Digital Biomarkers

Digit Biomark 2020;4:99–108
DOI: 10.1159/000510820
Received: June 12, 2020
Accepted: August 11, 2020
Published online: October 19, 2020

© 2020 The Author(s)
Published by S. Karger AG, Basel
www.karger.com/dib

OPEN ACCESS

This article is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND) (http://www.karger.com/Services/OpenAccessLicense). Usage and distribution for commercial purposes as well as any distribution of modified material requires written permission.

Validation – Review Article

Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations

Jessica Robin[a] John E. Harrison[b–d] Liam D. Kaufman[a] Frank Rudzicz[e–g] William Simpson[a, h] Maria Yancheva[a]

[a]Winterlight Labs, Toronto, ON, Canada; [b]Metis Cognition Ltd., Park House, Kilmington Common, Warminster, UK; [c]Alzheimer Center, AUmc, Amsterdam, The Netherlands; [d]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [e]Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada; [f]Department of Computer Science, University of Toronto, Toronto, ON, Canada; [g]Vector Institute for Artificial Intelligence, Toronto, ON, Canada; [h]Department of Psychiatry and Behavioural Neuroscience, McMaster University, Hamilton, ON, Canada

Keywords
Digital biomarkers · Validation · Speech · Language · Digital health · Dementia

Abstract
Speech represents a promising novel biomarker by providing a window into brain health, as shown by its disruption in various neurological and psychiatric diseases. As with many novel digital biomarkers, however, rigorous evaluation is currently lacking and is required for these measures to be used effectively and safely. This paper outlines and provides examples from the literature of evaluation steps for speech-based digital biomarkers, based on the recent V3 framework (Goldsack et al., 2020). The V3 framework describes 3 components of evaluation for digital biomarkers: verification, analytical validation, and clinical validation. Verification includes assessing the quality of speech recordings and comparing the effects of hardware and recording conditions on the integrity of the recordings. Analytical validation includes checking the accuracy and reliability of data processing and computed measures, including understanding test-retest reliability, demographic variability, and comparing measures to reference standards. Clinical validity involves verifying the correspondence of a measure to clinical outcomes which can include diagnosis, disease progression, or response to treatment. For each of these sections, we provide recommendations for the types of evaluation necessary for speech-based biomarkers and review published examples. The examples in this paper focus on speech-based biomarkers, but they can be used as a template for digital biomarker development more generally.

© 2020 The Author(s)
Published by S. Karger AG, Basel

Jessica Robin
Winterlight Labs
46 Hayden Street, Suite 400
Toronto, ON, M4Y 1V8 (Canada)
jessica@winterlightlabs.com

Karger

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024

Introduction

Research into digital biomarkers is an area of rapid growth in digital medicine. Recent reviews have shown how digital measures will bring benefits to clinical research and practice [1–6]. For example, digital biomarkers are more ecologically valid measures than many current clinical assessments of cognition and function, allowing assessments to occur during everyday activities or with little instruction. They also offer a promising solution for known problems associated with current clinical tools, especially repeated assessment effects, inter-rater variability, time investment, and expense. Importantly, the use of digital measures will facilitate remote testing, improving accessibility and reducing the risks inherent in visiting health care centers. The use of digital biomarkers facilitates frequent testing with its potential to provide richer and more detailed data. This approach holds considerable promise for yielding more sensitive measures of symptoms and disease, but many of these potential advantages must be borne out with empirical testing.

Systematic and rigorous evaluation of digital biomarkers is crucial to ensure that they are providing accurate measurement and can serve as suitable surrogate endpoints for detecting and monitoring disease. Following Strimbu and Tavel [7], we prefer to use the term "evaluation" rather than "validation" to refer to the overall process of assessing a biomarker, since this is a continuous process and may not have a single, conclusive outcome. We approach biomarker evaluation as a systematic series of studies that evaluate and quantify the suitability of a given biomarker and its context for use. In this paper, we review recommendations for digital biomarker evaluation using speech-based biomarkers as an illustrative case [2, 7–11].

