An interactional profile to assist the differential diagnosis of neurodegenerative and functional memory disorders

Revised Manuscript

Markus Reuber[**)1], **Daniel Blackburn**[**)1], Chris Elsey[2], Sarah Wakefield[1], Kerry Ardern[1], Kirsty Harkness[3], Annalena Venneri[1], Danielle Jones[4], Chloe Shaw[5], Paul Drew[6]

*): Corresponding author (Daniel Blackburn)
**): Markus Reuber and Daniel Blackburn contributed in equal measure to this study.

1) Department of Neuroscience, University of Sheffield, Sheffield, UK.
2) School of Allied Health Sciences, De Montfort University, Leicester, UK.
3) Department of Neurology, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK.
4) School of Dementia Studies, University of Bradford UK.
5) UCL Institute for Women's Health, London, UK.
6) Department of Social Sciences, Loughborough University, Loughborough, UK.

:

Corresponding Author:

*** Daniel Blackburn**, Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, 385A Glossop Road, Sheffield, S10 2HQ. **Telephone**: 0114 222 2267 **Email:** d.blackburn@sheffield.ac.uk

Other authors:

Markus Reuber, Academic Neurology Unit, University of Sheffield, Royal Hallamshire Hospital, Glossop Road, Sheffield, S10 2JF. **Telephone**: 0114 226 8688 **Email**: markus.reuber@sth.nhs.uk

Chris Elsey, School of Allied Health Sciences, De Montfort University, The Gateway, Leicester, LE1 9BH.**Email:** christopher.elsey@dmu.ac.uk

Sarah Wakefield, Department of Neuroscience, Medical School, University of Sheffield, Sheffield, S10 2TN. **Email**: s.wakefield@sheffield.ac.uk

Kirsty Harkness, Department of Neurology, Royal Hallamshire Hospital, Glossop Road, Sheffield, S10 2JF. **Email**: kirsty.harkness@sth.nhs.uk

Kerry Ardern, Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, 385A Glossop Road, Sheffield, S10 2HQ. kaardern1@sheffield.ac.uk

Annalena Venneri, Department of Neuroscience, Medical School, University of Sheffield, Beech Hill Road, Sheffield, S10 2RX. Telephone: 0114 2713430. Email: a.venneri@sheffield.ac.uk

Danielle Jones, School of Dementia Studies**,** University of Bradford**,** Bradford, BD7 1DP. **Telephone**: 01274 236469 **Email**: d.k.jones1@bradford.ac.uk

Chloe Shaw, Institute for Women's Health, University College London, 89-96 Chenies Mews, London WC1 6HX. **Telephone**: 020 7679 6036 **Email**: chloe.shaw@ucl.ac.uk

Paul Drew, Department of Social Sciences, Loughborough University, Epinal Way, Leicestershire, LE11 3TU. **Telephone**: 01509 223365 **Email**: p.drew@lboro.ac.uk

Objective: Specialist services for dementia are seeing an increasing number of patients. We investigated whether interactional and linguistic features in the communication behaviour of patients with memory problems could help distinguish between those with problems secondary to neurological disorders (ND) and those with Functional Memory Disorder (FMD).

Methods: In Part 1 of this study, a Diagnostic Scoring Aid (DSA) was developed encouraging linguists to provide quantitative ratings for 14 interactional features. An optimal cut-off differentiating ND and FMD was established by applying the DSA to 30 initial patient–doctor memory clinic encounters. In Part 2, the DSA was tested prospectively in ten additional cases analysed independently by two Conversation Analysts blinded to medical information.

Results: In part one, the median score of the DSA was +5 in ND and -5 in FMD ($p<0.001$). The optimal numeric DSA cut off (+1) identified patients with ND with a sensitivity of 86.7% and a specificity of 100%. In part two, DSA scores of rater one correctly predicted 10/10 and those of rater two 9/10 diagnoses.

Conclusions This study indicates that interactional and linguistic features can help distinguish between patients developing dementia and those with FMD and could aid the stratification of patients with memory problems.

INTRODUCTION

Demographic changes have increased pressure on specialist services for patients with dementia, causing healthcare professionals and service commissioners in many countries to focus on improvements to diagnostic pathways for people with memory complaints. In the UK, the National Dementia Strategy identified the closure of the 'dementia gap' (the difference between the predicted number of people with dementia versus those diagnosed as having dementia) as an area of particular concern [1]. An audit by the Royal College of Psychiatrists found that the number of people assessed in specialist clinics with memory concerns increased fourfold between 2010 and 2013[2]. A further 31% increase was seen from 2013 to 2014. While the 'dementia gap' has narrowed with this increase in activity, it has not been reduced at the same rate at which the number of memory clinic referrals has risen [3, 4]. One reason for this is that the proportion of patients with functional memory disorder (FMD) or other non-progressive memory disorders has also increased [3, 5, 6]. The referral of patients to memory clinics is costly and can cause avoidable distress [7, 8]. These observations – combined with studies showing that current screening procedures lack sensitivity[9] – suggest that case selection for referral to specialist clinics is suboptimal.

