Author:
Creavin, Sam T

Title:
Approaches to diagnosing dementia syndrome in general practice

Determining the value of clinical judgement and tests

Approaches to diagnosing dementia syndrome in general practice

Determining the value of clinical judgement and tests

By

SAMUEL THOMAS CREAVIN

Bristol Medical School
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Health Sciences.

14 JANUARY 2020

Word count: 65,283 words

ABSTRACT

People with cognitive symptoms often report a long wait to see a specialist in order to be given a diagnosis of dementia or some other disorder, or be told they are normal. This thesis investigates the accuracy of a range of tests potentially suitable for GP use, including GP clinical judgement, for the diagnosis of two target conditions, dementia and normal.

A systematic review was done in five electronic databases. A diagnostic test accuracy study was done in 21 GP surgeries in South West England with a total eligible population of 34,956. Qualitative interviews were done to explore the acceptability of GP diagnosis.

From 12,681 citations, 16 were included, referring to 10 studies, of which seven were included in a meta-analysis. For dementia, in the studies at lowest risk of bias, the sensitivity of GPs clinical judgement ranged from 34% (95% CI 18% to 54%) to 91% (95% CI 85% to 96%) and the specificity ranged from 58% (95% CI 51% to 66%) to 99% (95% CI 97% to 100%).

In 240 people aged 70 years or more without a previous diagnosis of dementia who had presented to their GP with cognitive symptoms, for dementia using ICD-10 criteria as judged by an expert the sensitivity of single tests ranged from 23% (95% CI 16% to 31%) for TAC to 100% (95% CI 97% to 100%) for MOCA; and specificity ranged from 4% (95% CI 1% to 11%) for Sniffin Sticks to 97% (95% CI 92% to 99%) for FAQ. GP judgement had sensitivity 56% (95% CI 47% to 65%) specificity 89% (95% CI 81% to 94%).

In 26 interviews, GP diagnosis of dementia was judged acceptable if the service was adequately resourced and met the specific needs of patients.

A number of candidate index tests were identified that could be further investigated to help select people with a high probability of dementia in general practice who may not need specialist evaluation for diagnosis.

DEDICATION AND ACKNOWLEDGEMENTS

I take responsibility for the design, conduct, analysis, interpretation and reporting of the work. I am very thankful for the support of colleagues and family. My supervisors, Professor Yoav Ben-Shlomo, Professor Sarah Purdy, Dr Sarah Cullum and Dr Lesley Wye, have inspired, encouraged, challenged and supported me.

The Wellcome Trust funded my salary and the research costs during the doctoral investigations (£321,248). The work also received pilot funding from 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). 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. I was the Principle Investigator on each of these grant applications.

Judy Haworth was remunerated for her work as the specialist at research clinics. Lizzie Elliot, Abi Sherlock and Alex Kwong were remunerated for their work in organising the research clinics. Rebecca Lefeuvre, Tom Sherlock, and Myles Hall were remunerated for doing data entry. Ryan Langdon was remunerated for data entry and scanning of case report forms. The staff at the West of England Clinical Research Network arranged for redaction, collection and transport of medical records from general practices.

My wife has been the best friend and team mate any person could hope for.

PUBLICATIONS ARISING FROM THE RESEARCH

The following papers have been published and are related to the work in the thesis.

1. Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people. **Creavin ST**, Noel-Storr AH, Richard E, Creavin AL, Cullum S, Ben-Shlomo Y, Purdy S. Cochrane Database of Systematic Reviews. 2017, Issue 2. Art. No.: CD012558.
2. Towards improving diagnosis of memory loss in general practice: TIMeLi diagnostic test accuracy study protocol. **Creavin ST**, Cullum SJ, Haworth J, Wye L, Bayer A, Fish M, Purdy S, Ben-Shlomo Y. BMC Fam Pract. 2016 Jul 19;17:79. Erratum in: BMC Fam Pract. 2016;17(1):119. PMID: 27430736

I conceived and wrote both papers, including drafting the manuscript, revising for important intellectual content, and responding to comments of co-authors.

The co-authors reviewed the draft manuscript for important intellectual content and approved the final manuscript.

Material from paper (1) is included in Chapter 2 Systematic Review. Material from paper (2) is included in Chapter 3 Methods.

SIGNED: 

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

TABLE OF CONTENTS

List of Tables

List of Figures

ACRONYMS

C_{95} 95th centile of duration. 113, 139

$\Downarrow$ lower scores indicate better cognition. 138, 139

$\Uparrow$ higher scores indicate better cognition. 138, 139

$\bar{x}$ mean. 135, 138, 139

95% CI 95% confidence interval. i, 14–16, 25, 66, 67, 69, 78, 81, 82, 86, 88, 93, 112, 120, 128, 131, 133–135, 137–139, 141, 143, 146, 147, 150, 152, 153, 155, 157–159, 162, 163, 166, 168–172, 199–201, 203, 212, 217, 222, 336, 337

•• threshold for dementia. 112, 138, 139

6CIT Six Item Cognitive Impairment Test. 104, 105, 112, 137–139, 145, 146, 150, 152, 153, 155, 157–159, 162, 163, 165–170, 172, 218, 222, 223, 333, 334, 336–341, 346

ACE3 Addenbrooke's Cognitive Examination III. xv, xviii, 107, 111, 112, 120, 131–136, 169, 170, 177, 179, 200, 201, 206, 215, 219, 220

AD8 Galvin AD8 Dementia Screening Interview. 40, 105, 112, 113, 133, 141, 143–146, 150, 153, 155–159, 162, 168, 169, 171, 172, 206, 336–338, 340, 341

ADDTC Alzheimer's Disease Diagnostic and Treatment Centers. 22

ADL Katz index of activities of daily living. 105, 112, 113, 141, 143, 144, 146, 150, 152, 153, 155, 157–159, 162, 163, 165, 166, 168–170, 172, 315, 332, 333, 336–341

AGECAT Automated Geriatric Examination for Computer Assisted Taxonomy. 52

AMT abbreviated mental test. 106

AUROC area under ROC curve. xvi, 34, 38–40, 118, 142, 144, 159, 162

BADL Bristol activities of daily living questionnaire. 107

BASDEC Brief Assessment Schedule Depression Cards. 107

BCA brief cognitive assessment. 218–220

BNSSG Bristol, North Somerset, and South Gloucestershire area. 97, 100, 109, 194, 207

CAMCOG Cambridge Cognitive Examination. 15

CAMDEX Cambridge Mental Disorders of the Elderly Examination. 52, 58–61, 64

CDR Clinical Dementia Rating scale. 23, 24, 52, 56, 58, 59, 61

CERAD Consortium to Establish a Registry for Alzheimer's Disease protocol. 59

CIDI World Health Organisation Composite International Diagnostic Interview. 56

CIND cognitive impairment, but not dementia. 62, 101, 113, 131, 133, 142, 155, 205, 206, 210, 218, 220

CSF cerebrospinal fluid. 22

DOR diagnostic odds ratio. 34, 39, 40, 67, 69, 87, 88, 113, 142, 147, 150, 153, 157, 158, 166, 336, 337

DSD delirium superimposed on dementia. 28

DSM-III-R Diagnostic and Statistical Manual of Mental Disorders (3rd ed., Revised). xvii, 4, 9, 12, 52, 55, 58–61, 71–73, 75–77, 79, 80

DSM-IV-TR Diagnostic and Statistical Manual of Mental Disorders (4th ed., Text Revision). xvii, 8, 9, 11, 12, 52, 55, 58, 59, 61, 71, 72, 217

DSM–5 Diagnostic and Statistical Manual of Mental Disorders (5th ed.). 7–9, 11, 12, 22, 26, 29, 129, 131, 135, 141

EPSS Extra pyramidal signs scale. 105, 112, 138, 139, 146, 150, 153, 157, 158, 162, 200, 336–341, 346

FAQ Functional Activities Questionnaire. i, 105, 112, 113, 141, 143–146, 150, 152, 153, 155, 157–159, 162, 163, 165–170, 172, 214, 218, 222, 223, 315, 332, 333, 336–341

FDG fluorodeoxyglucose (^{18}F). 22

FN false negative. 34–36, 40, 45, 66, 69, 78, 89, 90, 113, 145, 146, 152, 217

FP false positive. 34–36, 40, 45, 66, 69, 78, 87, 89, 90, 113, 145, 146, 152, 163, 217

GDS global deterioration scale. 24, 26

GP General Practitioner. i, iii, xii, xiii, xv–xviii, 31, 45–47, 49–53, 55–58, 60–64, 66, 67, 69, 70, 72–74, 76, 78, 82–85, 87–93, 95–101, 103, 104, 106, 107, 109, 111–113, 116–120, 122, 123, 125, 128, 131–134, 139, 141, 142, 145–150, 152, 153, 155–159, 162, 163, 165–172, 174–177, 180–182, 184–189, 191, 192, 194, 197–208, 210–223, 287, 310, 315, 316, 332–334, 336–343

GPCOG General Practitioner Assessment of Cognition. 104, 105, 112, 116, 137–139, 145, 146, 150, 152, 153, 155, 157–159, 162, 166, 169, 170, 172, 176, 218, 336–341, 346

HBM Health Belief Model. 183

IADL Lawton instrumental activities of daily living scale. 61, 105, 112, 113, 141, 143, 144, 146, 150, 152, 153, 157, 158, 162, 163, 165, 166, 169, 170, 334, 336–341

ICD-10 The International Classification of Diseases, Tenth Revision. i, xvii, 4, 8, 9, 11, 12, 52, 55, 58, 60, 61, 71–77, 79–81, 96, 107–109, 122, 123, 129, 131, 132, 135, 139, 141, 142, 165, 200, 217, 218

ICD-11 The International Classification of Diseases, Eleventh Revision. 9

IQCODE Short Form of the Informant Questionnaire on Cognitive Decline in the Elderly. 40, 86, 87, 89, 105, 112, 113, 133, 141, 143–146, 150, 152, 153, 155–158, 162, 163, 165–172, 176, 201, 206, 218, 219, 222, 223, 333, 334, 336–341

IQR interquartile range. 131–133, 135, 206, 211

kDa kilodaltons. 5, 18

LRN negative likelihood ratio. 34, 36, 113, 146, 150, 152, 153, 155, 157, 158, 166, 171, 172, 199–201, 217, 222, 336, 337

LRP positive likelihood ratio. 34, 36, 113, 118, 146, 150, 152, 153, 155, 157, 158, 163, 166, 171, 172, 199–201, 212, 216, 217, 222, 223, 336, 337

MAT Memory alteration test. 105, 106, 112, 137–139, 146, 150, 152, 153, 155, 157, 158, 162, 166, 168, 200, 201, 206, 287, 310, 315, 336–341, 346

