietary speech processing pipeline based on the Google Speech API [20]. The performance of ASR systems is determined using Word Error Rate (WER). To calculate WER, we compared the ASR performance against manually corrected transcripts from trained clinical personnel. Manual transcripts, ASR transcripts, and WER were obtained for all participants in both datasets. WER is computed as the error between the number of target words in the manual transcripts and that in the ASR transcripts. In the RAVLT, only correctly remembered words are considered for the WER calculation. Based on the previous literature in similar scenarios, a mean WER of 20% is considered acceptable [21–23]. However, across the clinical stages, WER can vary substantially on an individual level, which is why the median WER is a better measure. For verification, we evaluate ASR performance in two different languages: Dutch – DeepSpA and English – SPeAk.

Analytical Validation

Analytical validation of SBs evaluates performance of the algorithms to measure a certain concept of interest (similar to construct validity). The ki:e SB-C measures cognition as the concept of interest. For the analytical validation, we compared the ki:e SB-C score with a gold standard anchor measure for cognition in this population, evaluate its stability within a short time frame as well as its ability to detect change in a nonclinical sample.

For the comparison with a gold standard anchor, we compute Spearman rank partial correlation between the digital biomarker score and the MMSE score in DeepSpA T0, taking confounding age effects into account. Moreover, we calculated Spearman rank partial correlation between the digital biomarker sub-scores and the MMSE subdomain scores (executive function, processing speed, and MMSE attention/concentration; memory and MMSE delayed recall), to control for the effect of age.

For evaluating measurement stability, we calculate test-retest reliability within 3 months for DeepSpA (T12-T15) and SPeAk (T0-T3). It can be expected that there are little to no learning effects within 3 months as well as no significant clinical progression of patients [24].

Finally, we evaluate the score's ability to detect nonclinical, age-related cognitive change by comparing young and old cognitively unimpaired participants (SCI subgroups) in their digital biomarker scores. Therefore, we perform a median split within DeepSpA SCI population (participants that did T0 and T12 follow-up) and perform group comparison in biomarker score between the two age groups using nonparametric Kruskal-Wallis test. Age has a significant effect on cognition [25], which should result in observable but not clinically meaningful differences in the biomarker score.

Clinical Validation

Clinical validation of SBs evaluates whether algorithms validly measure clinically meaningful change within an intended scenario including a specified clinical population. The ki:e SB-C is built to measure clinically meaningful change in cognition with a focus on MCI and dementia population. We perform group comparisons in biomarker scores between the different diagnostic groups (ANCOVA controlling for age between SCI, MCI, and dementia groups) and between groups that differ minimally in a clinical anchor (ANCOVA controlling for age with grouping factor CDR-GS 0 vs. CDR-GS 0.5). Additionally, we analyze the correlation between the biomarker score and a clinical anchor measure in the DeepSpA T0 population (Spearman rank partial correlation between ki:e SB-C score and CDR-SOB, controlling for the effect of age). To evaluate the ability of the biomarker to detect disease progression, we compare biomarker score changes from T0 to T12 DeepSpA) between decliners (progression in CDR-GS from 0 to 0.5) and non-decliners (no change in the CDR total score). This was done comparing the SB-C score difference between timepoints T12 and T0 between decliners and non-decliners using nonparametric Kruskal-Wallis test.

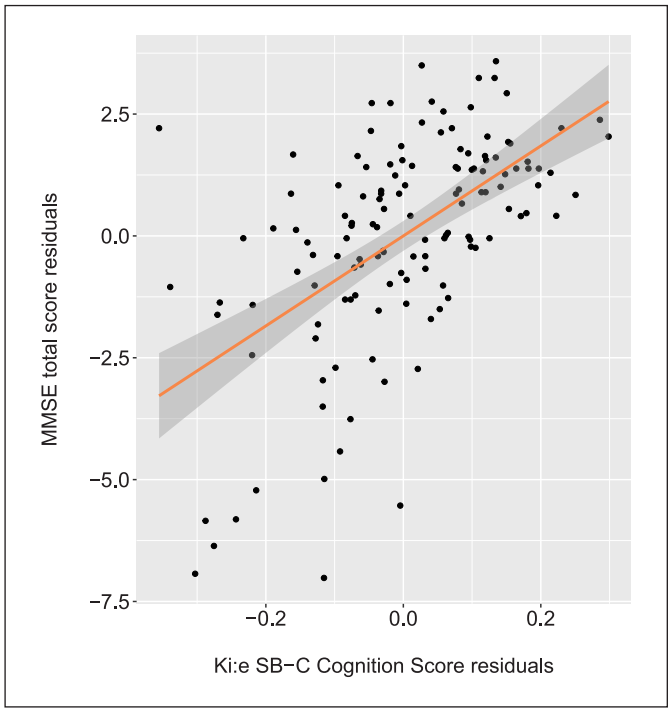

Fig. 3. Partial correlation between ki:e SB-C global biomarker score and MMSE total score, controlling for age.