It is well recognised that patients exhibit linguistic impairments and deficits in spontaneous speech even in the earlier stages of dementia [10, 11]. Language impoverishment, through grammatical simplification, loss of vocabulary, semantic paraphasias, and overuse of semantically empty words, becomes progressively evident in dementia, as do impaired semantic processing and classification errors [12-15]. While the analysis of patients' language may therefore contribute towards identifying those at risk of developing dementia, the detection of such language impoverishment i) requires complex linguistic analysis, ii) may be diagnostically ambiguous, and iii) does not take account of more directly observable conversational or interactional features of language. Automated analysis of spontaneous speech could address some of the practical problems with the assessment of language in routine practice[16, 17]. However, to date, it remains uncertain how well this method would perform

as a screening procedure in clinical situations. What is more, previous approaches have focused on vocal/aural features of speech, abstracted from their communicative context. Thus, they would not capture impairments in specifically interactional capabilities, that can be detected with methods focusing on the co-construction of conversation and which may well be particularly early and specific indicators of cognitive complaints secondary to neurological disorders (ND).

We aimed to use problems with communication between patients, doctors and third parties in medical consultations as a diagnostic tool. Our project was inspired by studies demonstrating the potential of Conversation Analysis (CA)-derived interactional and linguistic observations in the differentiation of epilepsy and (non-epileptic) dissociative seizures (DS) [18]. Using previously described conversational profiles of patients with seizures, CA experts were able to predict the medical "gold standard" diagnosis of epilepsy or DS with a sensitivity and specificity of around 85% [19].

Mirroring the study design pursued in seizure clinics, Conversation Analysts have previously described a number of interactional and linguistic features, which appeared to distinguish patients with ND from those with FMD in two qualitative studies [20, 21]. The current paper describes the initial validation and assessment of a quantitative Diagnostic Scoring Aid (DSA) guiding analysts to rate each doctor-patient encounter on a number of the interactional, topical and linguistic features described in these qualitative studies. Part 1 of the present study was designed to establish an optimal discriminatory DSA cut-off for the distinction of ND and FMD. In Part 2 of this study this numeric cut-off was applied prospectively to DSA ratings of clinic interactions with newly recruited patients.

METHODS

Participant recruitment and assessment

Participant recruitment and assessment have been described in a previous article exploring the scope of automated analysis of conversational interaction[22]. Briefly, all participants had been referred to the neurology-led memory clinic in Sheffield, UK. Patients are routinely encouraged to bring someone along to their memory clinic appointment if possible (accompanying person, AP). A member of the study team obtained written informed consent prior to the encounter with a neurologist. Participants and AP were only consented if they had capacity to make their own decision about participation and used English as their first language. Participants whose diagnosis remained uncertain and those whose cognitive problems were considered to be due to other causes than ND or FMD were excluded.

Participants were investigated and followed up by Consultant Neurologists specialising in memory disorders according to clinical need. Participants were referred for detailed neuropsychological testing and MRI brain imaging. The neuropsychological battery (see Wakefield et al 2014[23] for details) included the Mini Mental State Examination[24], tests of short and long term memory (verbal and non-verbal)[25], abstract reasoning[26, 27], attention and executive function[28], language comprehension, naming by confrontation, category and letter fluency[29].

All participants were recruited before their first ever appointment in the memory clinic. Most patients with ND were in the early disease stages, but some already had moderately severe dementia. Diagnoses were reached by multidisciplinary consensus; taking into account clinical history, neurological examination, neuropsychological scores and neuro-radiological findings. Alzheimer's disease was diagnosed according to the NINCDS-ADRDA criteria[30]. A diagnosis of mixed dementia (AD plus vascular cognitive impairment) was made if moderate to severe small vessel ischaemic changes or cortical infarctions were present on MRI brain imaging. Vascular Cognitive Impairment not demented (VCIND) was used to label those with extensive radiological evidence of vascular impairment who, however, did not reach the threshold of dementia- i.e. the deterioration in function sufficient to impair activities of daily living; these patients were not included in the ND group. The

diagnosis of behavioural variant Frontotemporal Dementia (bvFTD) was made according to the Rascovsky criteria [31]. Mild Cognitive Impairment (MCI) was diagnosed according to the Petersen criteria[32]. We did not use biomarkers for amyloid or neurodegeneration (tau or FDG PET) because these tests are currently not available at our institution for routine assessment and because they are currently not widely used for clinical decision-making in the NHS. Initial diagnoses in the ND group were, however, confirmed by clinical follow-up.

