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

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

Background: **The accuracy of General Practitioners' (GPs') clinical judgement for dementia is uncertain.** ***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 $\geq$ 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 for 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***

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; or 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, prospective 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). A GP could use a range of brief cognitive assessments (9) to evaluate a person with symptoms of dementia. National guidelines differ on which test to use, possibly because there is little evidence in a symptomatic primary care population (10, 11). Formally evaluating cognition takes time, and familiarity with the test. GPs report using non-standardised processes (12), such as their clinical judgement (13) to decide whether a person has dementia. The sensitivity of GP clinical judgement for diagnosing dementia has been reported between from 51% (14) to 100%

(15), and the specificity ranges from 58% (16) to 100% (15). Previous studies to investigate the accuracy of GP clinical judgement have typically suffered from one of two significant limitations (17). Firstly, a definition of clinical judgement which is of unclear relevance to practice, such as retrospective judgement, or indeed documentation of recorded diagnoses in the medical record, which are systematically incomplete (18). Secondly, sampling unselected people attending general practice regardless of symptoms, which is more akin to screening. To address these limitations, we investigated the prospective accuracy of GP clinical judgement for the diagnosis of dementia syndrome in people over 70 years (19).

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 for a lower bound of the specificity 95% confidence interval of 80%, based on a specificity of 95% in prior studies, and a 75% prevalence of dementia in local memory clinic data (20).

Inclusion and exclusion criteria. Participants were people with cognitive symptoms but no prior diagnosis of dementia, who were 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 (21). 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 main focus of the consultation. As is routine practice, GPs could initiate an enquiry about cognition if they perceived there was 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"*. GPs were not required to arrange any test and could also 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 with more than 20 years' experience in the field of dementia conducted a standardised assessment lasting approximately 60 minutes comprising clinical history, the Addenbrooke's Cognitive Examination III (ACE-III) (22), Brief Assessment Schedule Depression Cards (BASDEC) (23) and the Bristol Activities of Daily Living (BADL) Questionnaire (24). The specialist was not aware of other test results such as GP judgement or any investigations. The reference standard was based on the evaluation of the specialist physician for dementia according to ICD-10 criteria (25). Medical records were reviewed for all participants six months after the research clinic to identify any information that had come to light 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 the GP judgement. Study data were electronically entered and managed using REDCap (Research Electronic Data Capture) hosted at the University of Bristol (26).

Statistical methods. Characteristics of participants including age, sex and ACE-III score were tabulated by dementia status according to the reference standard. Separate logistic regression analyses were used with non-participation 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 (sensitivity, specificity, positive and negative likelihood ratios (LRP, LRN), predictive values) were calculated together with 95% confidence intervals, for GP judgement of dementia against reference standard of dementia. Decision curve analysis (27) was used

to show the net benefit (incorporating discrimination and calibration) of GP judgement at varying threshold probabilities (P_t). Decision curve analysis quantifies the net benefit of a test, in units of true positive, across a range of preferences (28) where a net benefit of 0.05 means "five true positives for every 100 patients in the target population" (29), compared to the "label:all" (worried about missing dementia) or "label:none" (worried about overdiagnosis). The net benefit equation incorporates true (TP) and false positives (FP), but not test negatives (30):

$$\text{net benefit} = \frac{\text{TP}}{\text{N}} - \frac{\text{FP}}{\text{N}} \times \frac{\text{P}_t}{\text{1-P}_t}$$

Sensitivity analyses were done to explore whether accuracy varied by age (<80 years | $\geq$ 80 years) since prediction models perform differently in these age groups (31), and sex. Cochran's Q test was used to test the hypothesis of no difference in likelihood ratios between groups (32).

Results

Participants. *Figure 1 goes here*
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 (33) and the demographics of the population in the participating practices (34,956 aged over 70 years (34). The number approached is unknown. One person who consented withdrew before any data collection was done because they were admitted acutely ill to hospital. 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.
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 a 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 and specificity to those aged under 80 years (LRP p=0.296 LRN p=0.798). There was weak evidence that clinical judgement in women had a higher LRP (p=0.074) and a lower LRN (p=0.064) than clinical judgement in men.
Figure 2 goes here
Figure 2 shows that clinical judgement has greater net benefit than a label:all approach at threshold probabilities of above 50%, and a label:none approach at threshold probabilities below 85%. At a threshold probability of 80%, indicating a preference for avoiding over-diagnosis, clinical judgement has a net benefit of 0.11 over the label:none approach, indicating an additional 11 true positives for every 100 people. If the doctor prefers to not miss dementia, with a threshold probability of 33%, then clinical judgement identifies five fewer true positives for every 100 people than the label:all approach (everyone is treated as if they have dementia: further tests or referral arranged).