Speech offers rich insights into cognition and function and is affected by many psychiatric and neurodegenerative diseases [12–16]. By requiring the coordination of various cognitive and motor processes, even a short sample of speech may provide a sensitive snapshot of cognition and functioning relevant to many disease areas. Speech can encompass a broad range of behaviors, from the simple production of sounds or single words to spontaneous, natural language produced in conversation (Fig. 1). In this paper, we consider speech to include any task involving the oral articulation of sounds and words, but we acknowledge that how speech is elicited and collected can affect its relevance to disease. For example, reading scripted passages may be well-suited to capture changes in acoustic and motoric aspects of speech, but unstructured, open-ended speech tasks may be needed to capture changes in the organization or complexity of language. Structured speech tasks require instruction and active participation, while less structured speech such as interviews or conversations can be collected passively. Speech can be collected with widely available technology, such as smartphones, thereby facilitating remote and frequent monitoring, which can reduce measurement error. With advances in natural language-processing and machine-learning techniques, speech can be automatically and objectively analyzed, producing high-dimensional data.

Together, we argue that the factors discussed above make speech ideally suited for use as a potential biomarker, with applications in several disease areas. Speech-based biomarkers could facilitate more efficient clinical research and more sensitive monitoring of disease progression and response to treatment. Systematic evaluation of the suitability of speech-based biomarkers is required, however, to achieve these goals. In the following sections, we summarize a recently proposed framework for digital biomarker evaluation, with recommendations of how each aspect could be applied to speech-based measures, highlighting examples from the field.

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024

Digit Biomark 2020;4:99–108
DOI: 10.1159/000510820
© 2020 The Author(s). Published by S. Karger AG, Basel
www.karger.com/dib

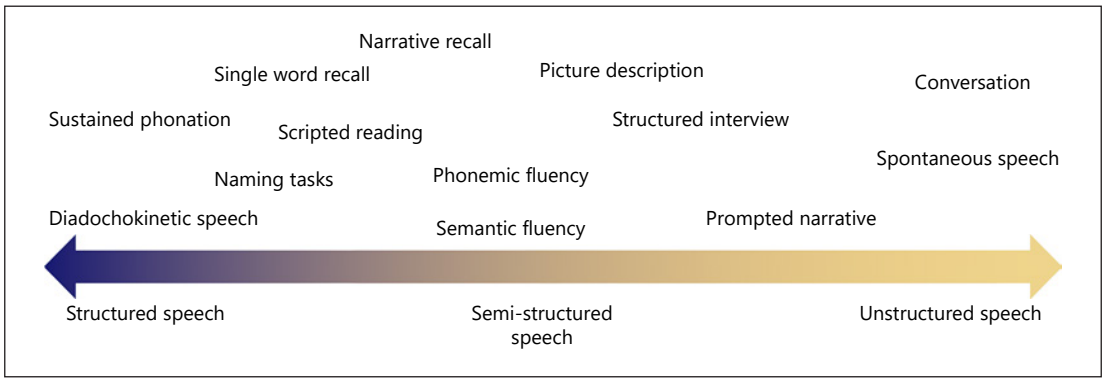

Fig. 1. Example of speech tasks ranging from short, structured speech, to unstructured, naturalistic conversation.

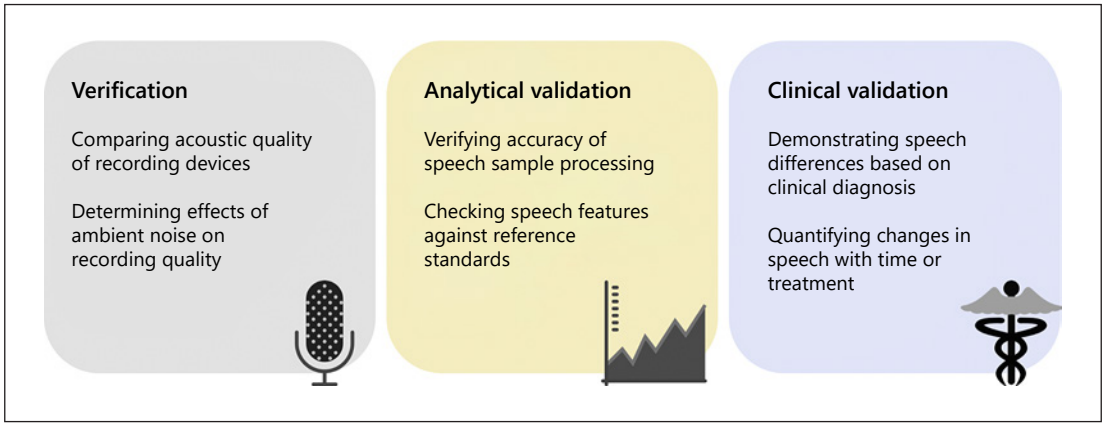

Fig. 2. Recommended steps of evaluation of a speech biomarker, based on the V3 framework (Goldsack et al. [9]).

