Article:
O'Malley, R.P.D., Mirheidari, B., Harkness, K. et al. (5 more authors) (2021) Fully automated cognitive screening tool based on assessment of speech and language. Journal of Neurology, Neurosurgery & Psychiatry, 92 (1). pp. 12-15. ISSN 0022-3050

https://doi.org/10.1136/jnnp-2019-322517

A fully automated cognitive screening tool based on assessment of speech and language

Ronan O'Malley[1&2], Bahman Mirheidari[3], Kirsty Harkness[2], Markus Reuber [1&2], Annalena Venneri[1], Traci Walker[4], Heidi Christensen[3] and Daniel Blackburn[1&2]

1. Department of Neuroscience, University of Sheffield, Sheffield, United Kingdom

2. Sheffield Teaching Hospitals NHS Foundation Trust, Department of Neurology.

3. Dept of Computer Science, University of Sheffield, Sheffield, United Kingdom

4. Human Communication Sciences, University of Sheffield, Sheffield, UK

Corresponding author

Daniel J. Blackburn, MBChB, BSc, PhD, Sheffield Institute of Translational Neuroscience (SITraN), University of Sheffield, 385A Glossop Road, Sheffield S10 2HQ, UK

e-mail: d.blackburn@sheffield.ac.uk

Tel 0114 2159101

Abstract

Introduction

The dramatic recent rise in referrals to specialist memory clinics has been associated with an increased proportion of patients referred with Functional Memory Disorder (FMD), i.e. non-progressive cognitive complaints. These referrals have exerted time and financial pressures on secondary care services, impairing their ability to deliver high-quality care for patients with neurodegenerative cognitive disorders. We have developed a fully automated system, "CognoSpeak", which enables risk stratification at the primary-secondary care interface and ongoing monitoring of patients with memory concerns.

Methods

We recruited 15 participants to each of four groups: Alzheimer's disease (AD), Mild Cognitive Impairment (MCI), FMD and healthy controls. Participants responded to 12 questions posed by a computer-presented talking head. Automatic analysis of the audio and speech data involved speaker segmentation, automatic speech recognition and machine learning classification.

Results

CognoSpeak could distinguish between participants in the AD or MCI groups and those in the FMD or healthy control groups with a sensitivity of 86.7%. Patients with MCI were identified with a sensitivity of 80%.

Discussion

Our fully automated system achieved levels of accuracy comparable to currently available, manually administered assessments. Greater accuracy should be achievable through further system training with a greater number of users, the inclusion of verbal fluency tasks and mood assessments. The current data supports CognoSpeak's promise as a screening and monitoring tool for patients with MCI. Pending confirmation of these findings, it may allow clinicians to offer patients at low risk of dementia earlier reassurance and relieve pressures on specialist memory services.

Introduction

Memory clinics assess patients with a variety of cognitive complaints disorders, including those related to Alzheimer's disease (AD), possible prodromal states (mild cognitive impairment, MCI) and those with Functional Cognitive Disorder (FCD, disabling but non-progressive cognitive complaints associated with emotional or psychological factors).

Early referral to an appropriate care pathway yields benefits for patient wellbeing and efficient resource allocation. Specialist memory clinics are a limited resource and patients with FCD can be successfully managed in other settings. Accurate pre-clinic

stratification tools are required to direct patients towards the most appropriate service.

AD is associated with subtle impairments in language that may precede deficits in episodic memory by decades (1). Previous work has shown that qualitative analysis of conversational profiles inspired by the methodology of Conversation Analysis (CA) can discriminate between patients with FCD and those with neuro-degenerative (ND) conditions (2). However, this approach depends on highly trained experts and is not easily scalable.

Whilst the use of automated speech analysis has been explored previously to identify cognitive impairment, most studies do not describe fully automated solutions. Instead, they rely on the automated analysis of data collected from human-human interaction (3) or used manually generated transcripts (4).