MCI Mild Cognitive Impairment. 7, 8, 23–26, 28, 52–54, 56, 58–61, 96, 107, 108, 112, 120, 122, 123, 129, 131–135, 137–142, 163, 165, 169, 170, 179, 200, 201, 203, 206, 209, 220, 223

MIS memory impairment screen. 106

MMSE Mini Mental State Examination. xviii, 56, 59–61, 64, 78, 86–89, 114–116, 218

MOCA Montreal Cognitive Assessment. i, 106, 112, 133, 137–139, 145, 146, 150, 152, 153, 156–159, 162, 166, 168, 169, 171, 172, 200, 201, 206, 212, 214, 219, 287, 315, 336–341, 346

MRI Magnetic Resonance Image. 43

NHS National Health Service. 98, 101, 103, 107, 109, 127, 176, 178, 203, 204, 207, 215, 216, 221, 281

NINCDS-ADRDA National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association. 19, 52, 53, 58, 61

NINDS-AIREN National Institute of Neurological Disorders and Stroke and the Association Internationale pour la Recherche et l'Enseignement en Neurosciences. 20, 22, 52

NPT Normalization Process Theory. 183

NPV negative predictive value. 34, 36, 38, 113, 150, 153, 157, 158, 166, 336, 337

PAF population attributable fraction. 15, 16

PPV positive predictive value. 34, 36, 38, 104, 113, 150, 153, 157, 158, 166, 216, 222, 336, 337

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses. xvii, 49, 57

Q1 first quartile. 134, 136, 140

Q3 third quartile. 134, 136, 140

QUADAS-2 Quality Assessment of Diagnostic Accuracy Studies. xvii, xviii, 49, 54, 55, 63, 65, 66, 69, 72–75, 78, 81–83, 85, 87, 88, 93, 95, 121, 202, 209, 211

RBD rapid eye movement sleep behaviour disorder. 22

REDCap Research Electronic Data Capture. 110, 111, 287

ROC receiver operating characteristic. xvii, xviii, 1, 34, 38–40, 54, 55, 66, 68, 69, 71–73, 75, 78, 113, 117, 125, 144, 147, 200, 315, 316, 332

s standard deviation. 131, 133, 135, 138, 139

SCI subjective cognitive impairment. 24, 25

SIDAM structured interview for the diagnosis of dementia of the Alzheimer type, multi-infarct dementia and dementias of other aetiology according to ICD-10 and DSM-III-R. 52, 58, 61

SOP standard operating procedure. 101–103, 238

SPECT Single-photon emission computed tomography. 21, 22

SPMSQ Short Portable Mental Status Questionnaire. 59

SPMT Scenery Picture Memory Test. 105, 112, 138, 139, 145, 146, 150, 152, 153, 155–159, 162, 166, 168–170, 172, 201, 214, 336–341, 346

TAC Time and change. i, 105, 112, 139, 145, 146, 150, 152, 153, 156–159, 162, 163, 165, 166, 168–170, 172, 200, 201, 214, 315, 332, 336, 338–341, 346

TDP-43 TAR DNA-binding protein with molecular weight 43 kilodaltons. 18

TDQ Telephone Dementia Questionnaire. 59

TFA Theoretical framework of acceptability. 183, 184, 188–193, 196, 201

TN true negative. 34–36, 40, 45, 66, 69, 78, 87, 90, 113, 145, 152, 217

TP true positive. 34–36, 40, 45, 66, 69, 78, 89, 90, 113, 116, 145, 152, 163, 217

TPB Theory of Planned Behavior. 183

TUG Timed up-and-go. 105, 112, 137–139, 145, 146, 150, 152, 153, 157, 158, 162, 166, 169, 336–341, 346

CHAPTER 1

INTRODUCTION

This Chapter introduces some important topics in the thesis, specifically dementia, diagnostic accuracy, and clinical judgement. Presented first is a summary of important literature regarding dementia, including definitions, epidemiology and aetiology, histology, clinical features, diagnosis, clinical staging, investigations, complications, and prognosis. Section two of this Chapter reviews the literature relating to diagnosis, diagnostic tests, diagnostic accuracy, and approaches to combining tests. Section three discusses literature regarding clinical judgement, explaining the concept and giving illustrative evidence for the accuracy of clinical judgement in different clinical contexts. The Chapter closes by discussing the scientific rationale for the empirical work in the context of some existing related literature, and sets out the aims and objectives for this thesis.

1.1 Background

1.1.1 Dementia

1.1.1.1 Definitions of dementia

The term dementia refers to a clinical syndrome of persistent cognitive impairment, from a previous higher level of functioning, that has a significant deleterious impact on the ability to perform activities of daily living [1]. Dementia may result from different pathological processes, which are outlined below in Section 1.1.1.2. Dementia should be distinguished from delirium, a fluctuating disturbance of attention, awareness and cognition that develops rapidly and is attributable to a disrupted physiology or multiple aetiologies [2].

Mild Cognitive Impairment (MCI), Section 1.1.1.7), is a clinical syndrome of cognitive decline that is greater than expected for age and educational attainment, but which is not associated with significant impact on activities of daily living [3]. The most recent edition of the Diagnostic and

Statistical Manual of Mental Disorders (5th ed.) (DSM–5) [2] refers to major and minor neurocognitive disorder in place of dementia and MCI respectively. MCI is discussed further in Section 1.1.1.7. DSM–5 defines major neurocognitive disorder as:

- *"A. Evidence of significant cognitive decline from a previous level of performance in one or more cognitive domains (complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition) based on:*
  - *Concern of the individual, a knowledgeable informant, or the clinician that there has been a significant decline in cognitive function; and*
  - *A substantial impairment in cognitive performance, preferably documented by standardized neuropsychological testing or, in its absence, another quantified clinical assessment.*
- *B. The cognitive deficits interfere with independence in everyday activities (i.e., at a minimum, requiring assistance with complex instrumental activities of daily living such as paying bills or managing medications).*
- *C. The cognitive deficits do not occur exclusively in the context of a delirium*
- *D. The cognitive deficits are not better explained by another mental disorder (e.g., major depressive disorder, schizophrenia)"*

In contrast the Diagnostic and Statistical Manual of Mental Disorders (4th ed., Text Revision) (DSM-IV-TR) [4], which is used in much of the published literature, offers the following definition: *"The essential feature of a dementia is the development of multiple cognitive deficits that include memory and at least one of the following: aphasia, apraxia, agnosia, or a disturbance in executive functioning. The cognitive deficits must be sufficiently severe to cause impairment in occupational or social functioning and must represent a decline from a previous high level of functioning. A diagnosis of dementia should not be made if the deficits occur exclusively during the course of delirium."* The third common definition in the literature is The International Classification of Diseases, Tenth Revision (ICD-10) [5] definition which states that for a diagnosis of dementia *"There is evidence of each of the following:*

- *G1*[1]
  - *(1). A decline in memory, which is most evident in the learning of new information, although in more severe cases, the recall of previously learned information may also be affected. The impairment applies to both verbal and non-verbal material. The decline should be objectively verified by obtaining a reliable history from an informant, supplemented, if possible, by neuropsychological tests or quantified cognitive assessments. The*

[1] G1 is one of the criteria

severity of the decline, with mild impairment as the threshold for diagnosis should be assessed[2].

– *(2). A decline in other cognitive abilities characterized by deterioration in judgement and thinking, such as planning and organizing, and in the general processing of information. Evidence for this should be obtained when possible from interviewing an informant, supplemented, if possible, by neuropsychological tests or quantified objective assessments. Deterioration from a previously higher level of performance should be established. The severity of the decline, with mild impairment as the threshold for diagnosis, should be assessed.*

- *G2. Awareness of the environment is preserved during a period of time sufficiently long to allow the unequivocal demonstration of the symptoms above.*
- *G3. There is a decline in emotional control or motivation, or a change in social behaviour manifest as at least one of the following: emotional lability, irritability, apathy, coarsening of social behaviour.*
- *G4. For a confident clinical diagnosis the symptoms in criteria G1 should have been present for at least six months."*

The three definitions outlined above are summarised and compared in Table 1.1. The three definitions are consistent in requiring there to be a decline in cognition from a previous level, highlighting that dementia may be diagnosed even if an individual performs in the normal range, or even above average on formal cognitive testing. The most significant differences between the definitions are: that ICD-10 and DSM-IV-TR require memory impairment in conjunction with impairment in other domains, whereas DSM–5 only requires impairment in one domain and does not require amnesia; that only ICD-10 requires impairment in emotional regulation or social interaction; and that DSM-IV-TR and ICD-10 require impairment of daily life to be attributable to *each* of the impaired cognitive domains. Finally ICD-10 states that for a "confident" diagnosis the symptoms should be present for at least six months. Two other definitions should be mentioned. Firstly, the Diagnostic and Statistical Manual of Mental Disorders (3rd ed., Revised) (DSM-III-R) [6] is used in some of the literature prior to 1995 and defines dementia as *"impairment in short- and long-term memory, associated with impairment in abstract thinking, impaired judgement, other disturbances of higher cortical function, or personality change. The disturbance is severe enough to interfere significantly with work or usual social activities or relationships with others. The diagnosis of Dementia is not made if these symptoms occur. . .in Delirium"* [6]. Secondly, The International Classification of Diseases, Eleventh Revision [7] (ICD-11) is a classification system that was defined in 2018 that aligns to DSM–5

[2] Mild: a degree of memory loss sufficient to interfere with everyday activities though not so severe as to be incompatible with independent living... The main function affected is the learning of new material. For example, the individual has difficulty in registering, storing and recalling elements in daily living, such as where belongings have been put, social arrangements, or information recently imparted by family members

in using the term *neurocognitive disorder* in conjunction with dementia, and defines this as *"an acquired brain syndrome characterised by a decline from a previous level of cognitive functioning with impairment in two or more cognitive domains (such as memory, executive functions, attention, language, social cognition and judgement, psychomotor speed, visuo-perceptual or visuo-spatial abilities). The cognitive impairment is not entirely attributable to normal ageing and significantly interferes with independence in the person's performance of activities of daily living. Based on available evidence, the cognitive impairment is attributed or assumed to be attributable to a neurological or medical condition that affects the brain, trauma, nutritional deficiency, chronic use of specific substances or medications, or exposure to heavy metals or other toxins."*