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 (35) and <88 for MCI (35) 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 people being excluded 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, which means that cognitive problems were one of the problems discussed in the initial GP consultation; typically 2.5 problems are discussed per appointment (36). The index test clinical judgement in this study reflects an average measure of diagnostic accuracy for an estimated 142 whole time equivalent GPs working in different settings (37). Responses indicated that clinical judgement was typically informed by "face to face presentation". GPs were told they need not use any formal test to inform their judgement and based on previous studies this is likely to be based on rules of thumb (12) and not formal tests (13). The interval between clinical judgement and the reference standard was relatively short, and unlikely to be associated with a significant progression in cognitive impairment (11). Clinical judgement was fully verified against the reference standard for all consenting people who were referred and there was no evidence of selective participation by cognitive status. Follow-up data after six months was obtained, and uncertain cases were adjudicated. 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 (38, 39). 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 is one of only two studies with complete verification by the reference standard. This study has lower sensitivity and higher specificity than the French study (16), but this could be because the other study verified only 26% of people who underwent the index test, or because other studies did not require participants to be symptomatic and consequently had a lower prevalence of dementia (ranging 2% to 29%).

Implications for Research and/or practice. Diagnosis can be conceptualised as a pragmatic method of classification that is fit for purpose in the clinical setting (40), and GP judgement may often use heuristics (rules of thumb) and system one (non-analytical (41)) cognitive processes. Clinical judgement may be systematically different to formal definitions, just as different formal definitions (generally formulated for research needs) select different groups of people (42). The GP heuristic of dementia in the older adult may be an individual with forgetfulness who also has sensory impairment, limited mobility, multi-morbidity, and needs additional assistance performing activities of daily life (43). It remains to be seen which definition is most useful in practice. Concerns about resources and lack of specialist expertise for GPs to diagnose and manage dementia well have been reported for many years (44), and GPs have been reported to frame dementia care as a specialist activity (45). However, the priority of patients and their kin who are seeking medical input is to get a prompt diagnosis, in an emotionally safe and personalised way (2). Approaches to diagnosis (3, 6, 46) and follow-up (47) have been reported in primary care but regrettably, a well-designed intervention to improve practice was not effective in improving documentation or increasing case identification (8). In this study clinical judgement was more likely to under than over identify dementia. However, instead of training GPs, patients may benefit more from additional practice-based multi-disciplinary dementia case workers (8) to help refine GP classification, especially where GPs clinical judgement is "not dementia"; in England these roles could be provided through Primary Care Networks (48).

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. Written in LaTeX using `zHenriquesLab-StyleBioRxiv.cls` available at https://www.overleaf.com/latex/templates/henriqueslab-biorxiv-template/nyprsybwffws

Bibliography

1. James Lind Alliance. Dementia Top 10 Priorities, 2014.
2. Jill Manthorpe, Kritika Samsi, Sarah Campbell, et al. From forgetfulness to dementia: clinical and commissioning implications of diagnostic experiences. *The British journal of general practice : the journal of the Royal College of General Practitioners*, 63(606):e69–75, jan 2013. ISSN 1478-5242. doi: 10.3399/bjgp13X660805.
3. Ian Greaves, Nicola Greaves, Elaine Walker, et al. Gnosall Primary Care Memory Clinic: Eldercare facilitator role description and development. *Dementia (London, England)*, aug 2013. ISSN 1741-2684. doi: 10.1177/1471301213497737.