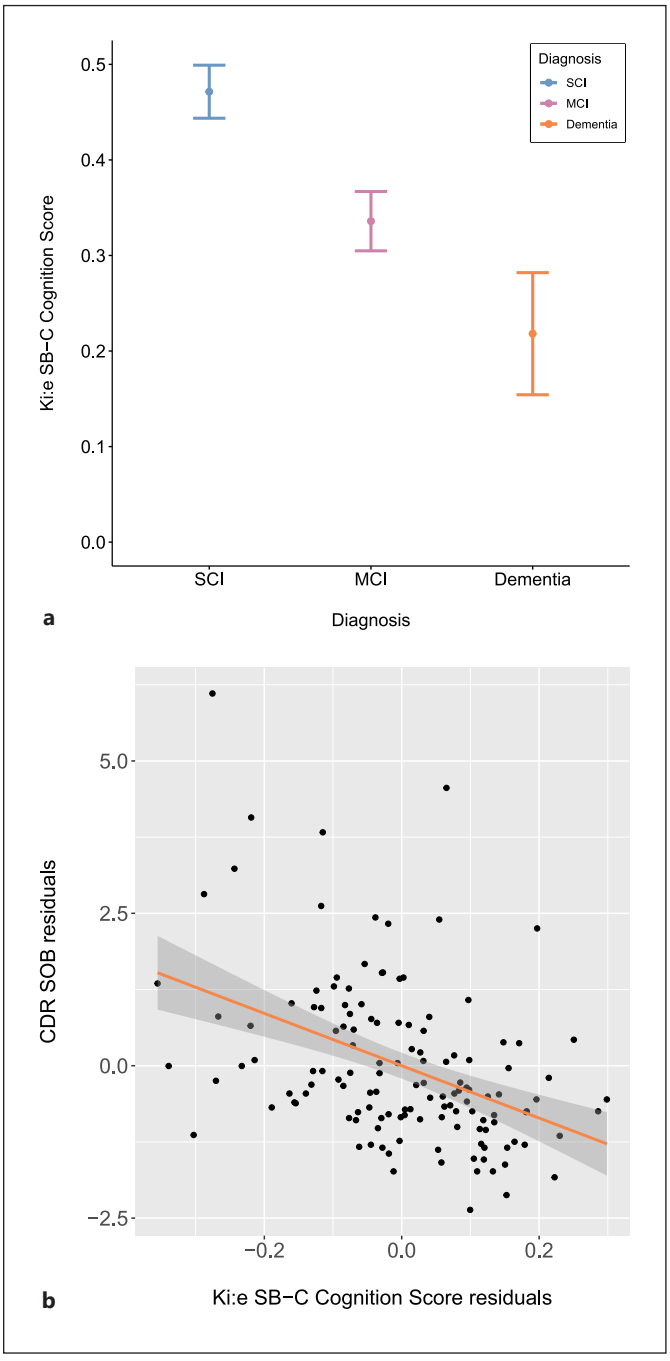

Fig. 4. a Group difference in ki:e SB-C score between diagnostic groups at DeepSpA T0, controlling for age. **b** Partial correlation between ki:e SB-C score and CDR-SOB at DeepSpA T0, controlling for age.

Results

We report results according to the V3 framework: verification, analytical validation, and clinical validation.

Verification

Results show for both Dutch and English language a median WER of below 16%. Median WER in Dutch DeepSpA sample was higher than in the English SPeAk sample, but both WERs were within the acceptable range (<20%). Additionally, there was no significant difference between the WERs of each task (SVF and VLT phases). Also compare Table 2.

Analytical Validation

In the Dutch DeepSpA T0 sample, MMSE was strongly correlated with the ki:e SB-C global score ($r = 0.54$, $p < 0.001$, $d = 1.28$; see Fig. 3). In addition, ki:e SB-C subscores were strongly correlated with MMSE subdomain scores (executive function and MMSE attention/concentration $r = 0.28$, $p < 0.01$, $d = 0.58$; memory and MMSE-delayed recall $r = 0.40$, $p < 0.001$, $d = 0.87$; processing speed and MMSE attention concentration $r = 0.28$, $p < 0.001$, $d = 0.59$).

