Creavin, S. T., Haworth, J., Fish, M., Cullum, S. J., Bayer, A., Purdy, S., & Ben-Shlomo, Y. (Accepted/In press). Clinical judgement of General Practitioners for the diagnosis of dementia: a diagnostic test accuracy study. *BJGP Open*.

This research was funded in whole, or in part, by the Wellcome Trust [108804/Z/15/z]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Clinical judgement of General Practitioners for the diagnosis of dementia: a diagnostic test accuracy study

Samuel Thomas Creavin MRCP MRCGP PhD[1,], Judy Haworth MB BS MSc[1], Mark Fish BSc MBBS MRCP MD[2], Sarah Cullum MRCPsych PhD [3], Antony Bayer FRCP[4], Sarah Purdy MBBS MPH MD[1], and Yoav Ben-Shlomo MBBS MSc PhD MRCP[1]

[1]Population Health Sciences, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS

[2]Royal Devon and Exeter NHS Foundation Trust, Barrack Road , Exeter, Devon, EX2 5DW

[3]Dept Psychological Medicine, School of Medicine, The University of Auckland, 22-30 Park Avenue, Grafton, Auckland 1142, New Zealand

[4]School of Medicine, Cardiff University, Cardiff, CF14 4YS

Samuel Thomas Creavin GP and Clinical Lecturer
Judy Haworth Dementia Specialist Doctor
Mark Fish Consultant Neurologist
Sarah Cullum Associate Professor and Old Age Psychiatrist
Antony Bayer Emeritus Professor of Geriatric Medicine
Sarah Purdy Professor and Pro Vice-Chancellor for Student Experience
Yoav Ben-Shlomo Professor of Clinical Epidemiology

Abstract

Background: General Practitioners (GPs) often report using clinical judgement to diagnose dementia. Aim: Investigate the accuracy of GPs' clinical judgement for the diagnosis of dementia. Design and Setting: Diagnostic test accuracy study, recruiting from 21 practices around Bristol. Method: The clinical judgement of the treating GP (index test) was based on the information immediately available at their initial consultation with a person aged over 70 years who had cognitive symptoms. The reference standard was an assessment by a specialist clinician, based on a standardised clinical examination and made according to ICD-10 criteria for dementia. Results: 240 people were recruited, with a median age of 80 years (IQR 75 to 84 years), of whom 126 (53%) were men and 132 (55%) had dementia. The median duration of symptoms was 24 months (IQR 12 to 36 months) and the median ACE-III score was 75 (IQR 65 to 87). GP clinical judgement had sensitivity 56% (95% CI 47% to 65%) and specificity 89% (95% CI 81% to 94%). Positive likelihood ratio was higher in people aged 70-79 years (6.5, 95% CI 2.9 to 15) compared to people aged ≥ 80 years (3.6, 95% CI 1.7 to 7.6), and in women (10.4, 95% CI 3.4 to 31.7) compared to men (3.2, 95% CI 1.7 to 6.2), whereas the negative likelihood ratio was similar in all groups. Conclusion: A GP clinical judgement of dementia is specific, but confirmatory testing is needed to exclude dementia in symptomatic people who GPs judge as not having dementia.

Dementia | General Practice | Sensitivity and Specificity | Medical History Taking | Symptom Assessment

Correspondence: sam.creavin@bristol.ac.uk

2,497 words

How this fits in

Previous studies in this area have investigated the accuracy of GP clinical judgement as a screening test for dementia in unselected people attending a primary care clinic or as a retrospective test based on their knowledge of their patient. Some have derived the accuracy of judgement from the medical records, which may not reflect the judgement of the clinician. The role of the GP in supporting a more effective route to diagnosis for people with dementia is a research priority for patients, carers and clinicians. This study shows that, in a symptomatic older adult, clinical judgement may be useful for helping to confirm a diagnosis of dementia, whereas GP judgement should not by itself be used to exclude dementia.

Introduction

The James Lind Alliance has identified the role of general practice in supporting a more effective route to diagnosis of dementia as a priority for health research (1). People with symptoms of dementia have historically faced long delays to get an assessment and an explanation for their symptoms (2). Approaches to address waiting lists have included psychiatrists supporting primary care memory clinics (3), integrated one-stop clinics (4), and training GPs to make a diagnosis in uncomplicated cases (5,6) which is supported by NICE (7). Some GPs have in the past been hesitant about diagnosing dementia when there is no disease modifying treatment (8) and disclosure of a diagnosis can still be problematic, especially if the affected person is not seeking help (9). The situation has been complicated in the UK by controversial policies that have funded case-finding for dementia (10–12). Formally evaluating cognition takes time, and familiarity with tests. A GP could use a range of brief cognitive assessments (13) to evaluate a person with symptoms of dementia and national guidelines differ on which test to use (14,15). Instead, GPs report using non-standardised processes (16), such as clinical judgement (17) to diagnose dementia. The sensitivity of GP clinical judgement for diagnosing dementia has been reported between from 51% (18) to 100% (19), and the specificity ranges from 58% (20) to 100% (19).