We have created a fully automated stratification tool. "CognoSpeak" consists of a virtual clinician, a computer screen-presented talking head, which asks questions and records the patients' spoken responses. The system uses Automatic Speech Recognition (ASR) and diarisation (segmenting the recording into contributions from different speakers) to extract acoustic and linguistic measures that are used by machine learning classifiers to select the most likely diagnostic category. (5).

Methods

We recruited 60 participants; 15 each from four different diagnostic groups: AD, MCI, FCD and healthy controls (HC). HCs were recruited via the "University of the Third Age" and patient participants from a specialist memory clinic in Sheffield, between May 2016 and January 2019. Patients could be recorded on their own or in the presence of an accompanying person. Ethical permission was granted by the NRES Committee South West-Central Bristol (Rec number 16/ LO/0737) in May 2016. Neurological diagnoses were made according to standard diagnostic criteria[1] and after multidisciplinary team review. Presence of significant mood disturbance (ascertained through clinical history and Patient Health Questionnaire-9 (PHQ-9 > 15) and significant cerebrovascular disease resulted in exclusion. All underwent cognitive assessment using Addenbrooke's Cognitive Examination-Revision (ACE-R) tool or detailed neuropsychological evaluation. Brain imaging (including CT, MRI and Tc99m HMPAO single-photon emission computed tomography) was performed based on clinical need. A proportion of the healthy control group had MRI as part of their involvement in a previous study (VPH-DARE@IT http://www.vph-dare.eu/).

The CognoSpeak assessment process has been comprehensively described in prior papers (5). In brief, Participants were directed verbally to respond to the questions posed by the virtual clinician and to use the "enter" key to proceed from one question to the next. Audio data were recorded using a Tascam DR-40 recorder.

[1] For AD, McKhann et al Alzheimer's & Dementia, 2011; for MCI Petersen et al Archives of Neurology 1999, Petersen Archives of Neurology 2001; for FCD Schmidtke et al American Journal of Geriatric Psychiatry, 2008

An ASR system was used to transcribe the audio into a string of words. A diarisation tool was then applied to provide annotation of which words were said by which speaker. The combined output of ASR and diarisation provided information on the content of the speech and the duration of the contributions of different speakers. The automatically transcribed text as well as the recorded audio were used to extract the range of features that the machine learning based classifier uses to assign participants to a diagnostic class.

72 features were extracted from the speech of patients and accompanying persons, including 17 CA -inspired features, 24 acoustic-only, 24 lexical-only, and 7 word vector features.

Results

Please see Table 1 for demographic variables, psychological measures performed and the results and a description of the imaging.

Table 1

Characteristics	HC	FCD	MCI	AD
Gender (% male)	40	40	66.7	66.7
Age (Mean)	69.5 (± 4.0)	54.9 (± 4.1)	63.4 (± 4.2)	67.8 (± 4.2) †
Years of Education (Mean) *	18.1 (± 1.0)	16.4 (±0.6)	17.3 (±1.1)	18 (±1.6)‡
Neuropsychology Performed (%)	100	93.3	100	100
ACE (%)	100	85.7	86.7	93.3
MMSE (%)		14.3	13.3	6.7
ACE Score (Mean)	95.3 (± 1.6)	88.7 (± 5.9)	81.3 (± 3.6)	69.5 (± 8.2)*
MMSE Score (Mean)		27.3 (± 1.9)	27.0 (± 0.8)	23 (± 3.3)
Structural Imaging Performed (%)	40	86.8	100	100
Abnormal (%)	0	20¥	20	66.7
Consistent with ND (%)	0	0	13.3	26.7
SPECT Performed (%)	0	0	46.7	66.7
Consistent with ND (%)	0	0	85.7	70

† Age: HC vs FCD, HC vs MCI, FCD vs MCI, FCD vs AD p <0.05. HC vs AD and MCI vs AD p >0.05

‡ Years of Education: HC vs FCD and FCD vs AD p <0.05. HC vs MCI, HC vs AD, FCD vs MCI and MCI vs AD p <0.05

* ACE Score: HC vs MCI, HC vs AD, FCD vs MCI, FCD vs AD and MCI vs AD p< 0.05. HC vs FCD p >0.05.