Moreover, we found strong correlations between the ki:e SB-C global score at two consecutive assessments 3 months apart both in Dutch (DeepSpA T12 & T15; $r = 0.72$, $p < 0.001$, $d = 2.05$) and English (SPeAk T0 & T3;

Table 2. Overview of WER between manually corrected transcripts (ground truth) and ASR transcripts based on the DeepSpA T0 and SPeAK T0 sample

DeepSpA T0		DeepSpA SCI T0	DeepSpA MCI T0	DeepSpA dementia T0	DeepSpA all T0	SPeAk
SVF	Mean	22.00	22.30	18.33	21.45	7.06
	SD	21.30	19.38	13.22	20.04	6.49
	Median	15.79	19.09	12.50	15.14	5.56
RAVLT-learning phase	Mean	17.10	17.76	30.15	16.71	12.95
	SD	17.11	14.57	21.48	17.20	11.59
	Median	11.90	16.06	26.32	12.12	11.48

MCI, mild cognitive impairment; SVF, semantic verbal fluency; VLT, verbal learning test; SD, standard deviation.

Table 3. Group differences in biomarker and MMSE scores at two timepoints between two age groups

	Age <65 years	Age ≥ 65 years	*p* value Kruskal test
N	21	22	
MMSE	29.19 (0.93)	28.64 (1.36)	0.197
ki:e SB-C at T0	0.56 (0.09)	0.44 (0.12)	<0.01
ki:e SB-C at T12	0.55 (0.14)	0.39 (0.13)	<0.01

SB-C, speech biomarker for cognition; MMSE, Mini-Mental Status Examination.

$r = 0.57$, $p < 0.01$, $d = 1.40$) data. As there are no disease-related changes to be expected in a 3-months' time frame, this test serves as check for test-retest reliability.

There was a significant difference in the ki:e SB-C global score between median-split age groups ([Age <65] < [Age ≥ 65]) in the Dutch DeepSpA sample both at T0 (χ^2 (1) = 10.40, $p < 0.01$, Cohen's $d = 1.09$) and T12 (χ^2 (1) = 9.82, $p < 0.01$, Cohen's $d = 1.05$)). These results were also followed by a similar group difference trend in the MMSE scores available from T0 (see Table 3).

Clinical Validation

Clinical validation was performed on the DeepSpA T0 and T12 samples. There was a significant difference of the ki:e SB-C score between the three clinical groups at DeepSpA T0 (SCI > MCI > dementia; $F(2,130) = 31.96$, $p < 0.001$, $d = 1.40$) (see Fig. 4a) and between groups that differ minimally in a clinical anchor (CDR-GS 0 vs. 0.5) ($F(1,118) = 8.79$, $p < 0.001$). Additionally, we observe a strong correlation of the biomarker score with the clinical anchor score CDR-SOB ($r = -0.42$, $p < 0.001$, $d = -0.93$); see Figure 4b. Comparing T12 and T0, the Kruskal-Wallis test revealed a significant difference between decliners and non-decliners in the biomarker score difference between timepoints T12 and T0 ($\chi^2 = 7.55$ (1), $p < 0.001$; see Figure 5).

Discussion

The digital speech biomarker ki:e SB-C was developed as a measure for cognition in elderly population with a special focus on dementia-related cognitive impairment. The current study aimed at evaluating whether this novel biomarker is fit for purpose in clinical trials through two independent samples using the V3 framework developed by the DiME Society. ki:e SB-C can be calculated automatically and robustly from two standard neuropsychological tests (SVF and RAVLT) and, as our results revealed, can validly measure cognitive function in an elderly population that is at risk of developing dementia.