Previous studies to investigate the accuracy of GP clinical judgement have typically suffered from one of two significant limitations (21). Firstly, a definition of clinical judgement which is of unclear relevance to practice, such as judgement in hindsight, or documentation of recorded diagnoses in the medical record which are systematically incomplete (22). Secondly, sampling unselected people attending general practice regardless of symptoms, which is more akin to screening. The aim of this study was to address these limitations of earlier studies and investigate the prospective diagnostic accuracy of GP clinical judgement for the diagnosis of dementia syndrome in symptomatic people over 70 years (23).

Methods.

Population.

We recruited participants from 21 participating GP surgeries in the Bristol, North Somerset, and South Gloucestershire (BNSSG) area, which is a diverse geographic area within 15 miles of the City of Bristol, covering a total population of around 900,000 people across 82 GP practices. Research clinics were in four participating GP surgeries, strategically located for accessibility. We calculated that a minimum sample size of 200 was needed, based on a specificity of 95% in prior studies, and a 75% prevalence of dementia in local memory clinic data (24).

Inclusion and exclusion criteria.

Participants were people with cognitive symptoms but no prior diagnosis of dementia, aged at least 70 years and had been referred by their GP to this research study. Cognitive symptoms were not specified but generally include disturbance in memory, language, executive function, behaviour, and visuospatial skills (25). Symptoms were required to be present for at least six months, and could be reported by the person themselves, a family member, a professional, or another person; there was no severity threshold. Cognitive problems did not need to be the focus of the consultation and (as routine practice) GPs could enquire about cognition if they perceived a problem. Symptom duration was determined from the clinical history. An accompanying informant was mandatory. All participants were offered free accessible transport and translation services. People were excluded if they had a known neurological disorder (i.e. Parkinsonism, Multiple Sclerosis, learning disability, Huntington's disease), registered blind, profound deafness (i.e. unable to use a telephone), psychiatric disorder requiring current secondary care input, or if cognitive symptoms were either rapidly progressive or co-incident with neurological disturbance. People with cognitive problems that were so advanced that they were unable to consent were excluded as they were judged by a lay

advisory group to find the research process overly burdensome. GPs were encouraged to make a clinical judgement and refer a consecutive series of all eligible patients with cognitive symptoms to the study, regardless of what their clinical judgement was or any test results. GPs gave study information including a leaflet, and obtained verbal consent to share contact details with the study on a referral form, including the persons age, sex, contact details, and the GPs clinical judgement. The study team contacted people referred by GPs to re-confirm eligibility, provide further written study details, and offer a research clinic appointment. The research team took written consent from all participants.

Index test of clinical judgement.

The referring GP recorded their clinical judgement using an electronic referral form during a consultation with their patient about cognitive symptoms. Clinical judgement was operationalised as normal, cognitive impairment not dementia (CIND), or dementia as options for response to the question "Is your gut feeling that this person has..". GPs were not specially trained, were not required to arrange any test, and could refer people simultaneously or subsequently to NHS services. The study team contacted the practice at least three times to obtain any missing referral data.

Reference standard.

At the research clinic, a single specialist physician conducted a standardised assessment lasting approximately 60 minutes comprising clinical history, the Addenbrooke's Cognitive Examination III (ACE-III) (26), Brief Assessment Schedule Depression Cards (BASDEC) (27) and the informant-completed Bristol Activities of Daily Living (BADL) Questionnaire (28). The specialist was not aware of other test results including GP judgement or investigations. The reference standard was based on the evaluation of the specialist physician for dementia according to ICD-10 criteria (29) for each individual patient; specific cut-offs on the aforementioned measures were not used and the expert used her integrated assessment to reach a diagnosis. Cognitive impairment not dementia (CIND), was diagnosed by the same expert and included Petersen MCI (30) and other causes of cognitive impairment that met neither criteria for ICD-10 dementia nor Petersen MCI, such as traumatic brain injury or affective disorder. Medical records were reviewed for all participants six months after the research clinic to identify any subsequent information that would contradict this judgement. A second specialist adjudicated cases where there was diagnostic uncertainty at the research clinic using the initial specialist assessment and the medical record review, but without access to GP judgement. Study data were electronically entered and managed using REDCap (Research Electronic Data Capture) hosted at the University of Bristol (31).