The diagnosis of FMD was based on the criteria proposed by Schmidtke et al. 2008[33] with the exception of the age cut-off of <70 years. We considered this criterion overly restrictive because there have been previous reports of cases of 'functional' memory problems in people aged over 70[34, 35]. All participants with a Patient Health Questionnaire-9 (PHQ9) score of >15 (indicative of current depression) or symptoms of active moderate depression as judged by the clinician (and whose memory symptoms may have been due to depressive pseudodementia, DPD), were excluded from further analysis. Participants were also screened for Generalised Anxiety Disorder using the GAD7. However, in keeping with the criteria for the diagnosis of FMD proposed by Schmidtke et al., they were not excluded from the study on the basis of GAD7 scores. This means that some patients with FMD included in this study will have had significant problems with anxiety symptoms other than anxiety about memory.

Like those with DPD, participants in whom memory problems were found to be due to vascular cognitive impairment (VCI) not related to dementia and those for whom the diagnosis remained uncertain were not further analysed. Although VCI and DPD are important diagnostic categories, at this early stage of development we were keen to test the DSA methodology in only two homogeneous diagnostic groups.

Part 1 of this study is based on analyses of the first 15 patients with ND and the first 15 patients with FMD whose conversational data were analysable. Part 2 is based on the blinded analysis of the consecutive five next patients

with ND and next five patients with FMD who agreed to participate in this study after the recruitment for Part 1 of this study had been completed.

Data preparation for Conversation Analysis (CA)

Video or audio recordings of the history-taking phase (from a patient's entry into the neurologist's office, to the start of cognitive testing) were transcribed in detail, using a transcription system capturing the timing of speech (e.g. overlaps between speakers, pauses both within and between speaker turns), certain intonation and prosodic features of speech (falling/rising intonation, loudness, emphasis, sound stretching) [36].

Initial CA approach

Interactions with patients with ND or FMD were subjected to examination using the perspective and methods of CA. CA is a micro-analytic, qualitative method, but it is well-suited to combination with statistical measures. It has been applied widely to medical interactions in exploratory research[37, 38], in research aimed at improving the effectiveness of doctor-patient communication,[39, 40] and in assisting the diagnostic process in other conditions [19, 38]. It is particularly useful for the identification of detailed aspects of language use and communicative practices (e.g. the ways in which patients with epilepsy 'normalise' their seizure experiences, whilst patients with dissociative (nonepileptic) seizures 'catastrophise' their seizure descriptions)[19]. In their analysis, the CA experts involved in this project initially identified the methods individual patients (and accompanying others if present) used to describe memory problems to the doctor (see [20, 21] for details).

Part 1: Quantitative examination of qualitative findings

Based on the findings of the initial qualitative analysis [20, 21], we developed a diagnostic scoring aid (DSA) to provide a guide for the rating of potentially diagnostic features of communication and in order to transform qualitative observations into a numeric score (see Table 1). The DSA encourages analysts to comment and rate nine separate items. An additional five items

focus on triadic features that can only be observed if patients are accompanied during their memory clinic appointment. The DSA describes findings for each item more in keeping with ND (associated with a score of +1) or observations more in keeping with FMD (given a numeric score of -1). Items can also be judged as un-ratable and would be given a score of 0. Items could be considered un-ratable because the interactional behaviour did not take place (for instance because the neurologist had not asked the specific question (e.g. “who is more concerned about the memory problems?”), the performance provided mixed evidence, or was neither typical of that expected from patients with ND nor that of patients with FMD. Free text fields are provided for each item allowing the analyst to describe the reasoning for their categorical judgement. Finally, the DSA asks the analyst to make a qualitative judgement taking account of the whole conversation profile.

The applicability of the DSA was initially tested by one CA expert using the DSA on the recordings and transcripts of the 30 cases previously analysed in a purely qualitative way. The analyst categorised each item as more in keeping with ND, more in keeping with FMD or un-rateable. This was translated into a numeric score for each item. Finally, item scores were added up to produce a total score for each patient. An AUROC statistic was carried out based on these numeric ratings to identify an optimal diagnostic cut off score.