Table 1.1: Comparison of three definitions of dementia

Domain	**Definition**		
	DSM–5	*DSM-IV-TR*	*ICD-10*
Cognition	Significant cognitive decline	Development of multiple cognitive deficits	Decline in multiple cognitive domains
Domains affected	One or more of: complex attention executive function learning and memory language perceptual-motor social cognition	Memory impairment AND aphasia apraxia agnosia executive functioning	Memory decline AND judgement and thinking general processing of information
Evaluation	Based on subjective concern AND objective measurement	Not specified	Reliable history AND quantitative tests if possible
Activities of daily living	Interference with independence in daily life	Significant impairment in both domains sufficient to interfere with social or occupational functioning to an extent that is a decline from a previous level	Memory and other cognitive abilities which both interfere with daily living
Delirium	Not exclusively in delirium	Not exclusively in delirium	Delirium not present when assessed
Emotions & social	Not specified	Not specified	Decline in emotional control or change in social behaviour manifest as emotional lability irritability apathy coarsening social behaviour
Duration	Not specified	Not specified	At least six months

DSM–5 Diagnostic and Statistical Manual of Mental Disorders (5th ed.) *DSM-IV-TR* Diagnostic and Statistical Manual of Mental Disorders (4th ed., Text Revision) *ICD-10* The International Classification of Diseases, Tenth Revision

Figure 1.1: Prevalence of dementia according to different definitions in five studies

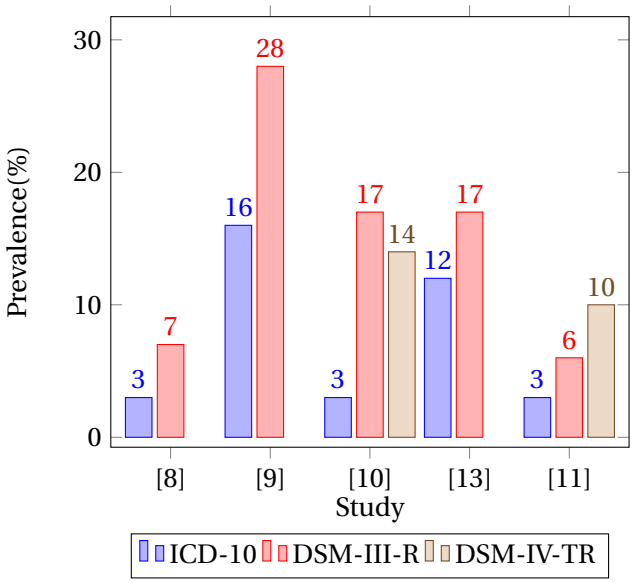


Figure 1.1 illustrates that studies have reported a lower prevalence of dementia according to ICD-10, (3% [8] 16% [9] and 3 % [10]) than DSM-III-R (7% [8], 28% [9] and 17% [10]) and DSM-IV-TR (14%[10]). Critically, while the different definitions have some overlap they classify different people as having dementia, with ICD-10 identifying people who have more advanced dementia [3] [10]; with implications for individuals, families, and society. Compared to ICD-10, DSM-III-R and DSM-IV-TR include more cases with mild dementia and have a trend towards a shorter duration of symptoms [10]. In the identified studies, DSM-III-R typically had higher prevalence than ICD-10 and DSM-IV-TR, with the exception of one study [11] which had a higher prevalence of DSM-IV-TR dementia. One study (not plotted) shows that DSM–5 and DSM-IV-TR dementia have similar prevalence [12].

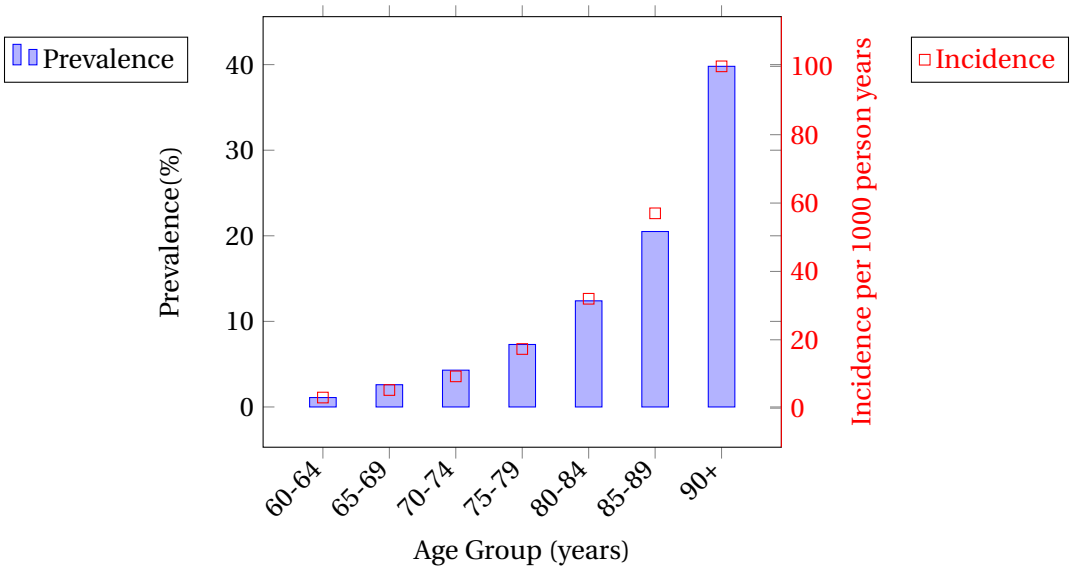
1.1.1.2 Epidemiology and Aetiology

Incidence and Prevalence Figure 1.2 illustrates that dementia is increasingly common with age, with the prevalence in Western Europe approximately doubling with every fifth birthday over 60 years [14]. Around 1% of the population are estimated to have dementia at ages 60-64 years, based on meta-analysis of 65 studies, and this rises to around 40% of the population aged 90 years and over [14]. In some areas, including Western Europe, the prevalence in men is around 15% lower than that in women (not shown in Figure 1.2). The incidence of dementia, based on 18 studies in a meta-analysis, also rises with age from 3 per 1000 person-years at ages 60-64 years to 122 per 1000 person-year at ages 90+ year [14].

The prevalence of dementia is similar in other parts of the world, regardless of the extent of economic development [14], and there is no definite indication of change over the past 30 years

[3] In [10] dementia was graded as mild, moderate or severe according to DSM-III-R

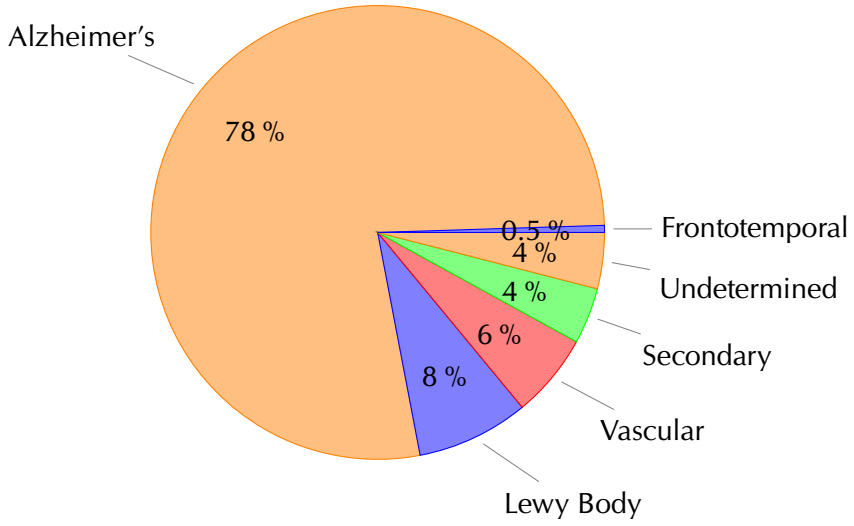
Figure 1.2: Prevalence and incidence of dementia by age in Western Europe [14]



[15]. In contrast, there is some evidence that in high income countries, the incidence of dementia has decreased over the past 30 years [15–17]. However, in Nigeria the incidence is reported to be stable over time [18] and indeed in China [19] and Japan [20] the incidence has been reported to have risen, though this is contested [21]. To the extent that incidence of dementia has fallen, there has been no reduction in the workload for clinicians that is associated with dementia, with recorded diagnosis of dementia remaining stable with an increasing trend from 1992-2014 in the Netherlands [22] and increasing in the UK from 2006-2018 [23]. These findings are possibly due to diagnosis at an earlier stage in the disease [22], a shift towards a milder disease profile with less deterioration over time [24], or improved life expectancy in late life resulting in larger numbers of older people compared to previous decades. To the extent that there *is* a fall in the incidence of dementia, it can be at least partly attributed to better education of the population and control of hypertension [15] and vascular risk factors, but there are also concerns that because obesity and diabetes are becoming more common [16] there could be a reversal of the decline in incidence [25]. Overall, the impact is that the number of people living with dementia worldwide is anticipated to increase from 50 million in 2018 to 152 million in 2050 [25] because of increased life expectancy [26].

Aetiology While there are a number of definitions for dementia as a syndrome, there are also a number of specific aetiologies. Clinical aspects of the different aetiologies are outlined in Section 1.1.1.4. However, attributing a fraction of dementia syndrome in the population to a particular aetiology is complicated because in older people with dementia, who comprise the majority [27], the clinical presentation may not correspond to the neuropathology [28–30], with discordance between the clinical diagnosis and the neuropathology [31, 32]. Indeed, the association between dementia and Alzheimer's pathology weakens with age [33] and it is common for people with dementia to have

Figure 1.3: Prevalence of dementia attributed to clinical aetiologies [37]



more than one pathology at the time of death [30]. These findings have led some investigators to advocate the idea that neuropathology should be regarded together with the clinical and imaging data, and *"trumped"* [34] by the clinical presentation in understanding which conditions might have affected a patient in life [34].

In general, the proportion of dementia attributable to different causes varies by age, with Lewy body dementias being more common in people aged over 65 years [35] and frontotemporal dementia occurring more often in people aged younger than 65 years [36], accounting for between 3%-26% of dementia in this age group [36]. Figure 1.3 shows the prevalence of the different clinical diagnoses of dementia in a community based sample of 2,170 people aged over 65 years in Spain [37]. Alzheimer's disease was the most common clinical diagnosis, which is typical [38], with other diagnoses being less common, though vascular dementia was unusually rare and more often accounts for around 15% of cases [39]. A second study of 1085 community-dwelling people in the UK found the prevalence of dementia was 10%, of whom 31% (95% 95% confidence interval (95% CI), 21%-43%) had probable Alzheimer's disease, 22% (95% CI 14%-34%) had vascular dementia, 11% (95% CI 5%-21%) had Lewy body dementia and 8% (95% CI 3%-17%) had frontotemporal dementia, with overlap between clinical diagnoses occurring not infrequently. Notwithstanding these studies of the patterns of clinical dementia by clinical aetiological type [37, 40], reports indicate that the majority of older community-dwelling people have neuropathological lesions at autopsy, regardless of cognitive status in life, that clinical dementia is associated with more than one type of neuropathology [41–44] and that in the oldest-old the extent of clinical cognitive impairment is positively associated with the number of neuropathologies [45]. Furthermore, as discussed in Section 1.1.1.3, increasing age attenuates the association between neuropathology and aetiological sub-type and it is uncommon for older people with dementia to have a single neuropathology.