Statistical methods.

Separate logistic regression analyses were used with non-participation (referred by GP but not taking part) as the dependent variable and GP judgement, age (in years) and female sex as the independent variables to test the hypothesis of no association with these variables. Time from referral to appointment was described using median and interquartile range and logistic regression was used to test the hypothesis of no association between time to appointment (in days) and dementia (as the dependent variable). Measures of diagnostic test accuracy were calculated together with 95% confidence intervals, for GP judgement of dementia against reference standard of dementia. Sensitivity analyses were done to explore whether accuracy varied by age (<80 years | ≥ 80 years) since prediction models perform differently in these age groups (32), and sex. Cochran's Q test was used to test the hypothesis of no difference in likelihood ratios between groups (33). We report this diagnostic test accuracy study in line with STARDdem guidelines.

Results

Participants.

Figure 1 goes here caption: STARDdem flowchart for inclusion of participants in the study

Recruitment took place between March 2015 and May 2017. Figure 1 shows a flowchart for inclusion in the study. The theoretically "eligible" figure of 1,735 people was derived from the age specific incidence of dementia (34) and the demographics of the population in the participating practices (34,956 aged over 70 years (35)). The number approached is unknown. One person who consented withdrew before any data were collected because they were acutely ill. Of the 240 with available data, there were 20 borderline cases that were adjudicated by a second specialist. The 240 people were classified by the reference standard as either Normal (47), Dementia (132) of whom 1 had DSM-5 but not ICD-10 because they had subjective but not objective amnesia, or were CIND (61) of 59 whom met criteria for MCI (1 affective disorder, 1 brain injury). Compared to people who participated, there was little evidence of an association between non-participation and a GP clinical judgement of CIND (odds ratio 1.2; 95% CI 0.55 to 2.41) or dementia (odds ratio 1.9; 95% CI 0.90 to 3.93). Compared to people who participated, non-participants were older (odds ratio per year 1.08; 95% CI 1.04 to 1.12), or more often female (odds ratio 1.88; 95% CI 1.21 to 2.92). The median time between referral (clinical judgement) and the clinic appointment (reference standard) was 47 days (IQR 30 to 72 days), the longest interval was 177 days, due to difficulties attending earlier appointments. There was no association between time from referral to appointment and dementia (odds ratio per day 1.0; 95% CI 0.99 to 1.01). Table 1 shows the demographics of participants and shows a cross tabulation of GP opinion against the reference standard, allowing derivation of diagnostic accuracy of clinical judgement for both CIND and dementia.

Table 1 goes here

Two people could not complete the ACE-III because English was not their first language; they had both declined an interpreter. In both cases sufficient information was available from other parts of the assessment for a categorisation about cognition to be made (one had normal cognition, one had dementia). For the 238 people who had an ACE-III score, the median was 75 (interquartile range 65 to 87). Referring GPs judged that 34 people were normal, 86 had dementia, and 120 were CIND; the one person who withdrew from the study due to acute illness was judged by the referring GP to have CIND. People that GPs judged as having dementia had a total ACE-III score IQR of 60 to 74, with a 90th centile of 81/100 and highest score of 95/100. Similarly, people that GPs judged as having CIND had an ACE-III score IQR 71 to 89.

Diagnostic accuracy.

Table 2 goes here

Table 2 shows the diagnostic accuracy for GP judgement for dementia. The sensitivity and specificity of GP judgement were respectively 56% (95% CI 47% to 65%) and 89% (95% CI 81% to 94%). Clinical judgement was more useful for ruling in dementia, than ruling it out, with higher specificity and positive predictive value than sensitivity and negative predictive value. In people aged 80 or more years, clinical judgement had similar sensitivity (p=0.296) and specificity (p=0.798) to those aged under 80 years. There was weak evidence that clinical judgement in women had a higher specificity (p=0.074) and a higher sensitivity (p=0.064) than clinical judgement in men.

Discussion

Summary.

From 21 participating GP surgeries, 456 people were referred and 240 were evaluated. Of these, 132 (55%; 95% CI 48% to 61%) had dementia. Clinical judgement as a single test had a LRP of 5 (95% CI 3 to 9) and a LRN of 0.5 (95% CI 0.4 to 0.6) for the target condition dementia. People that GPs judged as having dementia had a total ACE-III score

IQR of 60 to 74, and those that they judged as having MCI had a total ACE-III IQR 71 to 89. This compares to published ACE-III thresholds of <82 for dementia (36) and <88 for MCI (36) and suggests that in this study GPs are not being overly restrictive in their judgement for dementia, or liberal in their judgement for CIND.