The novel biomarker ki:e SB-C was systematically evaluated to determine fit for purpose by following the V3 framework. One important aspect and strength of digital biomarkers in general, and ki:e SB-C in particular, is automation. Speech data can be easily collected by telephone or by mobile front-ends face-to-face, and ki:e SB-C can be reliably calculated using the proprietary automatic speech analysis pipeline. Verification step of this study revealed that the automatic speech processing pipeline, including speech recognition, performs at an acceptable level across different languages and tasks, and therefore, automatic processing works reliably for the purposes of

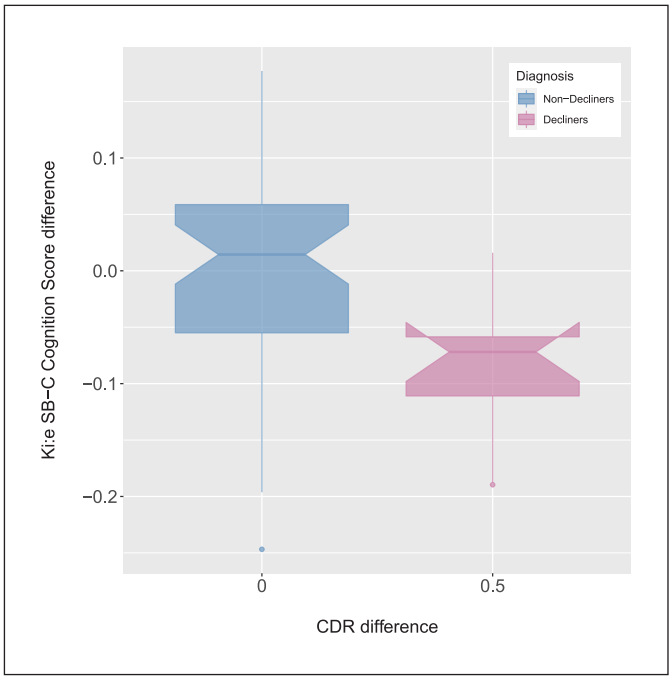

Fig. 5. SB-C score differences (Δ[SB-C_T12 – SB-C_T0]) between decliners and non-decliners. Those participants with a positive CDR difference (CDR_T12 – CDR_T^0 = 0.5) are considered decliners as their CDR score increased in T12 as compared to T0.

the biomarker. The aim of the analytical validation was threefold: to evaluate (1) the ki:e SB-C against a cognitive gold standard measure (MMSE), (2) the biomarker's retest reliability, and (3) how well ki:e SB-C reflects age-related – but clinically not relevant – changes. The results of the analytical validation analyses revealed that ki:e SB-C was a valid biomarker to measure cognitive abilities that are relevant for the target population, as seen by its high correlation with MMSE scores even if corrected for the effect of age. Furthermore, the automatically calculated biomarker was stable in retesting as assessed by a test-retest analysis.

It is well established that aging is a risk factor for cognitive decline and diseases [26]; however, not all changes are pathological. It is important for a biomarker that is a surrogate for cognition to reflect subtle changes that are at a subclinical level. Our results showed that ki:e SB-C can detect age-related changes in cognitive function. These changes are also reflected in MMSE scores as a trend. Regarding the direct comparison with the MMSE on the healthy DeepSpA population, the SB-C seems to vary more marked by a larger standard deviation (regarding the mean) as compared to the MMSE. This might be also due to the fact that the SB-C is a more fine-grained measure as compared to the MMSE, which might especially in the healthy population lack the resolution to measure cognitive changes at this higher functioning level and would elicit less variance.

Another important characteristic of a biomarker is its clinical meaningfulness, i.e., a biomarker should be related to the concept of interest that is clinically relevant for the target population and for the context of use. Our results show that ki:e SB-C score is associated with the scores of CDR scale, meaning that the biomarker is related to the existence and severity of dementia. This finding is further supported by the significant difference in the biomarker scores between the diagnostic groups, i.e., ki:e SB-C reflects clinically defined disease stages. Moreover, the biomarker score difference was different between decliners and non-decliners, which further confirms the biomarker's clinical relevance. Altogether, the ki:e SB-C is sensitive to minimal clinical differences between people and can detect clinical changes within a person.

The results of our validation study are in line with studies using speech as a marker of cognitive health and impairment [27–29]. Our work extends these previous findings by thoroughly and carefully validating (V3 framework) ki:e SB-C as a novel biomarker both in healthy and clinical samples and in two different languages (Dutch and English). To our knowledge, this is the first study that evaluates a speech-based digital marker systematically using the V3 framework [7].

One limitation of our work is that we could not perform predictive analyses due to the small number of decliners in our sample, which is likely to be related to the pandemic and small longitudinal sample size, and the relatively short follow-up period for neurodegenerative diseases. The next step would be to evaluate the performance of the biomarker ki:e SB-C as a predictor of change over time in a larger longitudinal sample with a longer follow-up period.