Risk factors Risk factors for dementia are considered as conferring a risk of dementia syndrome, rather than a specific aetiology, because public health is improved by the reduction of dementia regardless of aetiology [46]. There is some weak evidence that women are more susceptible to dementia than men [47] but other studies suggest that the risk is similar and that women may be affected at earlier ages [48, 49]. Some risk factors for dementia, such as genetic profile [48, 50], have no scope to be modified in the foreseeable future. Others, such as early life experiences and socioeconomic adversity [51], are difficult to modify without substantial social-political programmes and investment. Conversely, many risk factors for dementia are potentially modifiable, with the opportunity for up to one third of cases to be prevented [52] based on observational studies. Based on observational studies the biggest single factor in reducing preventable cases of dementia may be in improving crystallised intelligence, lifetime intellectual activity and cultural exposure [46]. Experimental designs investigating the causal effect of improved education on reduced dementia incidence report mixed findings, with one indicating a 10% (95% CI 4% to 15%) reduction in dementia risk per year of schooling [53], another finding no evidence of an effect [54], a systematic review of mendelian randomisation studies reporting no evidence, albeit potentially due to insufficient power [55], and a two-sample multivariable mendelian randomisation study reporting that with each standard deviation increase in years of schooling and intelligence, odds of Alzheimer's dementia were on average 37% and 35% lower [56]. Table 1.2 presents the population attributable fraction (PAF) for various risk factors at different stages of the life course and shows that in later life the most important modifiable risk factors for dementia are smoking, depression, physical inactivity, social isolation, and diabetes. Because the data in Table 1.2 are based on observational studies it is important to note that the associations may be (partly) attributable to residual or unmeasured confounding, or (with depression, inactivity and isolation) reverse causation. In addition to factors that increase risk for developing dementia, there are also factors which might improve cognitive reserve, *"the ability to optimize or maximize normal performance"* [57], which has been proposed as a concept to explain the individual variation in cognitive decline. In a population based study, a cognitive reserve score calculated using the formula

$$\text{Cognitive reserve score} = 1.7 \times \text{years of education} + 1 \times \text{occupational complexity level} \tag{1.1}$$

was found to mediate 21% of the variance in Cambridge Cognitive Examination (CAMCOG) scores that was attributable to four lifestyle factors (physical activity, diet, alcohol, cognitive and social activity) [58].

Genetics It is unusual for dementia syndrome to be caused by a single gene mutation, as seen in mendelian inheritance, and single genetic mutations have a low positive predictive value for dementia [59]. Despite this, family history is often relevant clinically as people with a first degree relative with non-mendelian Alzheimer's disease have a lifetime risk that is 2.5x the population risk [59]. A family history of psychiatric disorders and the age of dementia onset are important in determining the risk of genetically linked dementia, both Alzheimer's disease and frontotemporal dementia, with

Table 1.2: Modifiable risk factors for dementia [51]

Risk factor	**Relative risk for dementia** (95% CI)	**Prevalence of risk factor**	**PAF** ∗
	Early life age < 18 years		
Less education (none or primary school only)	1.6 (1.3-2.0)	40%	20%
	Mid life age 45-65 years		
Hypertension	1.6 (1.2-2.2)	8.9%	5.1%
Obesity	1.6 (1.3-1.9)	3.4%	2.0%
Hearing loss	1.9 (1.4-2.7)	32%	23%
	Later life age >65 years		
Smoking	1.6 (1.2-2.2)	27%	14%
Depression	1.9 (1.6-2.3)	13%	10%
Physical inactivity	1.4 (1.2-1.7)	18%	6.5%
Social isolation	1.6 (1.3-22)	11%	5.9%
Diabetes	1.5 (1.3-1.8)	6.4%	3.2%

∗ *PAF* population attributable fraction

the probability that a person has a genetic mutation rising with the number of affected first degree relatives and the age of onset [59]. Dementia may also be a feature of other neurodegenerative disorders, notably Huntington's disease, autosomal dominant Lewy body dementia (which is typically sporadic) [59], and Parkinson's disease.

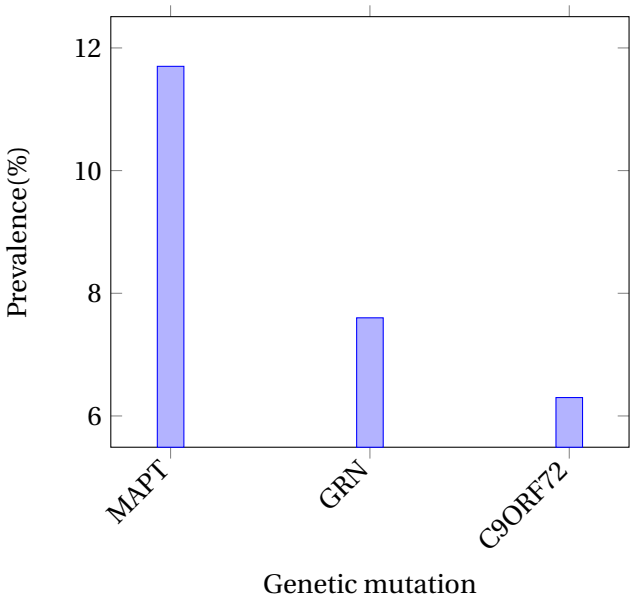
Most Alzheimer's dementia with an onset over age 65 years is due to sporadic forms of the disease, with genetic mutations found in only 1% of people who have two or more first degree relatives with onset at this stage in the life course [59]. Genetic mutations are also rare in people with early onset Alzheimer's disease who have no family history. The three most common genetic mutations in Alzheimer's disease are *PSEN1* on chromosome 14 (60% of mendelian cases), *APP* on chromosome 21 (15% - 23% of mendelian cases), and *PSEN2* on chromosome 1 (which is rare) [59]. Each of these three mutations has a role in the cellular processing of amyloid-β which is a key component of plaques (Section 1.1.1.3). Other genes contribute to the risk of Alzheimer's disease in a complex (i.e. non-mendelian) way. For example, people with ApoE E genotype E2/E2 have a lower risk of Alzheimer's disease (0.5x population risk) whereas those with genotypes E3/E4 and E4/E4 have respectively 3x and 8x the population risk; despite this up to 75% of people with the E4 allele do not develop Alzheimer's disease [59].

In contrast to Alzheimer's disease, 40%-50% of people with frontotemporal dementia have a relevant family history, and between 10%-30% of them have an autosomal dominant inheritance pattern [59]. The probability of identifying a genetic mutation in someone with a family history of frontotemporal dementia is positively associated with younger age at onset and increasing number of affected relatives [59]. Figure 1.4 shows that *MAPT* is the most common genetic mutation in

frontotemporal dementia.

There is overlap between the genetics of Lewy body dementia, Parkinsons's disease, Parkinson's disease dementia, and Alzheimer's disease. Mutations in the *SNCA* and *LRRK2* genes are associated with a rare form of autosomal dominant inherited Lewy body dementia [35]. There are few studies investigating the genetics of vascular dementia, and no conclusive evidence of definite genetic risk factors, other than that a frameshift mutation in the notch gene on chromosome 19 is related to cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) [39].

Figure 1.4: Prevalence of genetic mutations in frontotemporal dementia [59]



1.1.1.3 Pathology

The hallmark of Alzheimer's disease pathology is the accumulation of abnormally folded plaques of amyloid protein comprised of Amyloid-β and tau, together with neuronal tangles [38]. However, in older people with dementia the association with plaques and tangles is weaker, whereas the association with cerebral atrophy remains [33]. In dementia associated with *PSEN1* Amyloid-β accumulates in hydrophobic peptides of increasing length as amyloid precursor protein is cleaved less efficiently [38]. Neurodegeneration can result from accumulated Tau protein which may act in parallel or independently of Amyloid-β [38]. The medial temporal lobe is preferentially affected by atrophy in Alzheimer's disease [38].

In contrast frontotemporal dementia is associated with changes in the frontal lobes, anterior temporal lobes, and cortex (anterior cingulate and insular), with neurodegeneration manifesting as gliosis, neuronal loss and microvacular changes [36]. Nearly all frontotemporal lobar degeneration is attributable to three abnormal proteins. TAR DNA-binding protein with molecular weight 43 kilodaltons (kDa), TDP-43 accounts for almost 50% of cases of frontotemporal lobar degeneration and has three main subtypes which have characteristic cytoplasmic or intranuclear pathology [36]. Microtubule associated protein tau accounts for 35%-50% of frontotemporal lobar degeneration, most often manifest as Pick's disease, corticobasal degeneration, and progressive supranuclear palsy, each of which accounts for 1/3 of cases of frontotemporal lobar degeneration-tau and is associated with characteristic histology [36]. Finally fused-in-sarcoma accounts for around 10% of cases of frontotemporal lobar degeneration and is associated with abundant immunoreactive inclusion bodies in the dentate gyrus and severe atrophy of the striatum [36].

Lewy body dementias are associated with inclusion bodies of α-synuclein and neuronal loss,

which are proposed to spread from cell to cell, and appear to be synergistic with amyloid-β but not vascular pathologies [35]. Vascular dementia is associated with a variety of neuropathology, including both macro- and micro-vascular disease in cortical and subcortical regions, as well other vascular processes such as amyloid angiopathy [39].

Notwithstanding the typical or classic descriptions of neuropathology associated with each dementia aetiology, older adults with late-onset dementia typically have mixed pathology at death [33, 42]. Furthermore, abnormalities that are classically associated with dementia, such as neurofibrillary and vascular pathology, are found at death in people both with and without clinical dementia in life [42]. A greater pathological load is associated with a higher likelihood of clinical dementia [29, 44], though it is uncertain whether this association may [31], or may not [28], persist in people aged more than 90 years.

1.1.1.4 Clinical aspects

Table 1.3 presents the clinical features for each dementia aetiology. Often the specific symptoms of particular dementia aetiologies are more obvious in the initial or early stages of the process, and as the disease progresses there is often increasing disruption of cognitive function, leading to apathy, withdrawal, and dependency in activities of daily living. Behavioural and psychological symptoms of dementia are common with many dementia aetiologies, especially frontotemporal dementia, and are often particularly distressing for patients and their kin.