[¥] 3 Abnormal MRI scans demonstrating a previous contusional injury, generalised atrophy and mild small vessel disease.

ND = Neurodegeneration

The mean duration of interactions across all participant groups and CognoSpeak was 11 minutes 24 seconds. HCs spent an average of 9 minutes 58 seconds; FCD patients 11 minutes 2 seconds; MCI patients 9 minutes 34 seconds and AD patients 15 minutes 2 seconds. Only the difference between the MCI vs AD groups was significant (U=59, P=0.026).

Two-way automatic classification (AD & MCI v FMD & HC)

An effective cognitive stratification tool must be able to separate those with potential neurodegeneration from those without. In the two-way classification CognoSpeak system had an accuracy of identifying participants with MCI or AD of 87% (see Figure 1a) whilst the accuracy of correctly allocating participants as either HC or FMD was 77%.

Three-way automatic classification (AD v MCI v FMD & HC)

Overall correct classification was achieved in 65% of cases. The accuracy of identifying participants with AD was 80% (see Figure 1b). Two participants were incorrectly allocated as MCI and one as belonging to the FMD and HC group.

The identification accuracy of participants with MCI was 80%. One participant was incorrectly allocated as AD and two as FMD and HC. The accuracy of identifying the participants with either HC or FMD was 50%. Eight participants were incorrectly allocated as MCI and seven as AD.

Four-way automatic classification (AD v MCI v FMD v HC)

The more difficult task of identifying participants from all four groups revealed an overall classification accuracy of 60%. Accuracy in identifying AD participants was 80%; accuracy in identifying MCI participants was 60%; accuracy in identifying FCD participants was 47% and accuracy in identifying the HC participants was 53% (Figure 1-c), (of note the most frequent misclassifications occurred between the HC and FCD groups).

Discussion

The UK's National Dementia Strategy emphasises the importance of early diagnosis and provision of support for patients with progressive cognitive disorders. Hence, our approach has prioritised sensitivity over specificity. The two-way classification system achieved a sensitivity of 86.7% for neurodegenerative memory disorders, comparing favourably with most commonly employed screening tools. The specificity was slightly lower at 76.7% but given the emphasis on early identification and the fact that "false positives" will be investigated by the specialist memory services, this is acceptable. Using a three-way classification MCI, AD & FCD plus HC) CognoSpeak was able to identify participants with MCI with a sensitivity of 80.0% (95% CI: 51.9-95.7) and a specificity of 77.8% (95% CI: 62.9-88.8), exceeding sensitivity (66.34%) and specificity (72.94%) of the MMSE in discerning between MCI and HCs (6).

Accurate identification of patients with MCI may allow early intervention and facilitate participation in research studies. We are unaware of any other screening tool that has included patients with FCD in their development. However, the inclusion of this patient group is essential in validation studies of screening or stratification tools as those with FCD make up 24% of referrals to specialist memory services (7). The ability of our tool to distinguish between FCD and MCI / AD rather than only between HC and MCI/AD groups increases its ecological validity. The CognoSpeak 2-way classification system can distinguish between patients with MCI / AD and those without neurodegenerative pathology (HC & FCD) with a sensitivity of 86.7%.

Typical pre-clinic cognitive screening tools take between 5-10 minutes and require clinician time (8, 9). CognoSpeak compares favourably to this. A large 2016 systematic review described mean primary care physician consultation lengths across 67 countries. The group reported that 18 countries, accounting for approximately 50% of the world's population, had mean GP consultation lengths of 5 minutes or less (10). This suggests that even the most concise, traditional cognitive screening tool may not be applicable to a significant proportion of the world's population. CognoSpeak has the advantage of being automated, capable of remote administration via a tablet in a primary care office or at home using a secure website, without the necessity for direct clinician oversight. It has been estimated that in 2019, 58.8% of the global population have access to the internet (11). Results from at-home screening could be transferred and analysed by local secondary care memory services, thus reducing primary care time. By identifying FCD this tool could be used to provide early reassurance and reduce the need for unnecessary and stressful memory clinic consultations. Furthermore, the results/reporting from CognoSpeak could triage those most at risk, plan scans before consultation and reduce the time needed for initial testing in specialist clinics. This may be especially important due to the accelerated use of telephone and virtual clinics, during the Covid-19 pandemic. Furthermore, the capacity of CognoSpeak to distinguish between patients with AD and MCI raises the possibility that this automated tool could be used for longitudinal monitoring of patients with MCI.