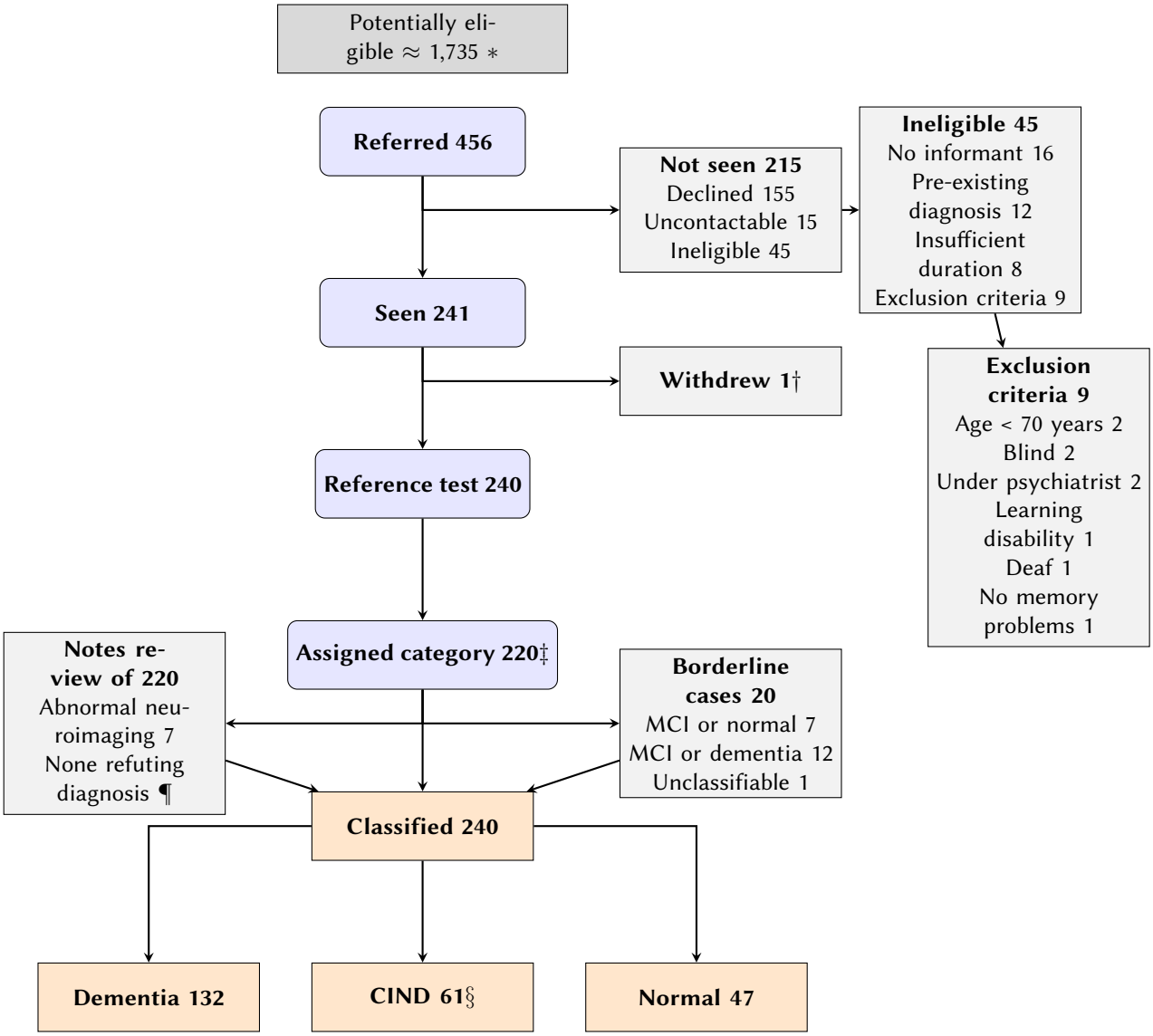
4. Christine E. Wells and Sarah J. Smith. Diagnostic Care Pathways in Dementia. *Journal of Primary Care & Community Health*, 8(2):103–111, apr 2017. ISSN 2150-1319. doi: 10.1177/2150131916678715.
5. Emily Dodd, Richard Cheston, Tina Fear, et al. An evaluation of primary care led dementia diagnostic services in Bristol. *BMC health services research*, 14(1):592, nov 2014. ISSN 1472-6963. doi: 10.1186/s12913-014-0592-3.
6. Emily Dodd, Rik Cheston, Tina Fear, et al. Primary Care Led Dementia services in Bristol (PiCLed): A participatory mixed method study. Technical report, 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*, chapter 5.1, page 64. London, 2013.
8. Steve Iliffe, Jane Wilcock, Vari Drennan, et al. Changing practice in dementia care in the community: developing and testing evidence-based interventions, from timely diagnosis to end of life (EVIDEM). *Programme Grants for Applied Research*, 3(3):1–596, 2015. ISSN 2050-4322. doi: 10.3310/pgfar03030.
9. Jeremy Brown. The use and misuse of short cognitive tests in the diagnosis of dementia. *J Neurol Neurosurg Psychiatry*, 86(6):680–685, 2015. ISSN 1468-330X. doi: 10.1136/jnnp-2014-309086.
10. Daniel Hj D.H. Davis, S.T. Sam T Creavin, J.L. Jennifer Ly Yip, et al. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. *The Cochrane database of systematic reviews*, 10:CD010775, jan 2015. ISSN 1469-493X. doi: 10.1002/14651858.CD010775.pub2.
11. Sam T Creavin, Susanna Wisniewski, Anna H Noel-Storr, 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 of Systematic Reviews*, 1(4): CD011145, jan 2016. ISSN 14651858. doi: 10.1002/14651858.CD011145.pub2.
12. Michael Pentzek, Angela Fuchs, Birgitt Wiese, et al. General practitioners' judgment of their elderly patients' cognitive status. *Journal of general internal medicine*, 24(12):1314–7, dec 2009. ISSN 1525-1497. doi: 10.1007/s11606-009-1118-2.
13. D W O'Connor, P A Pollitt, J B Hyde, et al. Do general practitioners miss dementia in elderly patients? *BMJ (Clinical research ed.)*, 297(6656):1107–10, oct 1988. ISSN 0959-8138.
14. H Kaduszkiewicz, T Zimmermann, H Van den Bussche, et al. Do general practitioners recognize mild cognitive impairment in their patients? *The journal of nutrition, health & aging*, 14(8):697–702, oct 2010. ISSN 1760-4788.
15. J. De Lepeleire, Bert Aertgeerts, I. Umbach, et al. The diagnostic value of IADL evaluation in the detection of dementia in general practice. *Aging and Mental Health*, 8(1):52–57, jan 2004. ISSN 13607863. doi: 10.1080/1360786310001613338.