Alzheimer's disease is typically associated with gradually progressive impairment in memory over months, sometimes years, with early decline in the formation of new memories subsequently leading to diminishing executive function and difficulties with other cognitive domains [38]. The clinical features of Alzheimer's disease are associated with the age at diagnosis, with higher odds of non-memory symptoms, especially abnormal visuospatial processing, at lower ages of first presentation [60]. Posterior cortical syndrome (not in table) [61] is a rare presentation (perhaps 5% of cases) often, but not always, associated with Alzheimer's disease pathology and may present with symptoms of anxiety, visual problems (especially with lines of text, judging distances and climbing stairs), and parietal lobe dysfunction such as dyspraxia, dyscalculia and dyslexia [61]. The standard diagnostic criteria are based on working groups from the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [62, 63].

Frontotemporal dementia, in contrast to Alzheimer's disease, is much less likely to manifest with memory impairment in the early stages. Frontotemporal dementia is described as having three presentations, one of which predominantly affects behaviour, and two of which (the progressive aphasias) are associated with progressive deficits in speech, grammar and word output which are the exclusive symptoms for the first few years. Firstly, *behavioural variant* frontotemporal dementia is associated with early behavioural deficits including apathy, personality changes and disinhibition [36]. Patients may be hyper-oral with an appetite for sweet food or alcohol, less sympathetic and

empathetic leading to difficulties in personal relationships and limited insight, socially unruly leading to public displays of nudity or overtly sexual comments (though often with impaired libido), and suffer from disrupted motivation resulting in repetitive stereotyped behaviour whilst becoming increasingly apathetic [36]. Secondly, *semantic variant primary progressive aphasia* is manifest as incomprehension and mispronunciation of isolated words, especially those infrequently used by the patient, but the syntax and fluency of speech is initially maintained so that repetition may not be impaired [36]. Behavioural symptoms develop as the disease progresses, and are influenced by the lateralisation of the neuropathology, right and left sided degeneration resulting in respectively verbal and visual compulsions [36]. Finally, *Non-fluent variant primary progressive aphasia* is characterised by awkwardness in the production of speech, with inappropriate additions, deletions or pauses impeding the fluency of speech, though understanding and pronunciation of single words is preserved [36].

Vascular dementia may manifest with a variety of presentations, depending on where the vascular damage has occurred in the brain, but is classically associated with impaired executive function and attention with relative preservation of memory [39]. Apathy and depression characterise the neuropsychiatric symptoms of vascular dementia, but there is considerable overlap with Alzheimer's disease and other more unusual symptoms may be present such as delusions and hallucinations [39]. Vascular dementia may present as part of vascular parkinsonism, a disease of older people often manifest as gradual onset gait disturbance particularly of the lower body, postural instability, falls, incontinence, and pseudobulbar effect [64]. The standard diagnostic criteria for much of the literature is based on a working group from the National Institute of Neurological Disorders and Stroke and the Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN) [65].

Lewy body dementias can be difficult to diagnose. Firstly, dementia with Lewy bodies can present similarly to Alzheimer's disease; secondly, the onset of Parkinson's disease *dementia* in someone with Parkinson's *disease* can be insidious [35]. Once established, Parkinson's disease dementia is indistinguishable clinically from dementia with Lewy bodies, because the conditions are differentiated on the basis of the time course: in dementia with Lewy bodies dementia presents before, or within one year, of spontaneous parkinsonism, whereas Parkinson's disease dementia presents after Parkinson's disease has been present for a number of years. Dementia with Lewy bodies is characterised by a gradually progressive cognitive impairment, perhaps especially affecting visuospatial skills, which is seen in 74% of people compared to 45% of those with Alzheimer's disease [35]. Executive function, and attention are other commonly affected domains in dementia with Lewy bodies, though these are likely to be less useful to diagnose the aetiology given the overlap with vascular and Alzheimer's dementia. Other features of dementia with Lewy bodies may include: fluctuating severity, well formed visual hallucinations, rapid eye movement sleep disorder (expressed as a person acting out or vocalising vivid, often persecutory dreams), spontaneous parkinsonism relatively early in the disease course, sensitivity to antipsychotics, falls, and autonomic dysfunction [35]. The clinical diagnostic criteria for

dementia with Lewy bodies identify only 32% of people who meet the neuropathological criteria at autopsy, with greater concordance associated with florid $\alpha-$synuclein pathology and sparse neuritic plaques [35]. Investigations can be helpful to refine the diagnosis of dementia with Lewy bodies from Alzheimer's disease. Single-photon emission computed tomography (SPECT) demonstrating low dopamine transporter uptake may be used to distinguish dementia with Lewy bodies from Alzheimer's disease, but not from other parkinsonian syndromes [35].

Table 1.3: Clinical aspects of dementia by aetiology

Aetiology	**Clinical features**	**Investigations**	**Definitions**	**Treatment**
Alzheimer's disease [38]	Impaired memory and executive function *Atypical presentations* with language, visual, practic, or executive function initially	*CSF* Amloid β_{42}, total-tau & phosphorylated tau (p-tau) *Imaging* medial temporal lobe atrophy *Nuclear imaging* FDG pattern temporoparietal and posterior cingulate	McKhann [62, 63]	acetylcholinesterase inhibitors Glutamate Antagonists
Frontotemporal dementia [36]	Three variants , converging over time, gradually developing global cognitive impairment & motor deficits: parkinsonism; motor neuron disease	*Nuclear imaging* FDG pattern anterior, asymmetric, or both	Neary [66] Brun [67] Boxer [68]	Symptomatic
Vascular dementia [39]	varied *subcortical* lesions result in inattention, processing & executive function.	imaging demonstrating changes to account for clinical presentation	NINDS-AIREN [65] ADDTC [69] & DSM–5 [2]	Symptomatic
Lewy body dementia [35]	Visuospatial defects and RBD common, well formed visual hallucinations	*CSF* $\alpha-$synuclein may be lower than in Alzheimers *Nuclear imaging* occipital hypometabolism, relatively preserved in posterior cingulate	McKeith [70, 71]	Acetylcholinesterase inhibitors Compression stockings

CSF cerebrospinal fluid *FDG* fluorodeoxyglucose (^{18}F) *ADDTC* Alzheimer's Disease Diagnostic and Treatment Centers
NINDS-AIREN National Institute of Neurological Disorders and Stroke and the Association Internationale pour la Recherche et l'Enseignement en Neurosciences
SPECT Single-photon emission computed tomography *DSM–5* Diagnostic and Statistical Manual of Mental Disorders (5th ed.)
RBD rapid eye movement sleep behaviour disorder

1.1.1.5 Diagnosis

Dementia usually presents as a gradual onset of progressive cognitive symptoms over months to years. A history of decline in a cognitive domain from a previous higher level is the hallmark symptom, which is often (though not always) verified by a person who knows the patient before their symptoms started. A clinical evaluation of someone with symptoms that could indicate dementia often begins with understanding who has raised concerns about the possibility of cognitive problems, the patient, their kin, or a health professional. An understanding of the nature and impact of the perceived cognitive deficits, onset, fluctuation, progression, trigger to seek help, and other affected domains is essential. Elicitation of the persons level of education, occupational role(s), family history (including of psychiatric disease), and exposure to risk factors (e.g. alcohol, smoking, trauma, and toxins) may also be helpful. Understanding the symptoms in the context of the persons medical, social and psychological history, as well as their medications, current comorbidities, and wishes for the future is an important part of providing holistic, responsive care, and may be relevant to the diagnostic evaluation.

Cognitive testing can be helpful in measuring objective impairment, though the evaluation should be targeted based on the history, so that there is confidence that the evaluation will test the cognitive domains where problems have been reported. In some cases a specialist neuropsychological evaluation is required to ascertain the full range of cognitive deficits. Often the physical examination of someone with dementia will be normal, but there may be signs of previous vascular insult, such as hemiplegia, or of a prior neurological disease, such as Parkinson's disease, and some investigators have reported clinical signs such as the head turn sign [4], applause sign, [5] or came alone sign [6] [72].

The most important alternative diagnoses to be considered in the evaluation of someone with possible dementia are delirium (Section 1.1.1.9), MCI (Section 1.1.1.7), and psychiatric conditions (such as depression). However, the differential diagnosis is broad and may include functional memory disorder, trauma (causing subdural haematoma), cerebral neoplastic disease (either metastatic cancer, or more rarely primary cancer), neurodegenerative conditions (such as normal pressure hydrocephalus), immunologically mediated disease (such as encephalitis or paraneoplastic phenomenon) or auto-immune (multiple sclerosis). Typically these alternative diagnoses are discounted on the basis that dementia is more in keeping with the clinical history, especially the time course, and of investigations, especially neuroimaging.

1.1.1.6 Clinical staging

The Clinical Dementia Rating scale (CDR) [73] was designed to stage Alzheimer's disease but can also be used to stage other dementias [74]. An interview is used to judge the capacity of an individual in six domains, memory, orientation, judgement and problem solving, community affairs, home and

[4] the patient looks to an accompanying person to help support them answering questions
[5] when asked to clap three times the person claps more than this indicating possible perseveration
[6] to clinic, supporting a a non-dementia diagnosis

hobbies, and personal care. The scale can be used to derive a global scale, which is calculated using an algorithm [75] and typically used for purposes of staging dementia with scores ranging from 0 (normal) through 0.5 (questionable), 1 (mild), 2 (moderate), to 3 (severe). An alternative approach is to score the sum of boxes, which results in a more detailed quantitative index ranging from 0 to 18 and provides more useful information in mild dementia [76–78]. An alternative clinical staging system is the global deterioration scale (GDS) which can be used to rate the severity of a primary neurodegenerative condition [79]. The GDS has seven stages from 0 to 7 (most severe), with stages 1-3 being pre-dementia and stages 4-7 being the stages in dementia. CDR 2 (moderate dementia) approximately corresponds to GDS 5 (moderate dementia), as both stages describe people at this stage as continent and being able to recall names but disorientated to time and place; GDS 6 relates approximately to CDR 3 described as forgetting names, disturbed diurnal rhythm, and dependency in activities of daily care, with variable incontinence; in contrast GDS 7 (describing generalised rigidity, incontinence and mutism) corresponds to a yet more advanced stage of disease than CDR 3.

A challenge for patients, clinicians, researchers and policy-makers is that in the hypothetical model of Alzheimer's disease pathology the disease develops many years before symptoms are present [80]. Alzheimer's disease has a long preclinical phase with pathology accumulating many years (perhaps decades) before symptoms develop, once the brain has accumulated a substantial load of Amyloid-β [81]. The pathological burden at which the disease manifests is not uniform, and may be influenced by cognitive reserve (Section 1.1.1.2) and comorbidities. As neuropathology develops a person may present to health services with subjective cognitive impairment (SCI), which is deficits in memory but objective cognitive tests scores within the normal range [82]. SCI in the absence of objective cognitive impairment is reported by around 17% of people in the community [83], and of this group 19% have dementia and 32% have MCI [84], (Section 1.1.1.7).