Strengths and limitations.

The patient selection in the current study closely reflects real world clinical practice in the United Kingdom, with efforts to avoid exclusion based on language, transport, or appointment availability. Participants were included with a range of GP opinions about the presence of cognitive impairment in people who had presented with symptoms in a consultation; typically 2.5 problems are discussed per appointment (37). The index test reflects an average measure of diagnostic accuracy for an estimated 142 whole time equivalent GPs working in different settings (38), who were not specially trained. We instructed GPs not to use any formal test to inform their judgement, but it is possible that brief cognitive tests such as GP Cog (39) may have been occasionally used. Based on previous studies, clinical judgement is likely to be based on rules of thumb (16), not formal tests (17), and information on referral forms indicated that judgement was informed by "face to face presentation". The interval between clinical judgement and the reference standard was unlikely to be associated with a significant progression in cognitive impairment (15). We fully verified the index test for all consenting participants, obtained follow-up data after six months, and adjudicated uncertain cases.

There was no evidence of selective participation by cognitive status, but non-participants may differ in other unmeasured ways which affect diagnostic accuracy. As reported in the Results we estimate that up to 1,735 people in the study population would have developed symptoms in the study period, but it is unknown how many of these would have presented to their GP. We have no data on recruitment bias, but dementia was less prevalent than we predicted based on local memory clinic data, suggesting a lower threshold for referral to the study. Any systematic selection bias in who GPs referred to the study (such as excluding more frail people) would limit the generalisability of our findings to that group. An important limitation is that despite providing translation services the population were largely white, native English speakers. In addition, the confidence intervals for our sub-groups are still wide. We excluded people with advanced cognitive impairment that could not consent, so our findings cannot be generalised to that group; though it is likely that GPs would be more sensitive in identifying cognitive impairment at a more advanced stage.

Comparison with existing literature.

Table 3 goes here

Table 3 summarises the features of this study compared to the existing literature (40,41). A major strength of this study for applicability to practice is that it is one of only two studies to evaluate symptomatic people. Our study has the smallest number undergoing the index test, but only one other study has complete verification by the reference standard (42). Our study has lower sensitivity and higher specificity than the French study (20), but this could be because the French study verified only 26% of people who underwent the index test (where participating GPs referred five patients per GP over two years), or because other studies did not require participants to be symptomatic and consequently had a lower prevalence of dementia (ranging 2% to 29%) (43–46).

Implications for Research and/or practice.

The accuracy of clinical judgement was comparable to other brief cognitive tests, many of which are now subject to licensing restrictions. The test characteristics of clinical judgement would support an approach to subsequent testing where highly sensitive tests are performed in people who GPs judge as not having dementia but there is significant patient concern (to rule out disease), and very highly specific, but minimally burdensome tests are done in people

who GPs do think have dementia. This would be a change to current practice where cognitive testing is typically done with the same tests regardless of GP judgement.

Additional information

Funding. The Wellcome Trust (Fellowship 108804/Z/15/z £321,248), Avon Primary Care Research Collaboration (£19,705), The Claire Wand fund (£5040), and the National Institute for Health Research School for Primary Care research (£9,971). This research was funded in whole, or in part, by the Wellcome Trust [108804/Z/15/z]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. For access to data please contact the corresponding author. The Western Clinical Research Networks approved an application for service support costs for practices to provide for the expense of room hire in GP surgeries and GPs referring people to the study. YBS is supported by the NIHR Applied Research Collaboration West (NIHR ARC West). The views expressed in this article are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care

Ethical approval. The National Research Ethics Service Committee London – Bromley (reference 14/LO/2025) gave a favourable ethical opinion on 25 November 2014. NHS Research and Development approvals were granted by Avon Primary Care Research Collaboration on behalf of Bristol, North Somerset and South Gloucestershire clinical commissioning groups. The University of Bristol acted as Sponsor.

COMPETING FINANCIAL INTERESTS

STC None JH None MF None SC None AB None SP None YBS None

ACKNOWLEDGEMENTS

The authors thank the participants and the staff at participating practices, without whom this work would not have been possible. The staff at the West of England Clinical Research Network arranged for redaction, collection and transport of medical records from general practices.