16. Virginie Rondeau, Herve Allain, Serge Bakchine, et al. General practice-based intervention for suspecting and detecting dementia in France: A cluster randomized controlled trial. *Dementia*, 7(4):433–450, nov 2008. ISSN 1471-3012. doi: 10.1177/1471301208096628.
17. S.T. Creavin, A.H. Noel-Storr, E. Richard, et al. Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people. *Cochrane Database of Systematic Reviews*, 2017(2), 2017. ISSN 1469493X. doi: 10.1002/14651858.CD012558.
18. Paul Russell, Sube Banerjee, Jen Watt, 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*, 3(12):e004023, jan 2013. ISSN 2044-6055. doi: 10.1136/bmjopen-2013-004023.
19. Sam T. Creavin, Sarah J. Cullum, Judy Haworth, et al. Towards improving diagnosis of memory loss in general practice: TIMeLi diagnostic test accuracy study protocol. *BMC Family Practice*, 17(1):79, 2016. ISSN 1471-2296. doi: 10.1186/s12875-016-0475-2.
20. Antoine Flahault, Michel Cadilhac, and Guy Thomas. Sample size calculation should be performed for design accuracy in diagnostic test studies. *Journal of clinical epidemiology*, 58(8): 859–62, aug 2005. ISSN 0895-4356. doi: 10.1016/j.jclinepi.2004.12.009.
21. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Author, Washington, DC, dsm-v edition, 2013.
22. Sharpley Hsieh, Samantha Schubert, Christopher Hoon, et al. Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease. *Dementia and geriatric cognitive disorders*, 36(3-4):242–50, jan 2013. ISSN 1421-9824. doi: 10.1159/000351671.
23. F Adshead, D D Cody, and B Pitt. BASDEC: a novel screening instrument for depression in elderly medical inpatients. *BMJ (Clinical research ed.)*, 305(6850):397, aug 1992. ISSN 0959-8138.
24. R S Bucks, D L Ashworth, G K Wilcock, and K Siegfried. Assessment of activities of daily living in dementia: development of the Bristol Activities of Daily Living Scale. *Age and ageing*, 25(2): 113–20, mar 1996. ISSN 0002-0729.
25. World Health Organization. *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. World Health Organisation, Geneva, 1993. ISBN 9241544554.
26. Paul A. Harris, Robert Taylor, Robert Thielke, et al. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2):377–381, apr 2009. ISSN 15320464. doi: 10.1016/j.jbi.2008.08.010.
27. Andrew J Vickers and Elena B Elkin. Decision curve analysis: a novel method for evaluating prediction models. *Medical decision making : an international journal of the Society for Medical Decision Making*, 26(6):565–74, 2006. doi: 10.1177/0272989X06295361.
28. Mark Fitzgerald, Benjamin R. Saville, and Roger J. Lewis. Decision Curve Analysis. *JAMA*, 313 (4):409, jan 2015. doi: 10.1001/jama.2015.37.