People with SCI who do not currently meet criteria for dementia or MCI have an annual risk of progression to these states which is approximately 2% for dementia and 6% for MCI [83]. At population level SCI is more likely to be an indicator of affective disorder than an organic cognitive disorder, and investigators have reported that functional impairment of activities of daily living may be a helpful way to identify a group of people who are most likely to have an underlying cognitive disorder [82], though necessarily this would mean that people with cognitive disease that is less advanced will not be identified. However, in the general practice population a history of depression and anxiety are also both associated with an increased risk of dementia [85].

One approach to identify people who have the highest risk of developing dementia is to evaluate the risk of future dementia [86] using a risk score, though many of these scores cannot be based on routinely collected data [87–89]. One model using routine data may be useful to identify people aged 60-79 years who are at high risk of a diagnosis of dementia within the subsequent five years, but unfortunately this did not extend to people aged 80+ years [90]. The model to predict risk of dementia in the subsequent five years in people aged 60-79 years comprised predominantly vascular risk factors (age, sex, smoking, current anti-hypertensives, body mass index, diabetes, cerebrovascular

disease, current aspirin use), together with harmful alcohol use, depression and atrial fibrillation [90]. The incidence of dementia in the group aged 60-79 years was 1.9 (95% CI 1.8–1.9) per 1000 person years at risk compared to 17 (95% CI 16-17) per 1000 person years at risk in the 80+ years age group. The discrimination and calibration (see Section 1.1.2.5) of the risk algorithm were good for the group aged 60-79 years: 2.0 (95% CI 1.9 to 2.1) C 0.84 (95% CI 0.81 to 0.87) slope 0.98 (95% CI 0.93 to 1.0) but were poor in the group aged 80-85 years: D 0.86 (95% CI 0.76 to 0.95) C 0.56 (95% CI 0.55 to 0.58) slope 1.0 (95% CI 0.89 to 1.2) [90]. Therefore while more of the older-old population develop dementia annually than than the younger-old, risk prediction is more difficult in the older group. An alternative model for estimating risk of dementia over a 10-year time frame using information that might be routinely available in primary care was derived from and validated in a population based cohort study of 2,170 people aged over 60 years (mean age 71 years) [91]. In this study, age, history of stroke, subjective memory decline, and need for assistance with finances or medication (which is less likely to be routinely available in general practice records) had good discrimination for predicting the risk of dementia with a c-statistic of 0.78 (95% CI 0.75-0.81), though once again the discrimination was lower in the 465 people aged over 80 years (c statistic 0.57, 95% CI0.49 - 0.63) [91].

People with SCI may derive some benefit from group sessions on education about cognitive symptoms, or cognitive training, but the evidence for these interventions is weak [92], and there is no evidence that people who are highest risk of progressing to dementia or MCI are more likely to benefit.

1.1.1.7 Mild Cognitive Impairment

MCI is a syndrome of cognitive decline where a person has cognitive function that is not normal for age, with minimal impairment of the activities of daily living to the extent that it is insufficient to meet criteria for dementia [3, 93–95]. MCI may occur either with memory impairment (amnestic-MCI) or without (non-amnestic-MCI), with more restrictive definitions unsurprisingly being reported to have a lower prevalence in the population [96]. Similarly to dementia, additional years of education reduce the risk of MCI, there is no clear evidence that sex is a risk factor [96], and the prevalence increases with age. MCI is identified in 6.7% (95% CI 3.4% - 13%) of people at ages 60-64, rising to 8.5% (95% CI 5.2% - 13%) at 65-69 years, 10% (95% CI 7.5% - 14%) at ages 70-74 years, 15% (95% CI 10% - 21%) at ages 75-79 years and 25% (95% CI 17% - 37%) at ages 80-84 years [96]. MCI is not *necessarily* a pre-dementia condition, and some people with MCI may find that their cognition remains stable. In a meta-analysis of 41 robust inception cohorts of people with MCI the annual conversion rate from MCI to dementia was 6.7% (95% CI 4.6%-9.1%), with between 30%-50% of people developing dementia with long term surveillance [97], suggesting that many people with MCI do not develop dementia.

Indeed, in a study of 357 people aged over 75 years with MCI in general practice 42% found their symptoms improved at three years [98]; other studies in different settings have reported that between 14% and 56% of people with MCI revert to normal [96]. The uncertainty about the disease course

can cause ethical questions about the value of designating a clinical presentation as MCI as this will inevitably raise concerns for the patient and their kin about the risk of developing dementia, perhaps unnecessarily. The uncertain natural course of MCI is compounded by the apparent reduction in the risk of conversion to dementia over time, with the risk of dementia being highest shortly after identification of MCI, perhaps especially in the first two years [99], and reducing with time [97]. Furthermore, compared to people with normal cognition at baseline, people with amnestic-MCI who have improvement in their cognitive symptoms remain at higher risk of subsequently developing dementia, though they are at lower risk than people with stable amnestic-MCI [99]. With the exception of exercise twice-weekly as part of a holistic approach to management that includes advance care planning and serial cognitive evaluation, there are no disease-modifying interventions that are recommended unequivocally for people with MCI, though it is is advisable to avoid any modifiable factors (such as medication) that may have an adverse impact on cognitive performance [96]. In DSM–5 [2] MCI is referred to as *Minor Neurocognitive Disorder*, which is defined as:

- *"A. Evidence of modest cognitive decline from a previous level of performance in one or more cognitive domains (complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition) based on:*
  - *Concern of the individual, a knowledgeable informant, or the clinician that there has been a mild decline in cognitive function AND*
  - *A modest impairment in cognitive performance, preferably documented by standardised neuropsychological testing or, in its absence, another quantified clinical assessment.*
- *B. The cognitive deficits do not interfere with capacity for independence in everyday activities (i.e., complex instrumental activities of daily living such as paying bills or managing medications are preserved, but greater effort, compensatory strategies, or accommodation may be required).*
- *C. The cognitive deficits do not occur exclusively in the context of a delirium.*
- *D. The cognitive deficits are not better explained by another mental disorder (e.g., major depressive disorder, schizophrenia)."*

1.1.1.8 Complications

As dementia progresses, the specific features of the individual aetiologies often become less apparent as the neuropathology becomes more generalised and there is more global cognitive impairment. Advanced dementia is described in GDS 7 as: "*All verbal abilities are lost ... Frequently there is no speech at all only unintelligible utterances and rare emergence of seemingly forgotten words and phrases. Incontinent of urine, requires assistance toileting and feeding. Basic psychomotor skills, e.g., ability to walk, are lost with the progression of this stage. The brain appears to no longer be able to tell the body what to do. Generalized rigidity and developmental neurologic reflexes are frequently*

Table 1.4: Common symptoms in advanced dementia [102]

Symptom	Cumulative incidence	Mortality within six months
Eating problems	85%	39%
Febrile episode	53%	45%
Agitation	54%	‡
Dyspnoea ∗	48%	‡
Pneumonia	41%	47%
Aspiration	41%	‡
Pain ∗	39%	‡
Pressure ulcer †	39%	‡

∗ ≥ five days per month † ≥ grade 2 ‡ not reported
Eating problems include weight loss, chewing problems, dysphagia, refusal to eat, suspected dehydration, or persistent food refusal
Pneumonia defined as documentation by physician, nurse practitioner, or physician assistant
Febrile episode, excluding pneumonia, defined as oral temperature ≥37.8°C; rectal ≥38.3°C; or axillary, ≥37.2°C

present." [79]. Table 1.4 shows that the most common clinical problems encountered in the later stages of dementia are with eating, pyrexia, agitation, dyspnoea and chest infection [100]. Colonisation with antimicrobial resistant organisms is also common, occurring in 48% of people with advanced dementia, and may be due to liberal use of broad spectrum antimicrobials with relatively little objective evidence of systemic infection [101]. Pain and agitation are also common symptoms, [102, 103], which are often managed with opiods, and anxiolytics respectively [103]. Opiods may also be used to manage terminal dyspnoea [103]. Advance care planning, aiming to maximise comfort and avoid unnecessary or potentially burdensome interventions, can help to improve the quality of care in the late stages of dementia [100] and is important because 96% of health care proxies consider comfort to be the principle goal for care at the end of life [102]. At the very end of life, delirium is also a common problem for people with advanced cognitive impairment [104].

Dementia shortens life expectancy, even allowing for the effects of age and co-morbidity [51], though some have hypothesised the secular trend that survival with dementia may be increasing [15]. Population based studies have reported that median survival after a diagnosis of dementia is between 3.3 years [105] and 4.4 years [106]. In contrast, a study based on primary care records of 22,529 people with dementia coded, and 112,645 people without dementia coded, reported that median survival from recorded diagnosis was 6.7 years in those aged 60-69 years and 1.9 years in those aged 90 years and over [107]. The important distinction of the primary care based study is that cases were those who were entered on the general practice record, which may be systematically under-recorded [108], and this effect may be greatest in older people. Older age at symptom onset, male sex, higher education, history of diabetes or depression symptoms, and greater cognitive decline have been associated with shorter life expectancy [106, 109].

1.1.1.9 Delirium

Delirium is a acute disturbance in attention and cognition, characterised by impaired consciousness, fluctuating severity, and typically occurring in the context of an acute illness. The interrelation between delirium and dementia is intricate. In a person who has MCI or dementia, superimposed delirium results in an acute deterioration in their clinical status. Delirium is a syndrome and is not, in itself, a diagnosis; it may be attributable to one or many aetiologies, commonly infection (e.g. urinary tract, respiratory tract), medication effect (especially those with sedative or anti-cholinergic action), trauma, metabolic disruption (e.g. electrolyte disturbance), or physiological abnormalities (e.g. hypoxia), or occur peri-operatively, among other causes [110]. Delirium may be identified using clinical tests such as the 3D-CAM [111], 4-AT [112] or OSLA [113].

Whether a person manifests delirium in an episode of acute illness is influenced by the background of the individual and characteristics of the acute illness episode. Prior factors influencing tendency to delirium include previous episodes, age, and underlying cognitive impairment [114]. Characteristics of the acute episode include the number of acute parallel stresses, their nature or likelihood of causing delirium, and the severity of the illness(es) [115]. While the probability of delirium is related to the prior susceptibility and the acute instigating factors of illness, it is a matter of ongoing research whether these factors also influence the probability of long term cognitive impairment after delirium [116, 117] independent of baseline cognition. However, it is clear that an episode of delirium in an older aged adult increases the risk of future dementia, and it is possible this is not mediated through classic dementia neuropathology [118, 119].