Part 2: Blinded analysis using a Diagnostic Scoring Aid

In order to test the discriminatory potential of the DSA, two CA experts independently rated an additional ten doctor-patient encounters (five consecutive cases in each group with ND and FMD). These cases had not been included in the initial qualitative or retrospective quantitative analyses. The Conversation Analysts were expected to predict the neurological diagnosis in these new cases on the basis of their qualitative analysis of the video recordings of the doctor-patient encounter and transcripts of these encounters, guided by the DSA. They were also asked to use the DSA to produce a numeric assessment score for each participant having provided a

score for each DSA feature. The analysts were blinded to all additional medical or demographic information about these patients. Analysts were not aware of the numeric cut-off calculated by the AUROC statistic at the time of this analysis and were encouraged not simply to base their diagnostic prediction on patients' numeric score. In their overall qualitative judgement, this enabled them to place more diagnostic emphasis on particularly outstanding features. However, in addition to the number of cases correctly diagnosed by their qualitative judgement, we also report the number of cases correctly categorised on the basis of the DSA scores using the diagnostic cut-off calculated in Part 1 of this study.

Statistical analysis

Routinely collected clinical data on consenting and non-consenting patients approached about this study were compared using t-tests to ascertain the representativeness of the patient group included.

In Part 1 of this study, the diagnostic potential of individual DSA items was examined using Fisher's Exact tests. The significance of differences in median scores of the ND and FMD groups was determined using the Mann-Whitney U test. An Area Under the Receiver Operated Characteristic Curve (AUROC) statistic was used to identify an optimal numeric cut-off for the differentiation of ND and FMD. In Part 2 of this study we report Kappa scores as a measure of the inter-rater reliability of focussed CA using the DSA.

Ethics

The study was approved by the NHS Research Ethics Committee (NRES Committee Yorkshire & The Humber - South Yorkshire). Ref 12/YH/0205.

RESULTS

Part 1

Of 353 patients referred to the specialist memory clinic during the recruitment period and of 148 eligible to take part in this study, 36 declined to participate and 112 were enrolled (see Figure 1 CONSORT diagram). Three withdrew their consent subsequently, leaving 109 who completed the study. There were

no significant differences in terms of age, gender, anxiety, depression or ACE-R scores between those who consented to take part and those who did not, suggesting that the participants were representative of the wider population served by this memory clinic. There was also no significant difference in terms of diagnostic mix between people who consented and people who did not consent (see supplementary Table 1). The ND patient group (n=20) comprised of eight patients with AD, four with amnestic MCI, two with vascular dementia, two with frontotemporal dementia, three mixed AD and vascular and one unspecified dementia (without detailed neuropsychology). Figure 1 also provides more information about participants who were excluded from the study.

Clinical details of the patients included in the ND and FMD groups described here are provided in Table 2. Two participants out of the twenty with FMD had MRI brain scans reported as possible atrophy. They were both followed up; One followed up at 24 months was aware that they had been working in a very stressful job at time of the first consultation. At follow-up they had changed jobs and now had no memory complaints. The second person was followed up at 18 months. The Montreal Cognitive Assessment (MoCA) was 26/30 at follow-up (prior ACE 85 and MoCA 22). They were functioning normally in a busy job. Two participants out of twenty cases with ND had normal structural scans but one of these had abnormal Single-Photon Emission Computed Tomography (SPECT). The other was seen for follow-up at 12 months and ACE-R had decreased from 87 to 82, the clinical picture at this stage being consistent with AD.

In Part 2 of the study, three out of five FMD cases did not attend for neuropsychology testing, hence the missing MMSE scores in table 2. However, their ACE-R scores were 87, 96 and 97. Two had entirely normal neuroimaging and were discharged. One had an old caudate head infarct and on follow-up one year later was still working, managing a team. Repeat ACE-R was 96 (97 one year earlier). Also in Part 2, one person with ND did not have detailed neuropsychological testing. Neuroimaging showed atrophy. On follow-up this patient showed significant cognitive impairment.

There were no significant differences in demographics, depression, anxiety or ACE-R scores between ND participants in Parts 1 and 2 or between FMD patients in the two parts of this study. 20 of the 30 participants included in Part 1, and seven of ten included in Part 2 of this study were accompanied. Feature 11 (patient's head turn encouraging accompanying person to answer a question directed at the patient) could not be rated in two of the accompanied encounters because participants had only consented to audio recording the interaction.

Table 3 shows the analyst's DSA ratings of the interactions included in Part 1 of this study. A more detailed description of the individual items can be found on the DSA form (additional web content). The median total score of the first nine items of the DSA was +5 in the ND (range +8 to -3) and -5 in the FMD group (range 0 to -9, difference $p<0.001$).

The median total of the five additional items to be rated in accompanied encounters was 2 (range +5 to -3) in the ND group and -1 (range 1 to -5) in the FMD group (difference $p=0.003$). The fact that only one of the additional items to be rated in accompanied encounters individually yielded a statistically significant between-group difference may (at least in part) be explained by the relatively small number of accompanied interactions available for analysis.