Bibliography

1. James Lind Alliance. Dementia Top 10 Priorities. Research Priorities: top 10s. 2014 [cited 2014 Jul 23]. Available from: http://www.lindalliance.org/top-tens.asp
2. Manthorpe J, Samsi K, Campbell S, et al. From forgetfulness to dementia: clinical and commissioning implications of diagnostic experiences. Br J Gen Pract. 2013 Jan;63(606):e69-75.
3. Greaves I, Greaves N, Walker E, et al. Gnosall Primary Care Memory Clinic: Eldercare facilitator role description and development. Dementia. 2013 Aug 5;
4. Wells CE, Smith SJ. Diagnostic Care Pathways in Dementia: A Review of the Involvement of Primary Care in Practice and Innovation. J Prim Care Community Health. 2016 Nov 22;
5. Dodd E, Cheston R, Fear T, et al. An evaluation of primary care led dementia diagnostic services in Bristol. BMC Health Serv Res. 2014 Nov 29;14(1):592.
6. Dodd E, Cheston R, Fear T, et al. Primary Care Led Dementia services in Bristol (PiCLed): A participatory mixed method study. Bristol; 2014.
7. National Institute for Health and Care Excellence. 5.1 Improving early identification, assessment and diagnosis. In: Support for commissioning dementia care CMG48. London; 2013. p. 64.
8. Iliffe S, Wilcock J, Drennan V, et al. Changing practice in dementia care in the community: developing and testing evidence-based interventions, from timely diagnosis to end of life (EVIDEM). Program Grants Appl Res. 2015;3(3):1–596.
9. Low L-F, McGrath M, Swaffer K, Brodaty H. Communicating a diagnosis of dementia: A systematic mixed studies review of attitudes and practices of health practitioners. Dementia. 18(7–8):2856–905.
10. Burns A. The benefits of early diagnosis of dementia. BMJ. 2012 Jan 22;344(may22_1):e3556.
11. Burns A. Alistair Burns and 51 colleagues reply to David Le Couteur and colleagues. BMJ. 2013 Jan

15;347(oct15_6):f6125.
12. Le Couteur DG, Doust J, Creasey H, Brayne C. Political drive to screen for pre-dementia: not evidence based and ignores the harms of diagnosis. BMJ. 2013 Jan 9;347(sep09_21):f5125.
13. Brown J. The use and misuse of short cognitive tests in the diagnosis of dementia. J Neurol Neurosurg Psychiatry. 2015;86(6):680–5.
14. Davis DHJ, Creavin ST, Yip JLY, et al. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. Cochrane database Syst Rev. 2015 Oct 29;10(10):CD010775.
15. Creavin ST, Wisniewski S, Noel-Storr AH, et al. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. Cochrane Database Syst Rev. 2016 Jan 13;1(4):CD011145.
16. Pentzek M, Fuchs A, Wiese B, et al. General practitioners' judgment of their elderly patients' cognitive status. J Gen Intern Med. 2009 Dec;24(12):1314–7.
17. O'Connor DW, Pollitt PA, Hyde JB, et al. Do general practitioners miss dementia in elderly patients? BMJ. 1988 Oct 29;297(6656):1107–10.
18. Kaduszkiewicz H, Zimmermann T, Van den Bussche H, et al. Do general practitioners recognize mild cognitive impairment in their patients? J Nutr Health Aging. 2010 Oct;14(8):697–702.
19. De Lepeleire J, Aertgeerts B, Umbach I, et al. The diagnostic value of IADL evaluation in the detection of dementia in general practice. Aging Ment Heal. 2004 Jan;8(1):52–7.
20. Rondeau V, Allain H, Bakchine S, et al. General practice-based intervention for suspecting and detecting dementia in France: A cluster randomized controlled trial. Dementia. 2008 Nov 1;7(4):433–50.
21. Creavin ST, Noel-Storr AH, Richard E, et al. Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people. Cochrane Database Syst Rev. 2017;2017(2).
22. Russell P, Banerjee S, Watt J, et al. Improving the identification of people with dementia in primary care: evaluation of the impact of primary care dementia coding guidance on identified prevalence. BMJ Open. 2013 Jan 23;3(12):e004023.
23. Creavin ST, Cullum SJ, Haworth J, et al. Towards improving diagnosis of memory loss in general practice: TIMeLi diagnostic test accuracy study protocol. BMC Fam Pract. 2016;17(1):79.
24. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. J Clin Epidemiol. 2005 Aug;58(8):859–62.
25. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: Author; 2013.
26. Hsieh S, Schubert S, Hoon C, et al. Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease. Dement Geriatr Cogn Disord. 2013 Jan;36(3–4):242–50.
27. Adshead F, Cody DD, Pitt B. BASDEC: a novel screening instrument for depression in elderly medical inpatients. BMJ. 1992 Aug 15;305(6850):397.
28. Bucks RS, Ashworth DL, Wilcock GK, Siegfried K. Assessment of activities of daily living in dementia: development of the Bristol Activities of Daily Living Scale. Age Ageing. 1996 Mar;25(2):113–20.
29. World Health Organization. The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research. The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research. Geneva: World Health Organisation; 1993.
30. Petersen RC. Mild cognitive impairment as a diagnostic entity. In: Journal of Internal Medicine. 2004. p. 183–94.
31. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019 Jul;95:103208.
32. Walters K, Hardoon S, Petersen I, et al. Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data. BMC Med. 2016 Jan 21;14(1):6.
33. Cohen JF, Chalumeau M, Cohen R, et al. Cochran's Q test was useful to assess heterogeneity in likelihood ratios in studies of diagnostic accuracy. J Clin Epidemiol. 2015 Mar;68(3):299–306.
34. Prince M, Wimo A, Guerchet M, et al. World Alzheimer Report 2015: The Global Impact of Dementia - An