Identifying delirium superimposed on dementia (DSD) can be very difficult because there is no agreed definition [120], but evaluating the extent to which motor function is compromised may help to differentiate DSD from dementia [121]. Typically delirium develops rapidly, over hours or days, and is associated with impaired attention and consciousness, whereas dementia without delirium develops over months or years, and is not usually associated with inattention or obtundation. Conversation in delirium is typically incoherent and disorganised, whereas in dementia while there may be aphasia in severe stages the speech is usually ordered and has some logic, even if it is not lucid. The crucial distinction for identifying DSD may be the report from relatives that the person is behaving abnormally and out of character [110].

Delirium in a person who is apparently cognitively normal when well is especially complex because it may herald incipient MCI or dementia, and a number of plausible mechanisms for this are debated. Firstly, delirium may compromise cognitive reserve strategies that have previously masked latent impairment, thus revealing undiagnosed dementia. Secondly, the triggering factors for an episode of delirium may themselves be harmful to the brain and these may act alone (e.g. through ischaemic, inflammatory, or apoptotic neural insults) or in synergism, to initiate or accelerate any existing dementia neuropathology, resulting in neurodegeneration [110]. Thirdly, delirium may itself mediate a directly deleterious impact on the brain regardless of the underlying aetiology [110]. Delirium may persist for an extended period, sometimes months or years following an acute insult

and is associated with expedited cognitive decline following the acute episode: even if dementia does not develop the person may not return to their baseline level of performance before the episode of delirium, though factors that mediate this are an ongoing area of research [110]. In DSM–5 [2] delirium is defined as:

- *"A. Disturbance in attention (i.e., reduced ability to direct, focus, sustain, and shift attention) and awareness (reduced orientation to the environment).*
- *B. The disturbance develops over a short period of time (usually hours to a few days), represents an acute change from baseline attention and awareness, and tends to fluctuate in severity during the course of a day.*
- *C. An additional disturbance in cognition (e.g.memory deficit, disorientation, language, visuospatial ability, or perception).*
- *D. The disturbances in Criteria A and C are not better explained by a pre-existing, established or evolving neurocognitive disorder and do not occur in the context of a severely reduced level of arousal such as coma.*
- *E. There is evidence from the history, physical examination or laboratory findings that the disturbance is a direct physiological consequence of another medical condition, substance intoxication or withdrawal (i.e. due to a drug of abuse or to a medication), or exposure to a toxin, or is due to multiple etiologies."*

Expert opinion has advised an inclusive approach to evaluating items (A) and (D) in DSM–5, because of concerns that it might be impossible to formally evaluate inattention, and because delirium is well recognised to cause impaired consciousness [122].

1.1.1.10 Treatment

Identifying a drug to modify the course of disease in dementia is a current focus of global research efforts [123]. Even if a disease modifying therapy existed, arguably there would be greater benefit for the population in investing in primary prevention, to reduce risk of dementia and improve cognitive reserve [51, 124].

In the absence of a disease modifying therapy, the focus of treatment for people with dementia should be a holistic care approach to the medical, social and psychological needs of the patient and their relatives [51]. The diagnosis of a progressive incurable condition may be an opportunity to consider future plans, including for future care at times of crises or in the terminal phase of life. Some specific symptoms, such as cognition and neuropsychiatric symptoms may benefit from interventions with drugs, but often a holistic approach thinking creatively about social approaches, problem solving, promoting inclusivity and independence can be as helpful. For people with dementia, music based stress reduction may improve depressive and behavioural symptoms [125], reminiscence

therapy may help to improve quality of life and communication [126], and exercise programs may support people with dementia to perform activities of daily living [127]. Group based cognitive stimulation programs improve cognition in people with mild-moderate dementia [128]. For people who are caring for those with dementia, mindfulness based stress reduction techniques may be helpful [129] and telephone support may help to reduce depressive symptoms [130]. Recommendations on treatment are limited by the lack of high quality primary research; even for interventions which seem valid and may be commonly utilised in clinical practice, such as respite care [131], case management [132], reducing antipsychotic prescribing [133] or thickening food to help with swallowing problems [134].

Acetylcholine is a neurotransmitter with roles in the peripheral and central nervous system; a deficit of cholinergic transmission is hypothesised to be part of the pathology in Alzheimer's disease [135]. The deficit of acetylcholine in dementia can be partially ameliorated by inhibition of the enzyme acetylcholinesterase which degrades the neurotransmitter in the synaptic cleft. Available cholinesterase inhibitors include donepezil (tablet or orodispersible tablet) [136], galantamine (capsule) [137] and rivastigmine (patch, capsule or liquid) [138]. The three medicines have slightly different pharmacological properties but all have the same end result [139] and have been shown to improve cognitive symptoms in people with Alzheimer's dementia and Lewy body dementia.

1.1.2 Diagnosis, diagnostic tests, and diagnostic accuracy

1.1.2.1 Diagnosis

One way to conceptualise diagnosis is that it is a label, or method of classification [140], which is applied to a particular group of patients to help to understand the aetiology, prognosis and potential treatment options of a particular clinical scenario. One definition of diagnosis is:

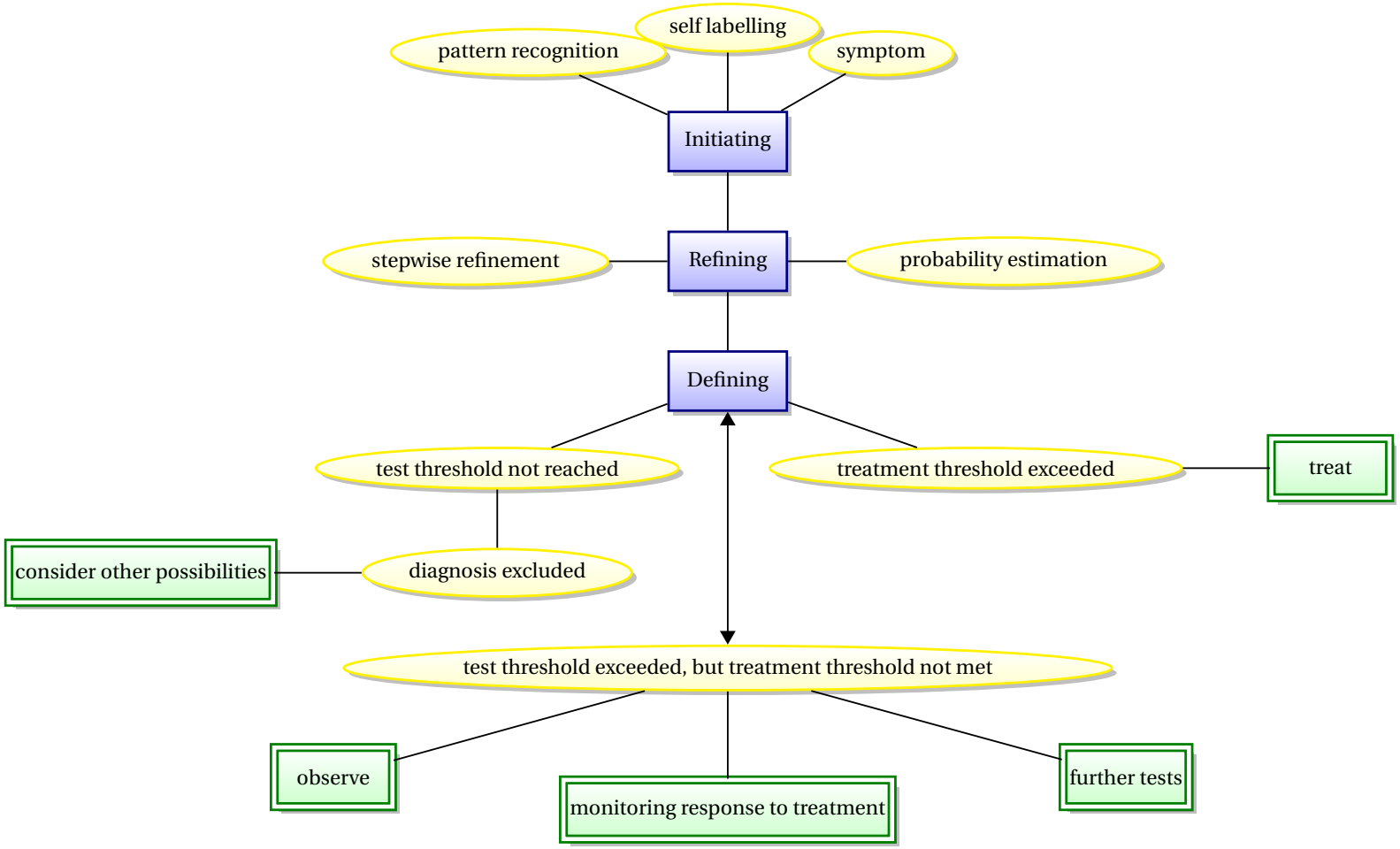
"the process of determining the health status and the factors responsible for it; [it] may be applied to an individual, family, group, society. The term is applied both to the process of determination and to its findings" [141].

A range of strategies are used by clinicians to help to formulate a diagnosis in any particular clinical scenario [142]. Broadly, the diagnostic process can be understood as having three stages, each with a range of options that may be used by the clinician. Firstly, *initiating the diagnostic hypotheses* may result from the patient self labelling with a diagnosis or stating a symptom which is parsed into a presenting complaint by the clinician. Alternatively the clinician may use pattern recognition based on experience to generate a range of hypotheses or make a "spot diagnosis", such as acne, where the refinement stage is skipped. A second stage in the diagnosis process *refines the diagnosis* by eliciting further information and iteratively testing the hypotheses against the evidence. Approaches at this second stage of the diagnostic process include stepwise refinement (e.g. of upper or lower urinary tract infection), or estimating the probability of diseases, either implicitly using pattern recognition or explicitly using a clinical prediction rule. The third stage of diagnosis, is to *"define the final diagnosis"* [142], though this could also be worded *define the management*. At the third stage

of diagnosis either the diagnosis is known and can be treated confidently, or further information is required either by using further tests, monitoring the response to treatment, or observing the scenario over time, when the symptoms may become more easily classified or spontaneously resolve. This third stage of diagnosis has been described as implicitly comparing the probability of disease against a treatment threshold, treatment is indicated when the probability of disease for a patient exceeds the threshold, and a test threshold, if the probability of disease in a patient falls below the threshold then further testing is not indicated [143]. An example:

*A 75 year old man attends his GP surgery concerned that he is forgetting the names of relatives and recently was unable to find his way home. He is concerned that he has dementia (**hypothesis initiation by self labelling**). The clinician parses this into a **hypothesis generating presenting complaint** of memory loss. The clinician elicits further detail to **refine the diagnosis**, establishing that there are no other cognitive concerns and that that no other person is concerned. The patient and the clinician weigh the probability of disease against the options for further testing or treatment and decide that further information is needed before a **final diagnosis** can be reached. The probability of disease is judged to not exceed the **threshold** for treatment or invasive testing at this stage and so it is agreed to use a test of time and review in two months.*

Figure 1.5: Stages in diagnostic reasoning



While the threshold for testing continues to be exceeded, defining the diagnosis iterates as the patient is observed, monitored in response to treatment, or investigated further until the threshold treatment is exceeded.