Conclusion

The aim of the current study was to validate a novel digital speech-based biomarker for cognitive impairment related to MCI and dementia diagnoses, the ki:e SB-C. The novel biomarker ki:e SB-C has undergone considerable validation following the V3 framework. The results of the validation analyses revealed that ki:e SB-C correctly measures cognitive impairment associated with MCI and dementia. The ki:e SB-C is a promising digital bio-

marker that has the potential to detect subtle prodromal changes longitudinally as well as to discriminate between diagnostic groups cross-sectionally.

The ki:e SB-C can be utilized to select target populations for clinical studies and may in the future function as a surrogate disease marker. The ki:e SB-C can be further used to identify patients in preclinical disease stages who have a low disease burden as targets for prevention and early treatment.

Statement of Ethics

Both studies that provided data to this research have been conducted in compliance with the Ethical Principles for Medical Research Involving Human Subjects, as defined in the Declaration of Helsinki and the European General Data Protection Regulation. For the Dutch DeepSpA study, the local Medical Ethical Committee (METC MUMC/UM) approved the study (MEC 15-4-100). For the English SPeAk study, the study has been approved by the Edinburgh Medical School Research Ethics Committee (REC reference 20-EMREC-007). All participants provided informed consent before completing any study-related procedures; participants had to have capacity to consent to participate in this study. For both studies, each participant gave written informed consent before the assessment.

Conflict of Interest Statement

Daphne ter Huurne, Nina Possemis, and Inez Ramakers have no conflicts of interest. Johannes Tröger, Ebru Baykara, Jian Zhao, Elisa Mallick, Simona Schäfer, Louisa Schwed, Mario Mina, and Nicklas Linz are employees of the digital biomarker company ki elements. Johannes Tröger and Nicklas Linz hold shares of the digital biomarker company ki elements. Craig Ritchie has received consultancy fees from Biogen, Eisai, MSD, Actinogen, Roche, and Eli Lilly, as well as payment or honoraria from Roche and Eisai.

Funding Sources

DeepSpA has been funded by EIT-Health project grant agreement number 19249, as well as supported by Janssen Pharmaceutica NV through a collaboration agreement (award/grant number is not applicable). SPeAk was supported by Janssen Pharmaceutica NV through a collaboration agreement (award/grant number is not applicable).

Author Contributions

Johannes Tröger conceptualized this work; he drafted the manuscript and edited the final version. Ebru Baykara contributed to the overall interpretation of the work and drafting of the manuscript. Elisa Mallick, Simona Schäfer, Louisa Schwed, and Mario Mina implemented the biomarker, analyzed the speech, conducted the statistical work, as well as drafted the methods and results sections of this article. Daphne ter Huurne and Nina Possemis acquired parts of the data, contributed to the clinical interpretation of the results, and revised the document. Jian Zhao oversaw the design of the V3 framework validation pipeline from a regulatory standpoint and revised the document. Nicklas Linz contributed to the overall concept of this research and revised the manuscript. Inez Ramakers is responsible for the DeepSpA study and data acquisition, drafted DeepSpA relevant parts of the document, and revised the manuscript. Craig Ritchie is the principal investigator of SPeAk, responsible for the concept and data acquisition, drafted SPeAk relevant parts of the document, and revised the manuscript.

Data Availability Statement

The datasets used for this research might be available to qualified researchers worldwide. Requests are handled by the respective PI or contact person: Craig Ritchie for SPeAk and Inez Ramakers for DeepSpA. Inquiries can be initially directed to the corresponding author.

References

1 McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011 May;7(3):263–9.

2 Öhman F, Hassenstab J, Berron D, Schöll M, Papp KV. Current advances in digital cognitive assessment for preclinical Alzheimer's disease. Alzheimers Dement Diagn Assess Dis Monit. 2021 Jan;13(1).

3 Gold M, Amatniek J, Carrillo MC, Cedarbaum JM, Hendrix JA, Miller BB, et al. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. Alzheimers Dement Transl Res Clin Interv. 2018 Jan;4(1):234–42.

4 Dorsey ER, Papapetropoulos S, Xiong M, Kieburtz K. The first frontier: digital biomarkers for neurodegenerative disorders. Digit Biomark. 2017 Sep-Dec;1(1):6–13.

5 Carlew AR, Fatima H, Livingstone JR, Reese C, Lacritz L, Pendergrass C, et al. Cognitive assessment via telephone: a scoping review of instruments. Arch Clin Neuropsychol. 2020 Nov;35(8):1215–33.

6 FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Bethesda (MD):Food Drug Administration US. 2016. Available from: https://www.ncbi.nlm.nih.gov/books/NBK326791/.