analysis of prevalence, incidence, cost And trends. London; 2015.
35. NHS England. Numbers of Patients Registered at a GP Practice - April 2018. 2018.
36. Beishon LC, Batterham AP, Quinn TJ, et al. Addenbrooke's Cognitive Examination III (ACE-III) and mini-ACE for the detection of dementia and mild cognitive impairment. Cochrane database Syst Rev. 2019;
37. Salisbury C, Procter S, Stewart K, et al. The content of general practice consultations: cross-sectional study based on video recordings. Br J Gen Pract. 2013 Nov;63(616):e751–9.
38. NHS Digital. Selected CCG information: NHS Bristol, North Somerset and South Gloucestershire CCG. In: General Practice Workforce. 2019.
39. Brodaty H, Pond D, Kemp NM, et al. The GPCOG: a new screening test for dementia designed for general practice. J Am Geriatr Soc. 2002 Mar;50(3):530–4.
40. van den Dungen P, van Marwijk HWM, van der Horst HE, et al. The accuracy of family physicians' dementia diagnoses at different stages of dementia: a systematic review. Int J Geriatr Psychiatry. 2012 Apr;27(4):342–54.
41. Mitchell AJ, Meader N, Pentzek M. Clinical recognition of dementia and cognitive impairment in primary care: a meta-analysis of physician accuracy. Acta Psychiatr Scand. 2011 Sep;124(3):165–83.
42. Valcour VG, Masaki KH, Curb JD, Blanchette PL. The detection of dementia in the primary care setting. Arch Intern Med. 2000 Oct 23;160(19):2964–8.
43. Cooper B, Bickel H, Schäufele M. The ability of general practitioners to detect dementia and cognitive impairment in their elderly patients: A study in Mannheim. Int J Geriatr Psychiatry. 1992 Aug;7(8):591–8.
44. Pond CD, Mant A, Kehoe L, et al. General practitioner diagnosis of depression and dementia in the elderly: can academic detailing make a difference? Fam Pract. 1994 Jun;11(2):141–7.
45. De Lepeleire J, Aertgeerts B, Umbach I, et al. The diagnostic value of IADL evaluation in the detection of dementia in general practice. Aging Ment Health. 2004 Jan;8(1):52–7.
46. Kaduszkiewicz H, Zimmermann T, Van den Bussche H, et al. Do general practitioners recognize mild cognitive impairment in their patients? J Nutr Health Aging. 2010 Oct;14(8):697–702.

Table 1. Characteristics of participants by cognitive category

Characteristic	Cognitive category * Dementia n=132	CIND n=61	Normal n=47
Sex n (column %)			
Male	68 (52)	35 (57)	23 (49)
Female	64 (49)	26 (43)	24 (51)
Age (years) Median (IQR)			
At clinic	82 (77-86)	80 (75-83)	75 (72-80)
Left school	15 (15-16)	15 (15-16)	16 (15-16)
Retired	60 (58-65)	60 (58-67)	61 (58-65)
Symptom onset Median (IQR) (months)			
Time ago	24 (12-36)	18 (12-24)	21 (12-36)
Type, n (column %)			
Gradual	111 (84)	55 (90)	43 (91)
Sudden	13 (10)	5 (8)	0 (-)
Uncertain	8 (6)	1 (1)	4 (9)
Symptom pattern n (column %)			
Course			
Progressive	111 (84)	42 (69)	29 (62)
Stepwise	2 (2)	0 (-)	0 (-)
Regressive	1 (1)	1 (2)	1 (2)
Static	5 (4)	7 (11)	9 (19)
Uncertain	13 (10)	11 (18)	8 (17)
Fluctuation			
None	112 (85)	53 (87)	45 (96)
Within one day	12 (9)	5 (8)	1 (2)
Over several days	8 (6)	3 (5)	1 (2)
ACE-III Score median (IQR)			
Total (max 100)	69 (61-74)	82 (76-87)	93 (90-95)
GP opinion n (row %)			
Normal	6 (18)	9 (26)	19 (56)
CIND	52 (43)	41 (34)	27 (23)
Dementia	74 (86)	11 (13)	1 (1)

Dementia according to ICD-10 definition; MCI according to Petersen definition.