29. Andrew J. Vickers, Ben van Calster, and Ewout W. Steyerberg. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic and Prognostic Research*, 3(1):1–8, 2019. ISSN 2397-7523. doi: 10.1186/s41512-019-0064-7.
30. Andrew J Vickers, Ben Van Calster, and Ewout W Steyerberg. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*, 352, 2016. doi: 10.1136/bmj.i6.
31. K. Walters, S. Hardoon, I. Petersen, et al. Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data. *BMC Medicine*, 14(1):6, jan 2016. ISSN 1741-7015. doi: 10.1186/s12916-016-0549-y.
32. Jérémie F Cohen, Martin Chalumeau, Robert Cohen, et al. Cochran's Q test was useful to assess heterogeneity in likelihood ratios in studies of diagnostic accuracy. *Journal of clinical epidemiology*, 68(3):299–306, mar 2015. ISSN 1878-5921. doi: 10.1016/j.jclinepi.2014.09.005.
33. Martin Prince, Anders Wimo, Maëlenn Guerchet, et al. World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost And trends. Technical report, Alzheimer's Disease International, London, 2015.
34. NHS England. Numbers of Patients Registered at a GP Practice - April 2018. Technical report, NHS England, 2018.
35. Lucy C. Beishon, Angus P. Batterham, Terry J. Quinn, et al. Addenbrooke's Cognitive Examination III (ACE-III) and mini-ACE for the detection of dementia and mild cognitive impairment. *The Cochrane database of systematic reviews*, 2019. ISSN 1469493X. doi: 10.1002/14651858.CD013282.pub2.
36. Chris Salisbury, Sunita Procter, Kate Stewart, et al. The content of general practice consultations: cross-sectional study based on video recordings. *British Journal of General Practice*, 63 (616):e751–e759, nov 2013. ISSN 0960-1643. doi: 10.3399/bjgp13X674431.
37. NHS Digital. Selected CCG information: NHS Bristol, North Somerset and South Gloucestershire CCG. In *General Practice Workforce*. 2019.
38. Pim van den Dungen, Harm W M van Marwijk, Henriëtte E van der Horst, et al. The accuracy of family physicians' dementia diagnoses at different stages of dementia: a systematic review. *International journal of geriatric psychiatry*, 27(4):342–54, apr 2012. ISSN 1099-1166. doi: 10.1002/gps.2726.
39. Alex J Mitchell, Nicholas Meader, and Michael Pentzek. Clinical recognition of dementia and cognitive impairment in primary care: a meta-analysis of physician accuracy. *Acta psychiatrica Scandinavica*, 124(3):165–83, sep 2011. ISSN 1600-0447. doi: 10.1111/j.1600-0447.2011.01730.x.
40. Peter Croft, Douglas G Altman, Jonathan J Deeks, et al. The science of clinical practice: disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice. *BMC medicine*, 13:20, jan 2015. ISSN 1741-7015. doi: 10.1186/s12916-014-0265-4.