Evaluation of Speech-Based Digital Biomarkers

In recent papers, members of the Digital Medicine (DiMe) society have proposed a framework and common vocabulary for the evaluation of digital biomarkers [2, 9, 17, 18]. This rigorous framework provides an excellent resource for researchers developing digital health tools, and adherence to a common vocabulary will allow for more consistency across evaluation studies. Based on this V3 framework [9], we summarize the components of evaluation of speech-based biomarkers in the subcategories: verification, analytical validation, and clinical validation. For each, we briefly define the evaluation step, discuss how it could be applied in the context of speech-based biomarkers, and review applicable examples (Fig. 2).

Verification

Verification describes the process of validating the hardware and sensors involved in recording a digital measurement [9]. This form of evaluation, sometimes referred to as bench-testing, is often performed without collecting human data. For speech-based measures, verification primarily involves evaluating the recording devices and determining the conditions required for adequate recording quality. This includes environmental conditions like setting, background noise, and microphone placement, and the properties of the recording tech-

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024

nology like sampling rate, audio format, and upload and storage parameters, which can vary across devices. It is therefore necessary to define the acceptable devices (e.g., smartphones, tablets, and computer microphones) and conditions (e.g., in the clinic, at home, and real-world environments) for recording speech in the context of a speech biomarker, and to perform quantitative comparisons of the quality of recordings across recording conditions and devices. In addition, the design of the user interface could also affect the quality of recordings, e.g., by including empty periods in the recordings to estimate ambient noise, or by providing feedback to the participant in the form of beeps, text, or voice prompts to improve task compliance.

Several recent papers have examined the comparability of audio recordings made across different devices, including smartphones [19–23]. While several studies found that smartphones yield recordings of acceptable quality compared to standard microphones [20, 22, 24], others reported mixed findings with the differences relating to the particular audio outcome measures used [19, 21, 23]. A few studies have examined the effects of ambient noise on voice recordings, with one recommending a threshold of 50 dBA as an acceptable level for their particular speech tasks and outcome measures [21, 25]. Thus, depending on the type of recording device used and the outcome measures, research is needed to establish which devices and recording parameters are acceptable for speech-based biomarkers. Ideally, these conditions should be able to be identified by the user of the biomarker, e.g., via an audio calibration task, to ensure that recordings are of acceptable quality and will yield reliable results.

Analytical Validation

Analytical validation involves checking that the measurements obtained via a digital biomarker are accurately measuring the intended phenomena [9]. For speech-based biomarkers, this requires verifying that any property or metric extracted from a speech sample, which we refer to as a feature, measures the associated aspects of speech accurately. The features that can be extracted from speech are numerous and diverse, and they vary according to the task and the processing procedures. Common features include acoustic parameters that reflect mathematical properties of the sound wave, such as fundamental frequency, shimmer, jitter, and Mel-frequency cepstral coefficients (MFCCs). Linguistic features, reflecting parts of speech, vocabulary diversity, syntactic structures, sentiment, and higher-level organization of language, can be obtained using natural language-processing tools or manual coding methods. The type of features used to create a speech-based biomarker will therefore depend on the speech task, processing method, and disease area in question. For example, assessing a motor speech impairment may be best accomplished by examining acoustic features in the context of a structured speech task, like sustained phonation. In contrast, assessing irregularities in the organization and content of speech, reflective of cognitive impairment, may require analyzing the linguistic features derived from a less structured speech task, like picture description (Fig. 1).

The elements of analytical validation will therefore vary based on the relevant features and methods used to compute a specific speech biomarker. As a first step, any speech processing, including segmenting audio based on speaker identity and transcription of the audio into text (via automated or manual methods), needs to be verified for accuracy. This can be accomplished by comparing multiple raters in the case of manual transcription, or by comparing automated to manual methods in the case of automated transcription. Speech processing must also be evaluated to determine how the properties of the recording affect its accuracy. Factors such as a speaker's age, education, diagnosis, or accent should all be tested to determine if and how they affect the accuracy of speech processing and the subsequent extraction of features. Examples of this type of analytical validation can be found in 2 recent studies comparing automatic speech recognition (ASR) to human transcription [26, 27].