ACE-III Addenbrookes' Cognitive Examination III; CIND Cognitive impairment, not dementia.

Table 2. Accuracy of GP judgement for the diagnosis of dementia

	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LRP (95% CI)	LRN (95% CI)
GP judgement (n=240)	56 (47 to 65)	89 (81 to 94)	86 (77 to 93)	62 (54 to 70)	5.1 (2.9 to 8.8)	0.49 (0.40 to 0.61)
Age ≥ 80 years (n=123)	57 (45 to 67)	84 (67 to 94)	89 (77 to 96)	46 (34 to 59)	3.6 (1.7 to 7.6)	0.52 (0.39 to 0.68)
Age < 80 years (n=117)	55 (40 to 70)	91 (82 to 97)	81 (64 to 93)	75 (65 to 84)	6.5 (2.9 to 15)	0.49 (0.35 to 0.68)
Men (n= 126)	50 (38 to 62)	85 (73 to 93)	79 (64 to 90)	59 (48 to 70)	3.2 (1.7 to 6.2)	0.59 (0.46 to 0.77)
Women (n= 114)	63 (50 to 74)	94 (84 to 99)	93 (81 to 99)	66 (54 to 77)	10.4 (3.4 to 31.7)	0.40 (0.29 to 0.55)

LRN negative likelihood ratio LRP positive likelihood ratio NPV

Table 3. Summary of seven studies investigating GP judgement for the diagnosis of dementia

	Mannheim (43)	Sydney (44)	Hawaii (42)	Antwerp (45)	AgeCoDe (46)	France (20)	This study
Participant selection*							
Series	C	C	C	C	R	C	C
Symptomatic	No	No**	No	No	No	Yes**	Yes
Characteristics of participants							
Number (index test)	3721	433	303	1003	3242	1453	240
Mean age (years)	76	85	75	75	80	81	80
% Female	70	84	63	63	66	71	47
% with dementia	29	25	9	2	2	50	55
Target condition and verification with reference standard							
Verified N	407	105	303	101	22	385	240
Verified %	11	24	100	1	70	26	100
GP Judgement (%)							
Not impaired	36	76	33	-	94	48	14
Cognitive impairment	41	-	-	-	-	-	40
Dementia	23	19	33	-	6	26	36
Uncertain	-	5	33	-	-	22	-
Diagnostic accuracy of clinical judgement for dementia							
Sensitivity	91	42	-	100	51	73	56
Specificity	76	89	-	100	96	58	89