In view of the fact that the additional item scores for accompanied interactions were only available for a subset of the encounters, only the first nine items were used for the AUROC analysis and the estimation of a quantitative diagnostic threshold. The area under the ROC curve was 0.98 (see Figure 2). At the optimal DSA score for the distinction of patients with ND from those with FMD of +1 (with DSA score above this threshold suggesting a diagnosis of ND) the DSA-derived total score identified patients with ND with a sensitivity of 86.7% and a specificity of 100%.

3.4 RESULTS – PART 2

3.4.1 Quantitative scoring using the DSA (blinded results)

Rater 1 was accurate in 10/10 cases, whilst Rater 2 correctly predicted 9/10 diagnoses on the basis of DSA-guided qualitative analysis (Rater 1 was more experienced because of his involvement in Part 1 of the project). The results were identical when the linguistic diagnostic prediction was based on the numeric DSA scores. The case misdiagnosed by Rater 2 as FMD (when the ultimate medical diagnosis was ND) attracted the lowest score Rater 1 gave to any of the patients assessed as having ND (+4) and the highest score Rater 2 gave to any patients thought to have FMD (-1). This suggests that the patient misdiagnosed by Rater 2 had an objectively ambiguous conversational profile, posing a particular discriminatory challenge. The differences in the two raters' diagnostic prediction was based on a single completely discordant judgment of DSA item 4 (ratings 1 vs. -1) and on non-concordant decisions (0 vs. 1 or 0 vs. -1) on DSA items 3, 7, 9, 13 and 14 (see Table 4 for further DSA scoring details).

3.4.2 Inter-rater reliability of the DSA

In terms of the final diagnosis (either based on the two raters' qualitative judgements or the quantitative procedure using the diagnostic cut-off derived from the AUROC analysis), the raters agreed in 9/10 cases. The Kappa value for the DSA procedure as a whole was therefore 0.8 (SE of Kappa = 0.19, 95% confidence interval 0.44 to 1.0) suggesting 'very good' inter-rater reliability. We also looked at the inter-rater reliability of the 123 individual +1, 0 or -1 ratings from Part 2 of this study. Both ratings were fully concordant for 87 numeric scores (the scores from raters 1 and 2 were 1/1, 0/0 or -1/-1), non-concordant for 30 and discordant for 6. This means that both raters agreed on 70.7% of the observations when agreement on 33.8% of the ratings would have been expected by chance. The Kappa value for all 123 DSA-based ratings combined was 0.56 (SE of Kappa = 0.06, 95% confidence interval 0.44 to 0.68), consistent with 'moderate' inter-rater agreement. The two raters'

scores for each item assessed in part 2 of the study and the Kappa-values of each individual item are shown in table 3.

Discussion

Our previous qualitative work has demonstrated that it is possible to describe characteristic conversational profiles of patients describing cognitive problems due to ND or FMD based on their interactional and linguistic contributions to initial encounters in a memory clinic[20, 21]. However, in these descriptive studies, the conversation analysts who analysed video- and audio recordings of memory clinic encounters between neurologists, patients and (sometimes) accompanying persons were always aware of the patients' medical diagnoses during the analytic process. The present study is the first to demonstrate that these linguistic and interactional features can be used diagnostically to predict diagnoses of ND or FMD made on the basis of standard medical criteria. What is more, we show that qualitative assessments can be structured and likely medical diagnoses formulated using a Diagnostic Scoring Aid with a numeric diagnostic cut-off. The fact that the linguistic raters involved in this study had no expertise in the medical assessment of patients presenting with memory problems, together with the relatively high level of agreement between the two raters, suggests that the raters did not base their diagnostic predictions on an ill-defined hunch but on robust and objectifiable interactional observations.

The correct classification of 9/10 by one rater and 10/10 by a second independent rater, and the very good inter-rater reliability of the DSA-guided procedure as a whole, suggest that the addition of the structured observation of interactional features can make a significant contribution to screening processes for ND. Importantly, this is one of the few studies of cognitive screening 'tools' to include participants with FMD; most previous studies compared patients with memory impairment with healthy controls. The inclusion of a group of patients with memory complaints but no neurological disorder adds ecological validity to our findings. Compared to other studies set in clinical situations (such as a study exploring the screening potential of

the 6CIT brief cognitive test in primary care) our approach appears to have greater reliability and validity [9].