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024

While neither study found ASR to be equivalent to human transcription in terms of accuracy, they both determined the accuracy of ASR in different clinical groups and compared algorithms to select the most accurate processing procedure.

Features extracted from speech vary in terms of the demands for analytical validation. Many acoustic features, such as MFCCs, are computed via mathematical transformations of the sound wave, and can therefore be validated using mathematical models of speech production. As an example of analytical validation of an acoustic feature, one study compared several pitch detection algorithms in terms of accuracy and robustness to background noise for analyzing the voice recordings of patients with Parkinson's disease, comparing against a gold standard calculation method [28]. This approach to analytical validation is, however, difficult to apply in cases in which gold standard referents do not exist.

Analytical validation can be less straightforward for higher level linguistic features such as the type and characteristics of the words used, the grammatical complexity, or markers of sentiment. Evaluation of such features often involves a comparison with judgments made by human raters or standard linguistic corpora [29, 30], but this can be expensive and time-consuming. In cases in which validation against such standards is impossible or impractical, concurrent validation may be performed by comparing speech measures to metrics obtained from other sensors, like comparing a vocal measure of fatigue to electrophysiological measurements [31], if available.

Finally, composite measures or machine-learning models based on multiple features present an even more challenging target for validation since there may not be any existing reference measures (i.e., a score reflecting word-finding difficulty could comprise features reflecting speech rate, pauses, errors, etc., and a classification model could be based on the weighted combination of hundreds of variables). In these cases, analytical validation can be carried out on the component features of the composite or model, if possible, with additional clinical validation of the overall outcome score. Analytical validation of models can be achieved by cross-validation, independent replication of results, testing the generalizability of complex speech models to new datasets, and ensuring that confounding factors do not bias the datasets. For example, analytical validation of a classification model for dementia based on speech could involve testing the model on an independent data set, ensuring that factors such as age, sex, education level, and accent were not unevenly distributed in training or testing data sets, leading to a bias in the model.

Another important aspect of analytical validation is determining the mathematical properties of values derived from a measurement. For example, examining the distribution of a measure across a population to determine if it is normally distributed, bimodal, or has significant skew will affect how this measure is interpreted. It is important to determine if there are floor or ceiling effects, and in what situations a measure may not be computed or yield missing data. Other important components of analytical validation include determining if a given measure demonstrates learning effects with repeated testing (and at what intervals), and how it may vary according to demographics such as age, sex, education level, or accent. An example of this type of analytical validation was provided in a study of language measures derived from a picture description task, which reported normative data for each language measure and the respective effects of age, sex, and education [32].

Clinical Validation

Clinical validation is the process of evaluating if a digital biomarker provides meaningful clinical information [9, 33]. For example, a digital biomarker could be used for disease diagnosis, measuring disease or symptom severity, monitoring change in disease/symptoms over time, predicting disease onset, or measuring the response to treatment or therapy [8, 9, 11]. An ideal biomarker might serve all of these functions, but some measures may be limited to

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024

Digit Biomark 2020;4:99–108
DOI: 10.1159/000510820
© 2020 The Author(s). Published by S. Karger AG, Basel
www.karger.com/dib

only one or two of these contexts and still offer significant clinical utility. To determine these types of clinical validity, a suitable clinical reference standard is necessary to define, and novel digital biomarkers should be compared against this standard using appropriate techniques, depending on whether the measures in question are binary, categorical, or continuous.

For some diseases, there may be a clear gold standard reference measure, such as a diagnostic test; however, for many diseases and disorders, the existing measures are themselves surrogate endpoints. For example, current gold standards for a diagnosis of Alzheimer's disease include the detection of amyloid pathology via the testing of cerebrospinal fluid, PET scan, or autopsy, all of which are invasive and expensive, and are therefore not practical for use in many trials [8]. As a result, a variety of neurological, cognitive, and functional assessments are used as surrogate endpoints in clinical research and practice [8]. Ultimately, the choice of reference measure limits the clinical validation, since a measure can only be shown to be as good as the measure used to validate it. Especially problematic are cases in which a selected reference measure has limited validity for a disease or high interrater variability, since it may be difficult to achieve consistent validity according to this measure and irregularities in the reference measure could introduce bias into the novel biomarker. Thus, it is also important to consider the validity of the reference measure used to assess clinical validity. In cases in which gold standard measures are not available, we recommend comparison with a number of surrogate measures, to provide corroborating validation checks and avoid the limitations of any single measure.