We acknowledge a number of limitations to this study. The relatively small sample size was limited by the time taken to recruit well described participants in each cohort. A larger cohort would provide more accurate information into the relative

sensitivity and specificity of CognoSpeak. This analysis used the first 9 of 12 questions posed. The remaining three questions, verbal fluency tasks and a picture description tasks, are extensively used in clinical practice, with good accuracy in detecting MCI. These will likely contribute to increasing accuracy of CognoSpeak. They have not been included in the current study as novel automated approaches to their analysis are still under development. Inherent to the iterative nature of the machine learning process, we anticipate attaining greater accuracy with access to larger sample sizes. This applies both to the accuracy of the automated speech recognition and the refinement of classifiers.

For this initial validation study we have limited recruitment to patients belonging to the diagnostic groups most commonly represented in specialist memory clinics. Future studies will also include patient groups with non-AD dementias.

In conclusion, CognoSpeak is a fully automated cognitive screening system that can discern between normal cognition and neurodegenerative memory disorders with sensitivity comparable to traditional screening methods.

Funding

This work was supported in part by MRC CiC grant.

Acknowledgments

This is a summary of independent research carried out at the NIHR Sheffield Biomedical Research Centre (Translational Neuroscience). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

We would like to thank all participants and care partners who took part in this study. We would like to thank all participating clinicians working in the memory clinic (Aijaz Khan). We would also like to thank students from the clinical neurology masters course (Imke Meyer, Casey Rutten, Thomas Swainston, Eric Brook and Lee-Anne Morris) and medical students (Isobel Sonksen, Alice Lewis) from the University of Sheffield for their contributions to data collection.

References

1. Amieva H, Le Goff M, Millet X, Orgogozo JM, Peres K, Barberger-Gateau P, et al. Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. Ann Neurol. 2008;64(5):492-8.
2. Reuber M, Blackburn DJ, Elsey C, Wakefield S, Ardern KA, Harkness K, et al. An Interactional Profile to Assist the Differential Diagnosis of Neurodegenerative and Functional Memory Disorders. Alzheimer Dis Assoc Disord. 2018.
3. Konig A, Satt A, Sorin A, Hoory R, Toledo-Ronan O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2015;1:112-24.

4. Tanaka H, Adachi H, Ukita N, Ikeda M, Kazui H, Kudo T, et al. Detecting Dementia Through Interactive Computer Avatars. IEEE J Transl Eng Health Med. 2017;5:2200111.
5. Al-Hameed S, Benaissa M, Christensen H, Mirheidari B, Blackburn D, Reuber M. A new diagnostic approach for the identification of patients with neurodegenerative cognitive complaints. PLOS ONE. 2019;14(5):e0217388.
6. Ciesielska N, Sokolowski R, Mazur E, Podhorecka M, Polak-Szabela A, Kedziora-Kornatowska K. Is the Montreal Cognitive Assessment (MoCA) test better suited than the Mini-Mental State Examination (MMSE) in mild cognitive impairment (MCI) detection among people aged over 60? Meta-analysis. Psychiatr Pol. 2016;50(5):1039-52.
7. McWhirter L, Ritchie C, Stone J, Carson A. Functional cognitive disorders: a systematic review. Lancet Psych. 2020;7(2): 191-207.
8. Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: A systematic review for the U.S. Preventive Services Task Force. Annals of internal medicine. 2013;159(9):601-12.

9. Ballard C B BA, Corbett A, Rasmussen J. Helping you to assess cognition: A practical toolkit for clinicians. In: Health Do, editor. London2015.
10. Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open. 2017;7(10):e017902.
11. Stats IW. Internet World Stats 2019 [Available from: https://www.internetworldstats.com/stats.htm.