7 Robin J, Harrison JE, Kaufman LD, Rudzicz F, Simpson W, Yancheva M. Evaluation of speech-based digital biomarkers: review and recommendations. Digit Biomark. 2020 Sep-Dec;4(3):99–108.

8 Tröger J, Linz N, König A, Robert P, Alexandersson J Telephone-based dementia screening I: automated semantic verbal fluency assessment. Pervasive Health; 2018. Vol. 9.

9 de la Fuente Garcia S, Ritchie CW, Luz S. Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: a systematic review. J Alzheimers Dis. 2020 Dec;78(4):1547–74.
10 Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). Npj Digit Med. 2020 Dec;3(1):55.
11 ki elements GmbH. The ki:e speech biomarker for cognition. 2022 [cited 2022 Jul 29]. Available from: https://ki-elements.de/sb-c/.
12 Bean J. Rey auditory verbal learning test, rey AVLT. In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer; 2011. p. 2174–5.
13 ter Huurne DBG, Ramakers IHGB, Linz N, König A, Langel K, Lindsay H, et al. Clinical use of speech and linguistic features automatically derived from the semantic verbal fluency test. 2021.
14 Linz N, ter Huurne DB, Langel K, Ramakers IH, König A, DeepSpA . Artificial Intelligence empowered recruitment for clinical trials. Alzheimers Dement. 2021 Dec;17(S8).
15 Ritchie CW, Molinuevo JL, Truyen L, Satlin A, Van der Geyten S, Lovestone S, et al. Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. Lancet Psychiatry. 2016 Feb;3(2):179–86.
16 Burns A, Brayne C, Folstein M. Key Papers in Geriatric Psychiatry: mini-mental state: a practical method for grading the cognitive state of patients for the clinician. M. Folstein, S. Folstein and P. McHugh, Journal of Psychiatric Research, 1975, 12, 189–198. Int J Geriatr Psychiatry. 1998 May;13(5):285–94.
17 Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int Psychogeriatr. 1997 Dec;9(Suppl 1):173–6; discussion 177-8.
18 Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry. 1982 Jun;140(6):566–72.
19 Gregory S, Linz N, König A, Langel K, Pullen H, Luz S, et al. Remote data collection speech analysis and prediction of the identification of Alzheimer's disease biomarkers in people at risk for Alzheimer's disease dementia: the Speech on the Phone Assessment (SPeAk) prospective observational study protocol. BMJ Open. 2022 Mar;12(3):e052250.
20 Google. Google Speech API. [cited 2022 Jul 29]. Available from: https://cloud.google.com/speech-to-text/.
21 Lehr M, Prud'hommeaux E, Shafran I, Roark B. Fully automated neuropsychological assessment for detecting Mild cognitive impairment, Vol. 4. 2012.
22 Pakhomov SVS, Marino SE, Banks S, Bernick C. Using automatic speech recognition to assess spoken responses to cognitive tests of semantic verbal fluency. Speech Commun. 2015 Dec;75:14–26.
23 König A, Linz N, Tröger J, Wolters M, Alexandersson J, Robert P. Fully automatic speech-based analysis of the semantic verbal fluency task. Dement Geriatr Cogn Disord. 2018;45(3–4):198–209.
24 Feinkohl I, Borchers F, Burkhardt S, Krampe H, Kraft A, Speidel S, et al. Stability of neuropsychological test performance in older adults serving as normative controls for a study on postoperative cognitive dysfunction. BMC Res Notes. 2020 Dec;13(1):55.
25 Deary IJ, Corley J, Gow AJ, Harris SE, Houlihan LM, Marioni RE, et al. Age-associated cognitive decline. Br Med Bull. 2009 Dec;92(1):135–52.
26 Niccoli T, Partridge L. Ageing as a risk factor for disease. Curr Biol. 2012 Sep;22(17):R741–52.
27 Lin H, Karjadi C, Ang TFA, Prajakta J, McManus C, Alhanai TW, et al. Identification of digital voice biomarkers for cognitive health. Explor Med. 2020 Dec;1(6):406–17.
28 Martínez-Nicolás I, Llorente TE, Martínez-Sánchez F, Meilán JJG. Ten years of research on automatic voice and speech analysis of people with Alzheimer's disease and mild cognitive impairment: a systematic review article. Front Psychol. 2021 Mar;12:620251.
29 Robin J, Xu M, Kaufman LD, Simpson W. Using digital speech assessments to detect early signs of cognitive impairment. Front Digit Health. 2021 Oct;3:749758.