A growing body of work is highlighting the ongoing clinical validation of speech-based measures in a variety of clinical contexts. Speech has been demonstrated to have diagnostic validity for Alzheimer's disease (AD) and mild cognitive impairment (MCI) in studies using machine-learning classification models to differentiate individuals with AD/MCI from healthy individuals based on speech samples [34–41]. Additionally, speech analysis has been shown to be able to detect individuals with depression [42–45], schizophrenia [46–49], autism spectrum disorder [50], and Parkinson's disease [51, 52], and can differentiate the subtypes of primary progressive aphasia and frontotemporal dementia [53–55]. Classification models provide diagnostic validity for speech measures and could be used to develop tools for disease screening and diagnosis. In these types of diagnostic clinical validation studies, it is important to report the accuracy of the classification, as well as other metrics such as sensitivity and specificity. It is also valuable to compare performance with current clinical standards, both for distinguishing the disease from healthy controls and from related diagnoses, to demonstrate the utility of a novel measure. In addition, when possible, exploring what variables drive classification models can help with clinical interpretability, by providing a better understanding of the symptoms and changes that accompany a disease. Interpreting the variables that contribute to a model can help guard against "black box" models and detect artifacts in the data. For example, if a disease is more prevalent in women, and women tend to have higher-pitched voices, the finding that vocal pitch was driving a classification model may only reflect the higher incidence of that disease in women. For diagnostic validation, we therefore recommend reporting both model accuracy and the features that contribute to classification.

Other forms of clinical validation of speech-based measures include measuring disease severity and tracking changes over time as a measure of disease progression, prognosis, or the response to treatment. Various studies have shown how speech measures relate to disease severity by demonstrating associations between speech features and the presence of pathology in primary progressive aphasia [56] as well as between clinician-rated symptoms and speech features in depression and schizophrenia [57–59]. Several longitudinal case studies suggest that speech features may have prognostic validity for predicting the onset of Alzheimer's disease and show changes in years prior to the diagnosis [60–65]. This research requires further validation in more general populations. We highlight the need for the

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024

collection of longitudinal speech data in clinical populations and the comparison of these measures with current clinical standards. Validated measures for tracking speech changes over time could provide continuous measurement of disease risk, more frequent assessment of disease progression, and better detection of response to treatment.

Conclusion

This article summarizes recommended approaches to the evaluation of digital biomarkers in the context of recent research into speech-based measures in neurological diseases and psychiatric disorders. Speech biomarkers potentially offer many advantages for clinical research and practice; they are objective, naturalistic, can be collected remotely, and require minimal instruction and time compared to current clinical standards. Examples from the literature illustrate the active research in this area, offering promising results for the development of speech-based measures as biomarkers in multiple disease areas including Alzheimer's disease, Parkinson's disease, frontotemporal dementia, depression, and schizophrenia. While these findings demonstrate the potential utility of speech-based biomarkers, it is important to note that the particular speech features analyzed differ widely across studies, and no speech measure has yet been comprehensively evaluated across all 3 categories, i.e., verification, analytical validation, and clinical validation. While the measures under development can provide exploratory insights for clinical research, it is necessary to continue to rigorously evaluate speech-based digital biomarkers to achieve valid surrogate endpoints for use in clinical research and practice. The recommendations in this paper are targeted for speech-based biomarkers, but they generalize to other novel digital biomarkers and can be broadly applied.

Conflict of Interest Statement

J.R., L.D.K., W.S., and M.Y. are employees of Winterlight Labs. J.E.H. reports receipt of personal fees in the past 3 years from AlzeCure, Aptinyx, Astra Zeneca, Athira Therapeutics, Axon Neuroscience, Axovant, Biogen Idec, BlackThornRx, Boehringer Ingelheim, Cerecin, Cognition Therapeutics, Compass Pathways, CRF Health, Curasen, EIP Pharma, Eisai, FSV7, G4X Discovery, GfHEU, Heptares, Lundbeck, Lysosome Therapeutics, MyCognition, Neurocentria, Neurocog, Neurodyn Inc, Neurotrack, Novartis, Nutricia, Probiodrug, Regeneron, Rodin Therapeutics, Samumed, Sanofi, Servier, Signant, Syndesi Therapeutics, Takeda, Vivoryon Therapeutics, vTv Therapeutics, and Winterlight Labs. Additionally, he holds stock options in Neurotrack Inc. and is a joint holder of patents with My Cognition Ltd. F.R. is an employee of Surgical Safety Technologies.