Especially when integrated in care pathways that include structural brain imaging and conventional cognitive testing, interactional and linguistic observations may increase the confidence of non-expert clinicians to diagnose clear cases of FMD, enabling them to treat patients in primary care or to refer them on to services providing appropriate treatment for functional neurological disorders. Importantly, the interactional and linguistic observations contributing to the diagnosis of FMD may allow clinicians to provide more effective reassurance by allowing them to demonstrate to patients that they are displaying good memory function in interaction. It should be possible for clinicians to pick up these features during routine clinic encounters. Previous studies in patients presenting with epileptic or dissociative (non-epileptic) seizures have demonstrated that doctors can learn to change their history-taking style to optimise patients' opportunities to demonstrate particular conversational behaviours and to make diagnostically useful interactional observations as they take a patient's history[41]. The DSA developed here can be the basis of similar studies in clinical settings in which patients present with memory problems.

This study has a number of limitations. First and foremost, we were only able to explore the potential of CA as a tool capable of predicting medical diagnoses in the setting of initial memory clinic encounters in a modest number of patients. Whilst the consecutive recruitment and the levels of statistical significance in between-group tests on a range of separate conversational features observed even in such a small patient group make it unlikely that our findings are spurious, it would be desirable to replicate our findings in a larger and more diverse group of patients. One particular limitation of our findings in this regard is that we excluded patients with depression and those with VCI from this first quantitative study of our method. These are important differential diagnoses, which will need to be picked up by screening procedures. Future larger studies will need to demonstrate that the inclusion of interactional and linguistic observations can contribute to

screening or stratification procedures in which patients with these problems are allocated to the correct management pathways. We also recognised that the ND and FMD groups in this study were not age-matched since those with ND were significantly older. This is not surprising as the biggest risk factor for ND is increasing age. However as younger patients with memory concerns are increasingly referred to specialist memory clinics it is important to include and compare all age groups in studies of this nature. Furthermore the mean MMSE score of the ND group was lower than that of the FMD group (20.4 versus 28.2), reflecting the relatively late stage in the development of cognitive disorders at which patients are currently first referred to specialist services. Only four of the patients in the ND groups had MCI. An optimal screening tool for the earliest stages of ND would need to be capable of picking up patients with MCI and near normal MMSE scores). This means that confirmatory studies capturing more patients at an earlier stage of ND will be required before the method described here can be embedded in screening procedures. It is also relevant that the mean MMSE in the ND sample was 20.44, which suggest that the MMSE scores of the MCI patients were already in the borderline range. We do not know how effective this tool will be at distinguishing MCI from FMD. Because a clinical diagnosis of MCI refers to a very heterogeneous symptom profile, the distinction between MCI and FMD might be difficult and will require larger number of participants, along with prospective follow-up to investigate whether it can predict those who at high risk of developing AD or other dementias.

We have only studied native English speakers; findings may have been different in patients speaking other languages or those using English as a second language. Although several different doctors were involved in the clinic conversations studied here, it would also be important to test this procedure in different clinical settings (for instance in community-based clinics, during home visits and in elderly-care settings). The fact that the ND group included patients with memory problems of different aetiologies should not be considered a weakness of this study. Although it is likely the method employed in this study could also be deployed to identify interactional

differences between different ND (such as Alzheimer's disease or frontotemporal dementia) the fact that we were able to distinguish clearly between patients with a range of ND and those with FMD demonstrates its potential for screening or stratifying patient management.

We did not have access to investigations confirming clinical diagnosis with tests documenting the presence of amyloid or tau (PET or cerebrospinal studies), but this reflects current NICE guidelines (http://www.nice.org.uk/guidance/CG42) and our 'medical' diagnoses were based on multidisciplinary assessment by experts including detailed neuropsychological testing and structural brain imaging as well as clinical follow up.

Future studies will need to demonstrate how much diagnostic value the observation of interactional features can add to conventional brief cognitive screening tools. A combined approach with an automated low cost, high-speed system to analyse speech will require the use of technology rather than Conversation Analysts. One way in which the research described here can be taken forward involves the computerised analysis of speech which has shown some promise in distinguishing AD, MCI and healthy controls[16]. Early indications are that computerised speech analysis and machine learning algorithms can also be used to produce an automated system to pick up and evaluate the sort of interactional observations described here and can discriminate between ND and FMD [22].

More immediately, the findings described here can be used in the training of clinicians working with patients with memory problems.

Acknowledgements: This article presents independent research funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0211-24079). 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. We would also like to thank all participating clinicians working in the memory clinic (Aijaz Khan, Alice Brockington, Dilraj Sokhi and Muhammad Rafiq).
We also acknowledge the support of the NIHR Sheffield Clinical Research Facility'.