41. Daniel Kahneman. A perspective on judgment and choice: mapping bounded rationality - London School of Economics and Political Science. pages 1–35, 1996.
42. Timo Erkinjuntti, Truls Østbye, Runa Steenhuis, and Vladimir Hachinski. The Effect of Different Diagnostic Criteria on the Prevalence of Dementia. *New England Journal of Medicine*, 337 (23):1667–1674, dec 1997. ISSN 0028-4793. doi: 10.1056/NEJM199712043372306.
43. Michael Pentzek, Anja Wollny, Birgitt Wiese, et al. Apart from nihilism and stigma: what influences general practitioners' accuracy in identifying incident dementia? *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry*, 17 (11):965–75, nov 2009. ISSN 1545-7214. doi: 10.1097/JGP.0b013e3181b2075e.
44. Steve Iliffe, Jill Manthorpe, and Alison Eden. Sooner or later? Issues in the early diagnosis of dementia in general practice: A qualitative study. *Family Practice*, 20(4):376–381, aug 2003. ISSN 02632136. doi: 10.1093/fampra/cmg407.
45. S. Iliffe and Jane Wilcock. The identification of barriers to the recognition of, and response to, dementia in primary care using a modified focus group approach. *Dementia*, 4(1):73–85, feb 2005. ISSN 1471-3012. doi: 10.1177/1471301205049191.
46. Linda Lee, Loretta M Hillier, Paul Stolee, et al. Enhancing dementia care: a primary care-based memory clinic. *Journal of the American Geriatrics Society*, 58(11):2197–204, nov 2010. ISSN 1532-5415. doi: 10.1111/j.1532-5415.2010.03130.x.
47. Els J Meeuwsen, René J F Melis, Geert C H M Van Der Aa, et al. Effectiveness of dementia follow-up care by memory clinics or general practitioners: randomised controlled trial. *BMJ (Clinical research ed.)*, 344:e3086, may 2012. ISSN 1756-1833. doi: 10.1136/BMJ.E3086.
48. Beccy Baird. Primary care networks explained. Technical report, Kings Fund, London, 2019.
49. R. C. Petersen. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3):183–194, 2004. doi: https://doi.org/10.1111/j.1365-2796.2004.01388.x.
50. Brian Cooper, Horst Bickel, and Martina Schäufele. The ability of general practitioners to detect dementia and cognitive impairment in their elderly patients: A study in Mannheim. *International Journal of Geriatric Psychiatry*, 7(8):591–598, aug 1992. ISSN 0885-6230. doi: 10.1002/gps.930070809.
51. C D Pond, A Mant, L Kehoe, et al. General practitioner diagnosis of depression and dementia in the elderly: can academic detailing make a difference? *Family practice*, 11(2):141–7, jun 1994. ISSN 0263-2136. doi: 10.1093/fampra/11.2.141.
52. V G Valcour, K H Masaki, J D Curb, and P L Blanchette. The detection of dementia in the primary care setting. *Archives of internal medicine*, 160(19):2964–8, oct 2000. ISSN 0003-9926.