Funding Sources

F.R. is supported with a CIFAR Chair in Artificial Intelligence.

Author Contributions

J.R. drafted the original manuscript. J.E.H., L.D.K, F.R., W.S., and M.Y. all substantively revised the manuscript. All authors approved the submitted version.

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024

References

1 Babrak LM, Menetski J, Rebhan M, Nisato G, Zinggeler M, Brasier N, et al. Traditional and Digital Biomarkers: Two Worlds Apart? Digit Biomark. 2019 Aug;3(2):92–102.
2 Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med. 2019;2(1):14.
3 Sabbagh MN, Boada M, Borson S, Chilukuri M, Doraiswamy PM, Dubois B, et al. Rationale for Early Diagnosis of Mild Cognitive Impairment (MCI) supported by Emerging Digital Technologies. J Prev Alzheimers Dis. 2020; 7(3):158–64.
4 Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity. NPJ Digit Med. 2019;2(1):9.
5 Piau A, Wild K, Mattek N, Kaye J. Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review. J Med Internet Res. 2019 Aug;21(8):e12785.
6 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. 2018 May;4(1):234–42.
7 Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010 Nov;5(6):463–6.
8 Frisoni GB, Boccardi M, Barkhof F, Blennow K, Cappa S, Chiotis K, et al. Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. Lancet Neurol. 2017 Aug;16(8):661–76.
9 Goldsack J, Coravos A, Bakker J, Bent B, 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 Apr 14;3:55.
10 Leptak C, Menetski JP, Wagner JA, Aubrecht J, Brady L, Brumfield M, et al. What evidence do we need for biomarker qualification? Sci Transl Med. 2017 Nov;9(417):eaal4599.
11 Puntmann VO. How-to guide on biomarkers: biomarker definitions, validation and applications with examples from cardiovascular disease. Postgrad Med J. 2009 Oct;85(1008):538–45.
12 Boschi V, Catricalà E, Consonni M, Chesi C, Moro A, Cappa SF. Connected Speech in Neurodegenerative Language Disorders: A Review. Front Psychol. 2017 Mar;8:269.
13 Cohen AS, Elvevåg B. Automated computerized analysis of speech in psychiatric disorders. Curr Opin Psychiatry. 2014 May;27(3):203–9.
14 Cohen AS, McGovern JE, Dinzeo TJ, Covington MA. Speech deficits in serious mental illness: a cognitive resource issue? Schizophr Res. 2014 Dec;160(1-3):173–9.
15 Poole ML, Brodtmann A, Darby D, Vogel AP. Motor Speech Phenotypes of Frontotemporal Dementia, Primary Progressive Aphasia, and Progressive Apraxia of Speech. J Speech Lang Hear Res. 2017 Apr;60(4):897–911.
16 Szatloczki G, Hoffmann I, Vincze V, Kalman J, Pakaski M. Speaking in Alzheimer's Disease, Is That an Early Sign? Importance of Changes in Language Abilities in Alzheimer's Disease. Front Aging Neurosci. 2015 Oct;7:195.
17 Coravos A, Goldsack JC, Karlin DR, Nebeker C, Perakslis E, Zimmerman N, et al. Digital Medicine: A Primer on Measurement. Digit Biomark. 2019 May;3(2):31–71.
18 Coravos A, Doerr M, Goldsack J, Manta C, Shervey M, Woods B, et al. Modernizing and designing evaluation frameworks for connected sensor technologies in medicine. NPJ Digit Med. 2020 Mar;3(1):37.
19 Jannetts S, Schaeffler F, Beck J, Cowen S. Assessing voice health using smartphones: bias and random error of acoustic voice parameters captured by different smartphone types. Int J Lang Commun Disord. 2019 Mar; 54(2):292–305.
20 Manfredi C, Lebacq J, Cantarella G, Schoentgen J, Orlandi S, Bandini A, et al. Smartphones Offer New Opportunities in Clinical Voice Research. J Voice. 2017 Jan;31(1):111.e1–7.
21 Maryn Y, Ysenbaert F, Zarowski A, Vanspauwen R. Mobile Communication Devices, Ambient Noise, and Acoustic Voice Measures. J Voice. 2017 Mar;31(2):248.e11–23.
22 Uloza V, Padervinskis E, Vegiene A, Pribuisiene R, Saferis V, Vaiciukynas E, et al. Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening. Eur Arch Otorhinolaryngol. 2015 Nov;272(11):3391–9.
23 Vogel AP, Rosen KM, Morgan AT, Reilly S. Comparability of modern recording devices for speech analysis: smartphone, landline, laptop, and hard disc recorder. Folia Phoniatr Logop. 2014;66(6):244–50.
24 Rusz J, Hlavnicka J, Tykalova T, Novotny M, Dusek P, Sonka K, et al. Smartphone Allows Capture of Speech Abnormalities Associated with High Risk of Developing Parkinson's Disease. IEEE Trans Neural Syst Rehabil Eng. 2018 Aug;26(8):1495–507.
25 Lebacq J, Schoentgen J, Cantarella G, Bruss FT, Manfredi C, DeJonckere P. Maximal Ambient Noise Levels and Type of Voice Material Required for Valid Use of Smartphones in Clinical Voice Research. J Voice. 2017 Sep; 31(5):550–6.
26 Jacks A, Haley KL, Bishop G, Harmon TG. Automated Speech Recognition in Adult Stroke Survivors: Comparing Human and Computer Transcriptions. Folia Phoniatr Logop. 2019;71(5-6):286–96.
27 Pakhomov SV, 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.