References

1. Banerjee S. Living well with dementia--development of the national dementia strategy for England. Int J Geriatr Psychiatry 2010;25:917-922.
2. Audit RMC. http://www.rcpsych.ac.uk/workinpsychiatry/qualityimprovement/nationalclinicalaudits/memoryservicesaudit.aspx. 2015

3. Larner AJ. Impact of the National Dementia Strategy in a neurology-led memory clinic: 5-year data. Clin Med 2014;14:216.
4. Mukadam N, Livingston G, Rantell K, Rickman S. Diagnostic rates and treatment of dementia before and after launch of a national dementia policy: an observational study using English national databases. BMJ Open 2014;4:e004119.
5. Menon R, Larner AJ. Use of cognitive screening instruments in primary care: the impact of national dementia directives (NICE/SCIE, National Dementia Strategy). Fam Pract 2011;28:272-276.
6. Bell S, Harkness K, Dickson JM, Blackburn D. A diagnosis for £55: what is the cost of government initiatives in dementia case finding. Age Ageing 2015.
7. Robinson L, Gemski A, Abley C, et al. The transition to dementia--individual and family experiences of receiving a diagnosis: a review. Int Psychogeriatr 2011;23:1026-1043.
8. Samsi K, Abley C, Campbell S, et al. Negotiating a labyrinth: experiences of assessment and diagnostic journey in cognitive impairment and dementia. Int J Geriatr Psychiatry 2014;29:58-67.
9. Hessler J, Brönner M, Etgen T, et al. Suitability of the 6CIT as a screening test for dementia in primary care patients. Aging Ment Health 2014;18:515-520.
10. Forbes-McKay KE, Venneri A. Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. Neurol Sci 2005;26:243-254.
11. Garrard P, Lambon Ralph MA, Patterson K, Pratt KH, Hodges JR. Semantic feature knowledge and picture naming in dementia of Alzheimer's type: a new approach. Brain Lang 2005;93:79-94.
12. Bayles KA, Boone DR. The potential of language tasks for identifying senile dementia. J Speech Hear Disord 1982;47:210-217.
13. Bayles KA. Effects of working memory deficits on the communicative functioning of Alzheimer's dementia patients. J Commun Disord 2003;36:209-219.

14. Tomoeda CK, Bayles KA, Trosset MW, Azuma T, McGeagh A. Cross-sectional analysis of Alzheimer disease effects on oral discourse in a picture description task. Alzheimer Dis Assoc Disord 1996;10:204-215.
15. 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;45:959-963.
16. Konig A, Satt A, Sorin 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;1:112-124.
17. López-de-Ipiña K, Alonso JB, Travieso CM, et al. On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. Sensors (Basel) 2013;13:6730-6745.
18. Schwabe M, Howell SJ, Reuber M. Differential diagnosis of seizure disorders: a conversation analytic approach. Soc Sci Med 2007;65:712-724.
19. Reuber M, Monzoni C, Sharrack B, Plug L. Using interactional and linguistic analysis to distinguish between epileptic and psychogenic nonepileptic seizures: a prospective, blinded multirater study. Epilepsy Behav 2009;16:139-144.
20. Elsey C, Drew P, Jones D, et al. Towards diagnostic conversational profiles of patients presenting with dementia or functional memory disorders to memory clinics. Patient Educ Couns 2015.
21. Jones D, Drew P, Elsey C, et al. Conversational assessment in memory clinic encounters: interactional profiling for differentiating dementia from functional memory disorders. Aging Ment Health 2015:1-10.
22. Mirheidari B, Blackburn D, Harkness K, et al. Toward the Automation of Diagnostic Conversation Analysis in Patients with Memory Complaints. J Alzheimers Dis 2017;58:373-387.
23. Wakefield SJ, McGeown WJ, Shanks MF, Venneri A. Differentiating normal from pathological brain ageing using standard neuropsychological tests. Curr Alzheimer Res 2014;11:765-772.
24. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189-198.
25. D W. Wechsler Adult Intelligence Scale-III. San Antonio, TX, USA: The Psychological
Corporation, 1997.
26. A. R. L'examen clinique en psychologie. Paris: Presses Universitaires de France, 1964.
27. Raven J. Coloured Progressive Matrices Sets A, Ab, B. Manual Sections 1 & 2 Oxford Oxford
Psychologists Press, 1995.
28. JR S. Studies of interference in serial verbal reactions. Journal of Experimental Psychology 1935
18:643-662.
29. De Renzi E, Faglioni P. Normative data and screening power of a shortened version of the Token Test. Cortex 1978;14:41-49.
30. McKhann GM, Knopman DS, Chertkow H, 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;7:263-269.
31. Rascovsky K, Hodges JR, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 2011;134:2456-2477.
32. Petersen RC. Mild cognitive impairment: current research and clinical implications. Semin Neurol 2007;27:22-31.
33. Schmidtke K, Pohlmann S, Metternich B. The syndrome of functional memory disorder: definition, etiology, and natural course. Am J Geriatr Psychiatry 2008;16:981-988.
34. Kirby HB, Harper RG. Team assessment of geriatric mental patients: the care of functional dementia produced by hysterical behavior. Gerontologist 1987;27:573-576.
35. Kirby HB, Harper RG. Team assessment of geriatric mental patients. (II): Behavioral dynamics and psychometric testing in the diagnosis of functional dementia due to hysterical behavior. Gerontologist 1988;28:260-262.
36. Jefferson G. A sketch of some orderly aspects of overlap in natural conversation. Philadelphia: John Benjamins, 2004.
37. Frankel R. From sentence to sequence: Understanding the medical encounter through micro-interactional analysis. Discourse processes 1984;7:135-170.
38. Drew P, Chatwin J, Collins S. Conversation analysis: a method for research into interactions between patients and health-care professionals. Health Expect 2001;4:58-70.
39. Maynard DW, Heritage J. Conversation analysis, doctor-patient interaction and medical communication. Med Educ 2005;39:428-435.
40. Robinson JH, J. Intervening With Conversation Analysis:
The Case of Medicine RESEARCH ON LANGUAGE AND SOCIAL INTERACTION 2014;47:201-218.
41. Jenkins L, Cosgrove J, Ekberg K, Kheder A, Sokhi D, Reuber M. A brief conversation analytic communication intervention can change history-taking in the seizure clinic. Epilepsy Behav 2015;52:62-67.