C Consecutive R Random Symptomatic: symptoms required for participation

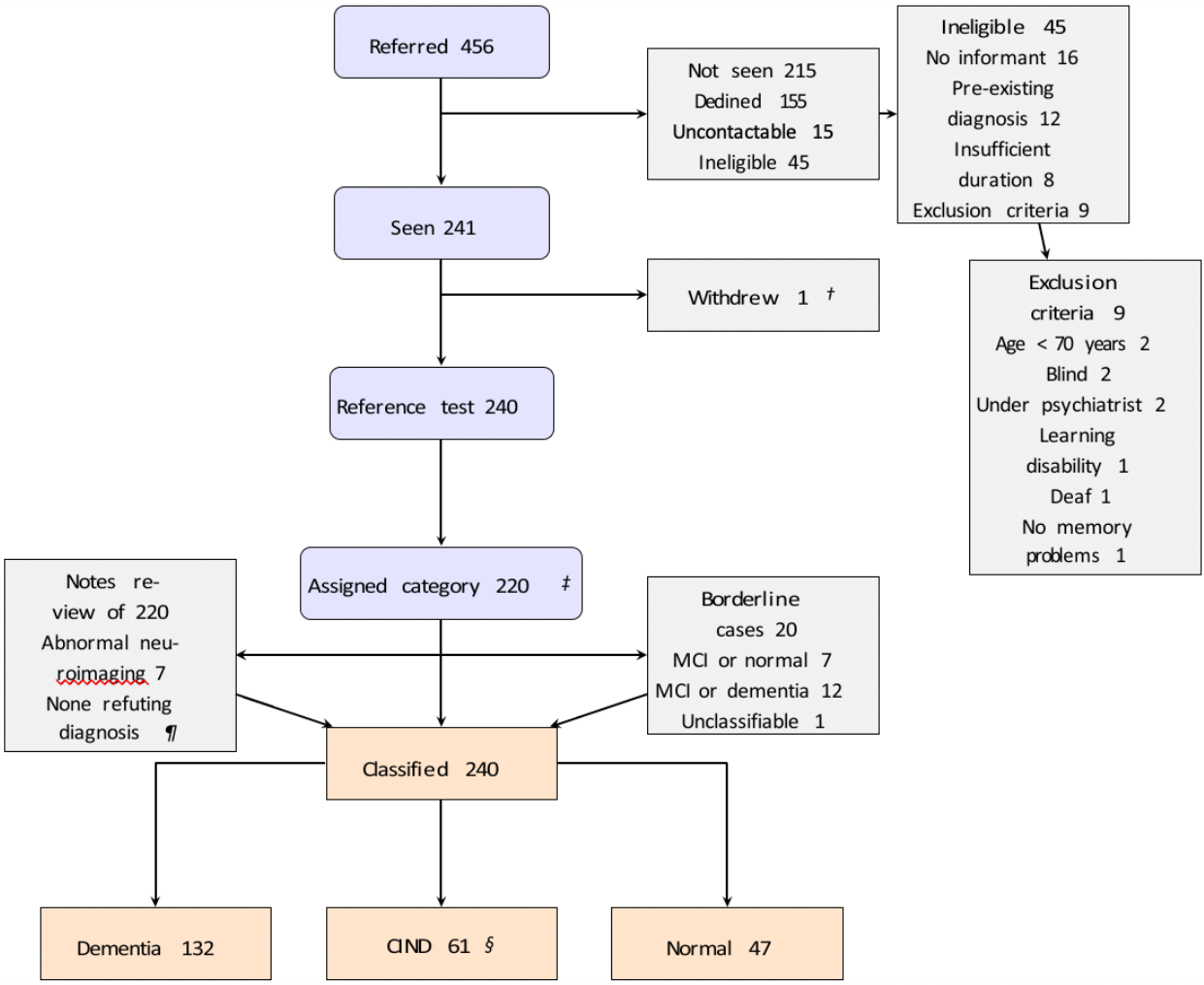
% verified = number underwent reference test / number underwent index test % with dementia = number with dementia / number verified

** Participants were not presenting with symptoms but GPs were asked to maximise the inclusion of people with suspected dementia – not reported

Supplementary Note 1: STARDdem checklist

Section, topic, and item number	STARD checklist item	Page
1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading " sensitivity and specificity")	1
2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups	2
3	The study population: the inclusion and exclusion criteria, setting and locations where data were collected. See also item 4 on recruitment and item 5 on sampling	2
4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard? See also item 5 on sampling and item 16 on participant loss at each stage of the study	2
5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected. See also item 4 on recruitment and item 16 on participant loss	2
6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	2
7	The reference standard and its rationale	3
8	Technical specifications of materials and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard. See also item 10 concerning the person(s) executing the tests	3
9	Definition of and rationale for the units, cutoffs, and/or categories of the results of the index tests and the reference standard	3
10	The number, training, and expertise of the persons executing and reading the index tests and the reference standard. See also item 8	3
11	Whether or not the readers of the index tests and reference standard were blinded (masked) to the results of the other test and describe any other clinical information available to the readers. See also item 7	3
12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g., 95% confidence intervals)	3
13	Methods for calculating test reproducibility, if done	NA

14	When study was performed, including beginning and end dates of recruitment	4
15	Clinical and demographic characteristics of the study population (at least information on age, sex, spectrum of presenting symptoms). See also item 18	9
16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended). See also items 3–5	14
17	Time interval between the index tests and the reference standard, and any treatment administered in between	4
18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition	9
19	A cross-tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard	9
20	Any adverse events from performing the index tests or the reference standard	4
21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g., 95% confidence intervals). See also item 12	8
22	How indeterminate results, missing data, and outliers of the index tests were handled	3
23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers, or centers, if done	10
24	Estimates of test reproducibility, if done. See also item 13	NA
25	Discuss the clinical applicability of the study findings	5-6



† one person had to withdraw part way through the reference test as they were acutely ill
‡ Dementia according to ICD-10 [38]

§ of 61 with CIND 59 met criteria for Peterson MCI (30), 1 affective disorder, 1 brain injury) ¶ One person met criteria for ICD-10 dementia and also had features of normal pressure hydrocephalus. Expert review endorsed a reference standard diagnosis of dementia.