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024

28 Illner V, Sovka P, Rusz J. Validation of freely available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease. Biomed Signal Process Control. 2020 Apr; 58:101831.
29 Provoost S, Ruwaard J, van Breda W, Riper H, Bosse T. Validating Automated Sentiment Analysis of Online Cognitive Behavioral Therapy Patient Texts: An Exploratory Study. Front Psychol. 2019 May;10:1065.
30 Rudkowsky E, Haselmayer M, Wastian M, Jenny M, Emrich Š, Sedlmair M. More than Bags of Words: Sentiment Analysis with Word Embeddings. Commun Methods Meas. 2018 Apr;12(2–3):140–57.
31 Dhupati LS, Kar S, Rajaguru A, Routray A. A novel drowsiness detection scheme based on speech analysis with validation using simultaneous EEG recordings. IEEE International Conference on Automation Science and Engineering; 2010; Toronto (ON).
32 Forbes-McKay KE, Venneri A. Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. Neurol Sci. 2005 Oct;26(4):243–54.
33 Allinson JL. Clinical biomarker validation. Bioanalysis. 2018 Jun;10(12):957–68.
34 Asgari M, Kaye J, Dodge H. Predicting mild cognitive impairment from spontaneous spoken utterances. Alzheimers Dement (N Y). 2017 Feb;3(2):219–28.
35 Noorian Z, Pou-Prom C, Rudzicz F. [Internet] On the importance of normative data in speech-based assessment. arXiv171200069 Cs. [cited 2017 Nov 30]. Available from: http://arxiv.org/abs/1712.00069
36 Fraser KC, Meltzer JA, Rudzicz F. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. J Alzheimers Dis. 2016;49(2):407–22.
37 Konig A, Satt A, Sorin A, Hoory R, Derreumaux A, David R, et al. Use of Speech Analyses within a Mobile Application for the Assessment of Cognitive Impairment in Elderly People. Curr Alzheimer Res. 2018;15(2):120–9.
38 Toth L, Hoffmann I, Gosztolya G, Vincze V, Szatloczki G, Banreti Z, et al. A Speech Recognition-Based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech. Curr Alzheimer Res. 2018;15(2):130–8.
39 Themistocleous C, Eckerström M, Kokkinakis D. Voice quality and speech fluency distinguish individuals with mild cognitive impairment from healthy controls. PLoS One. 2020;15(7):e0236009.
40 Pulido ML, Hernández JB, Ballester MÁ, González CM, Mekyska J, Smékal Z. Alzheimer's disease and automatic speech analysis: a review. Expert Syst Appl. 2020;150(Jan):113213.
41 Gosztolya G, Tóth L, Grósz T, Vincze V, Hoffmann I, Szatlóczki G, et al. Detecting Mild Cognitive Impairment from Spontaneous Speech by Correlation-Based Phonetic Feature Selection. Proceedings of the 17th Annual Conference of INTERSPEECH; 2016 Sept 8-12; San Francisco (CA).
42 Low LS, Maddage NC, Lech M, Sheeber LB, Allen NB. Detection of clinical depression in adolescents' speech during family interactions. IEEE Trans Biomed Eng. 2011 Mar;58(3):574–86.
43 Pan W, Flint J, Shenhav L, Liu T, Liu M, Hu B, et al. Re-examining the robustness of voice features in predicting depression: compared with baseline of confounders. PLoS One. 2019;20;14(6):e0218172.
44 Taguchi T, Tachikawa H, Nemoto K, Suzuki M, Nagano T, Tachibana R, et al. Major depressive disorder discrimination using vocal acoustic features. J Affect Disord. 2018 Jan;225:214–20.
45 Wang J, Zhang L, Liu T, Pan W, Hu B, Zhu T. Acoustic differences between healthy and depressed people: a cross-situation study. BMC Psychiatry. 2019 Oct;19(1):300.
46 Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 2015 Aug;1(1):15030.
47 Corcoran CM, Carrillo F, Fernández-Slezak D, Bedi G, Klim C, Javitt DC, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry. 2018 Feb;17(1):67–75.
48 Mota NB, Copelli M, Ribeiro S. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. NPJ Schizophr. 2017 Apr;3(1):18.
49 Mota NB, Vasconcelos NA, Lemos N, Pieretti AC, Kinouchi O, Cecchi GA, et al. Speech Graphs Provide a Quantitative Measure of Thought Disorder in Psychosis. PLoS One. 20129;7(4):e34928.
50 Cho S, Liberman M, Ryant N, Cola M, Schultz RT, Parish-Morris J. Automatic Detection of Autism Spectrum Disorder in Children Using Acoustic and Text Features from Brief Natural Conversations. Proceedings of the Annual Conference of INTERSPEECH; 2019 Sept 15-19; Graz, Austria.
51 García AM, Carrillo F, Orozco-Arroyave JR, Trujillo N, Vargas Bonilla JF, Fittipaldi S, et al. How language flows when movements don't: an automated analysis of spontaneous discourse in Parkinson's disease. Brain Lang. 2016 Nov;162:19–28.
52 Wroge TJ, Ozkanca Y, Demiroglu C, Si D, Atkins DC, Ghomi RH. Parkinson's Disease Diagnosis Using Machine Learning and Voice. IEEE SPMB; 2018 Dec 1; Philadelphia (PA).
53 Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex. 2014 Jun;55:43–60.
54 Jarrold W, Peintner B, Wilkins D, Vergryi D, Richey C, Gorno-Tempini ML, et al. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; 2014 June; Baltimore (MD).
55 Peintner B, Jarrold W, Vergyri D, Richey C, Tempini ML, Ogar J. Learning diagnostic models using speech and language measures. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2008; Vancouver (BC).