Figure 1. CONSORT diagram showing recruitment to study. FMD- Functional Memory Disorder; DPD- Depressive Pseudo Dementia; ND- memory problems secondary to neurological disorders (ND) – this includes neurodegenerative dementias and mild cognitive impairment due to likely underlying neurodegenerative aetiology.

Figure 2: Area under the Receiver Operator Characteristic Curve
A ROC curve was constructed for the sample of 15 ND and 15 FMD cases.

Table 1 Diagnostic Scoring Aid (DSA)
For each interactional feature scores range between 1 and -1 (1: in favour of ND; 0: undecided or unable to rate; -1: in favour of FMD). There are 9/14 (unaccompanied/accompanied) items to score, so the maximum score is 9/14 (unaccompanied/accompanied) and minimum score -9/-14 (unaccompanied/accompanied). A high score corresponds to a stereotypical ND description; a low score to a stereotypical FMD description.

Supplementary Table 1 Comparison of patients eligible for participation in the study who did and did not consent to have their clinic interaction recorded and analysed.

There is no significant difference in demographics or test scores between the people who consented and the people who did not consent. Addenbrooke's Cognitive Examination; Patient Health Questionnaire-9 PHQ9- Generalised Anxiety Disorder scale 7 GAD7

Table 2 Demographic and neuropsychological results
ACE-R Addenbrooke's Cognitive Examination. MMSE Mini Mental State Examination. PHQ9- Patient Health Questionnaire 9 item depression scale. GAD7General Anxiety Disorder 7-item Scale CF- Confrontational Naming. VPA- Verbal Paired Associates, P&PT-Pyramid & Palm Trees, Rey's CF-Rey's Complex Figure, SF- Semantic Fluency, PF - Phonemic Fluency, DS - Digit Span, VCA- Visuoconstructive Apraxia, TT-Token task, PM - Prose Memory. * 3 missing scores. + 1 missing score. #
Three missing scores from ND group (2 due to different protocol and one participants from part 2). Three missing scores from FMD group due to not attending appointments. Twenty participants with ND; comprised eight with AD, four with amnestic MCI, two 2 with vascular dementia, two with fronto temporal dementia, three with mixed AD and vascular and one unspecified dementia (without detailed neuropsychology).

Table 3: Diagnostic Scoring Aid (DSA) results
Profiles of 30 patients attending a specialist clinic with memory secondary to neurological disorders (ND) – this includes neurodegenerative dementias and mild cognitive impairment due to likely underlying neurodegenerative aetiology or functional memory disorder (FMD). 3a: Items rated in all

encounters / 3b: Items only rated in encounters also involving an accompanying person (AP).
For a full description of the items see supplementary appendix. Some items were unratable because a particular question was not asked (eg. “who is most concerned?”).

Table 4 Diagnostic Scoring Aid (DSA) results of blinded analysis.
Two independent linguistic raters (L1 and L2) of interactions with five patients with a medical diagnosis of FMD (5a) and five patients with medical diagnosis of ND (5b).