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.

Fig. 1. STARDdem flowchart for inclusion of participants in the study



∗ Potentially eligible population derived age specific incidence of dementia (28) and demographics of the population in the participating practices (34,956 aged over 70 years (29).
† one person had to withdraw part way through the reference test as they were acutely ill
‡ Dementia according to ICD-10 definition
§ of 61 with CIND 59 met criteria for Peterson MCI (49), 1 affective disorder, 1 brain injury
§ see Section
¶ 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.

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 $\geq$ 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 negative predictive value PPV positive predictive value

Fig. 2. Decision curve showing net benefit of clinical judgement for diagnosis of dementia at a range of threshold probabilities

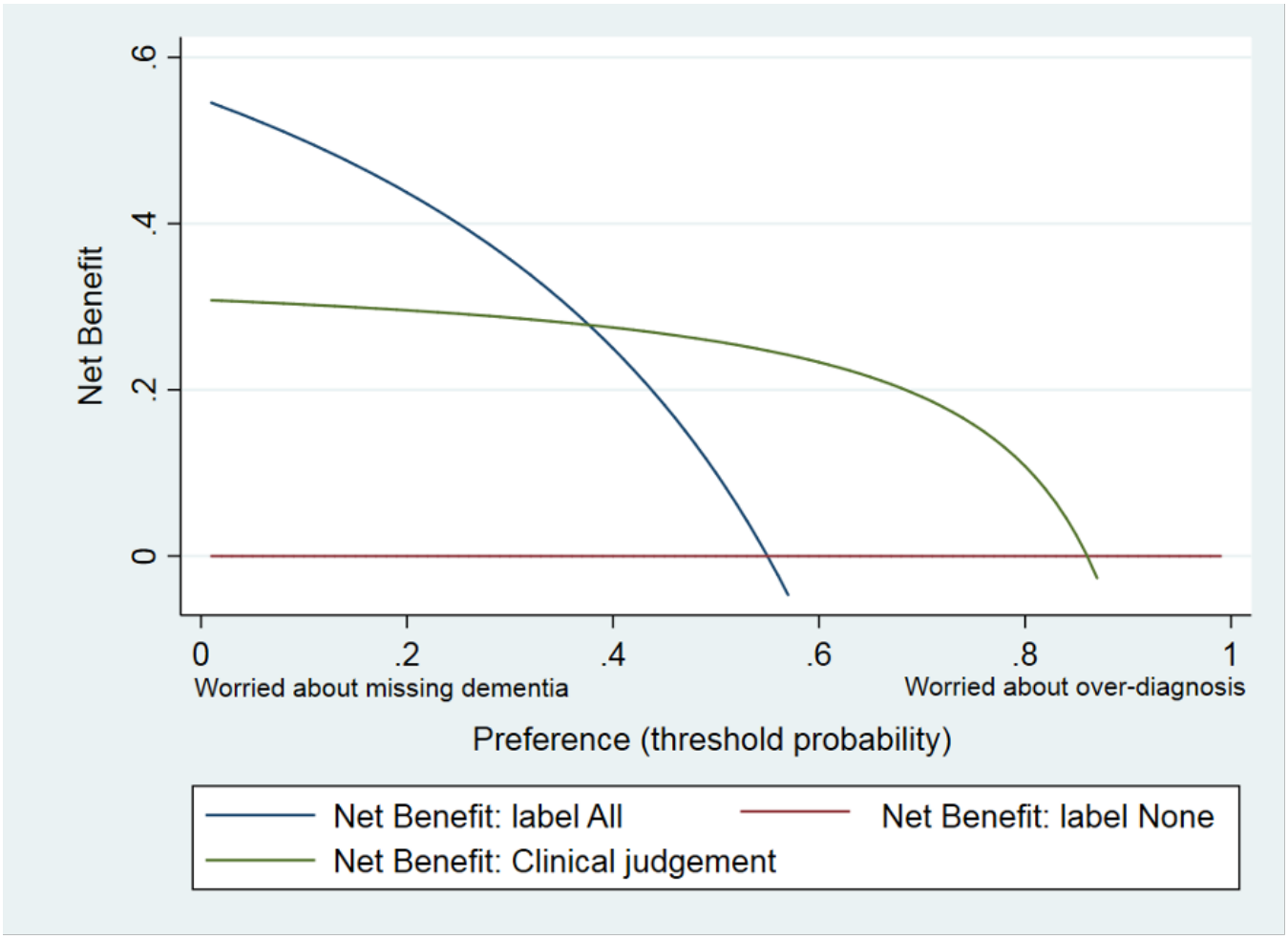


Table 3. Summary of seven studies investigating GP judgement for the diagnosis of dementia

	Mannheim (50)	Sydney (51)	Hawaii (52)	Antwerp (15)	AgeCoDe (14)	France (16)	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	1
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	1
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	2
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	2
9	Definition of and rationale for the units, cutoffs, and/or categories of the results of the index tests and the reference standard	2
10	The number, training, and expertise of the persons executing and reading the index tests and the reference standard. See also item 8	2
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	2
12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g., 95% confidence intervals)	2
13	Methods for calculating test reproducibility, if done	NA
14	When study was performed, including beginning and end dates of recruitment	3
15	Clinical and demographic characteristics of the study population (at least information on age, sex, spectrum of presenting symptoms). See also item 18	6
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	7
17	Time interval between the index tests and the reference standard, and any treatment administered in between	3
18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition	6
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	6
20	Any adverse events from performing the index tests or the reference standard	3
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	8
24	Estimates of test reproducibility, if done. See also item 13	NA
25	Discuss the clinical applicability of the study findings	3-4