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024

56 Nevler N, Ash S, Irwin DJ, Liberman M, Grossman M. Validated automatic speech biomarkers in primary progressive aphasia. Ann Clin Transl Neurol. 2018 Nov;6(1):4–14.
57 Minor KS, Willits JA, Marggraf MP, Jones MN, Lysaker PH. Measuring disorganized speech in schizophrenia: automated analysis explains variance in cognitive deficits beyond clinician-rated scales. Psychol Med. 2019 Feb;49(3):440–8.
58 Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. Biol Psychiatry. 2012 Oct;72(7):580–7.
59 Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J Neurolinguist. 2007 Jan;20(1):50–64.
60 Berisha V, Wang S, LaCross A, Liss J. Tracking discourse complexity preceding Alzheimer's disease diagnosis: a case study comparing the press conferences of Presidents Ronald Reagan and George Herbert Walker Bush. J Alzheimers Dis. 2015;45(3):959–63.
61 Garrard P, Maloney LM, Hodges JR, Patterson K. The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. Brain. 2005 Feb;128(Pt 2):250–60.
62 Le X, Lancashire I, Hirst G, Jokel R. Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. Lit Linguist Comput. 2011 Dec;26(4):435–61.
63 Riley KP, Snowdon DA, Desrosiers MF, Markesbery WR. Early life linguistic ability, late life cognitive function, and neuropathology: findings from the Nun Study. Neurobiol Aging. 2005 Mar;26(3):341–7.
64 Snowdon DA. Aging and Alzheimer's disease: lessons from the Nun Study. Gerontologist. 1997 Apr;37(2):150–6.
65 Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, Markesbery WR. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. JAMA. 1996 Feb;275(7):528–32.

Downloaded from http://karger.com/dib/article-pdf/4/3/99/2575454/000510820.pdf by guest on 14 June 2024