The three stage diagnostic process outlined above is typically executed rapidly and without conscious thought by experienced clinicians, which can be described as being a type 1 process in the dual model of cognition [144, 145]. The topic of the dual model of cognition is discussed further in Section 1.1.3.

1.1.2.2 Diagnostic tests

As described in Section 1.1.2.1 in the broadest sense a clinical test could be any instrument that helps a clinician and patient to refine the options that are available to management of a particular clinical scenario, such as a test of time or a response to treatment. However, as more commonly understood diagnostic tests are defined as:

"A test to diagnose whether or not a person has a disease or disorder. Not all biomedical (laboratory, imaging, genetic, other) tests have this aim. To protect individuals and groups from avoidable adverse effects of diagnostic activities, it is essential to distinguish related but different aims, such as diagnosing clinically overt disease; discovering occult disease (new process, recurrences); secondary and tertiary prevention; determining the stage, characteristics, and activity of the disease; counselling; or monitoring the clinical course and effects of therapies" [141].

The term diagnostic tests underestimates the role of tests in clinical medicine, because tests are not synonymous with diagnosis. Some clinical tests, such as blood tests, imaging, or cellular pathology are investigations. Investigations may have an important role in diagnosis to help refine the probability of a particular disease, but they may also be used for other purposes in clinical practice, such as: screening for disease, refining diagnosis within a category (e.g. at the molecular or genetic level), guiding treatment and response to treatment, helping to determine prognosis, or providing reassurance [146]. This thesis will not consider the use of tests for purposes other than diagnosis.

1.1.2.3 Diagnostic accuracy

The range of options for clinicians to reach a diagnosis is instructive as it highlights the broad range of possible "tests" that could be used in practice. A test does not need to be a blood test, or a component of the clinical evaluation, but could be a response to treatment or a change in the presentation over time. Some have advocated that in the absence of indicators for immediate action such as serious illness, the test of time is the most important diagnostic strategy in low prevalence settings such as general practice [147]. As trivial symptoms without a significant disease aetiology resolve spontaneously the prevalence of serious pathology in a given clinical profile increases, this has the benefit of avoiding unnecessary actions in the majority of people who have a low probability of disease. As the prevalence and spectrum of disease in a particular clinical profile changes, the characteristics of test accuracy also vary.

For a single test, the test accuracy can be conveniently displayed in a two-by-two table. The result of the test that is under investigation, termed the *index test*, for a particular *target condition* is tabulated against the best available information about the presence of absence of that target

disease, based on the *reference standard* [148]. Table 1.5 presents cross tabulated hypothetical data on the presence of disease against the result of a test. Where there is discordance between the index test and the reference standard the results of the index test are said to be false, and where there is concordance they are said to be true, resulting in the terms true positives (TP), which is test positive disease positive, and analogously false positives (FP), true negatives (TN) and false negatives (FN).

Table 1.5: Two by two table of test accuracy characteristics

	Disease positive (n)	**Disease Negative (n)**	**Total (n)**
Test positive	84 (a)	11 (b)	95 (c)
Test negative	16 (d)	89 (e)	105 (f)
Total	100 (h)	100 (i)	200 (j)

a true positives; b false positives; d false negatives; e true negatives
c test positives; f test negatives; h disease positives; i disease negatives
j population; $\frac{h}{j}$ prevalence
in a strict epidemiological sense $\frac{h}{j}$ is *not* a prevalence but is more accurately a % of cases in a sample; however prevalence is the recognised term in the diagnostic literature e.g. see [148]

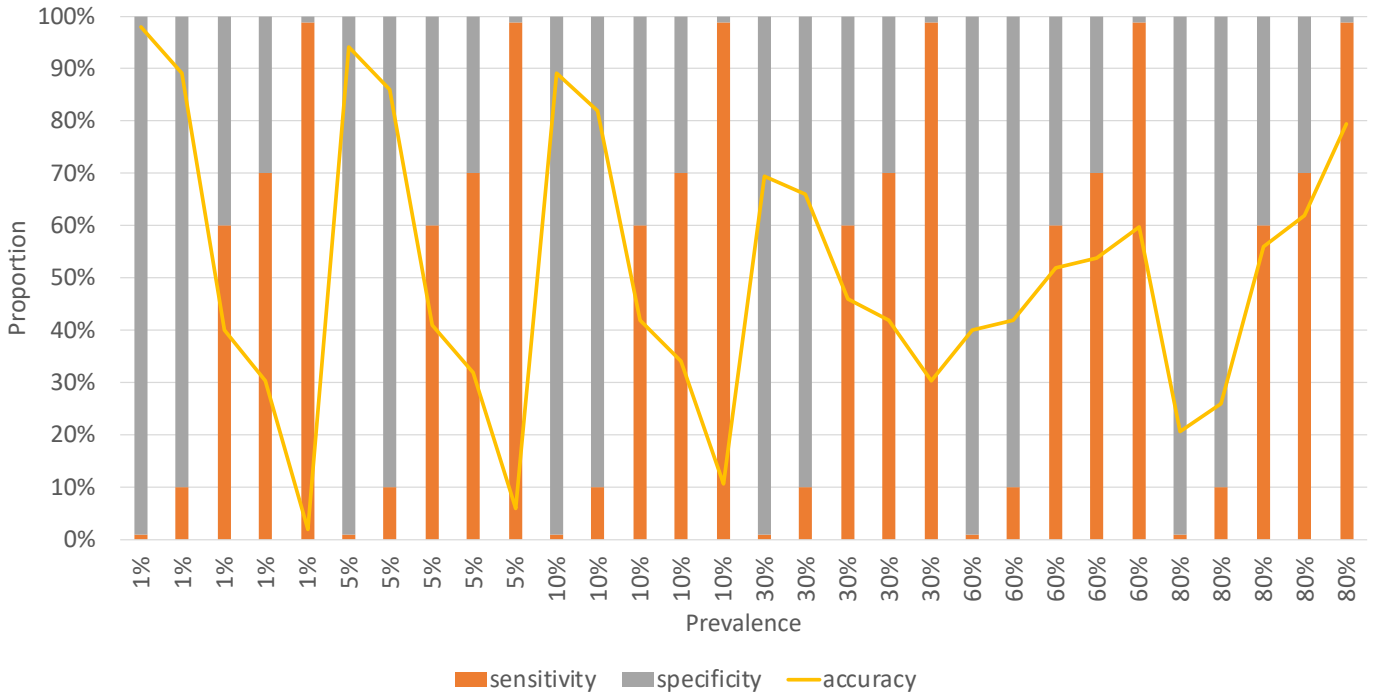
Table 1.6 presents some measures that are used to assess accuracy of a single test and its performance at a single threshold. Single tests can be judged on the basis of the paired test accuracy, expressed as sensitivity and specificity, positive predictive value and negative predictive value, negative likelihood ratio (LRN) and positive likelihood ratio (LRP). In addition to the measures described in the table, other measures include the *Youden index* [149, 150] which is derived from (sensitivity + specificity -1) and *accuracy* [151], which is described as the proportion correctly classified and calculated as $\frac{TP+TN}{TP+TN+FP+FN}$ where TP, TN, FP and FN are respectively true positive, true negative, false positive, false negative. The error rate, a weighted average of classification errors in people with disease and those without the disease is given by [148]:

$$\text{error rate} = (P(T-|D+) \times P(D+)) + (P(T+|D-) \times P(D-)) \tag{1.2}$$

For a test that gives results on a continuous scale, each of these measures is calculated for a single threshold. In contrast, the area under ROC curve (AUROC) and diagnostic odds ratio (DOR) are described in Section 1.1.2.6.

Figure 1.6 shows that for a hypothetical index test in a low prevalence setting (1%) when sensitivity is low (1%), and specificity is high (99%), accuracy is high (98%); when sensitivity is high (99%), and specificity is low (1%), accuracy is low (2%). Conversely in a high prevalence setting (80%) when sensitivity is low (1%), and specificity is high (99%), accuracy is modestly low (21%); when sensitivity is high (99%), and specificity is low (1%), accuracy is modestly high (79%).

Figure 1.6: Variation of accuracy, sensitivity and specificity with disease prevalence



The figure plots the accuracy ($\frac{TP+TN}{TP+TN+FP+FN}$) of a hypothetical index test and shows how accuracy, sensitivity and specificity vary with disease prevalence.

Table 1.6: Paired measures to assess test accuracy [148]

Measure	**Definition**	**Notation** ∗
Sensitivity	Probability of an abnormal test in people with disease	P(T+|D+)
Specificity	Probability of a normal test in people without disease	P(T-|D-)
PPV†	Probability of disease given abnormal test	P(D+|T+)
NPV ‡	Probability of non-disease § given normal test	P(D-|T-)
LRP ‖	Ratio of probability of abnormal test in diseased to abnormal tests in non-diseased	$\frac{P(T+\|D+)}{P(T+\|D-}$
LRN ¶	Ratio of probability of normal test in diseased to normal tests in non-diseased	$\frac{P(T-\|D+)}{P(T-\|D-}$

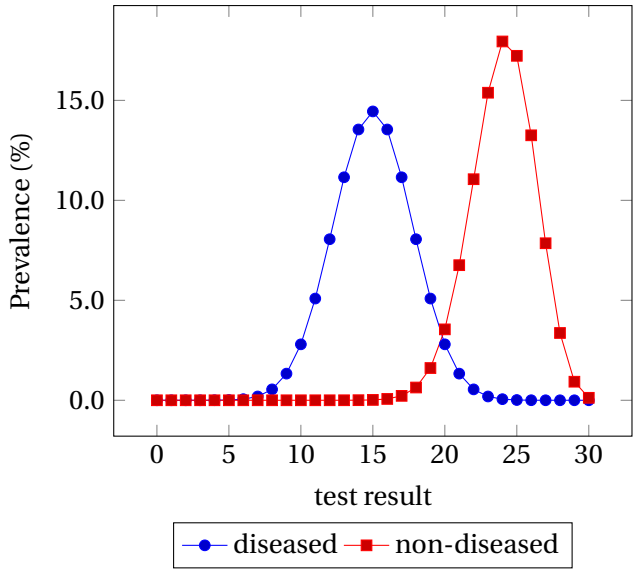
∗ T+ test positive T- test negative D+ disease positive D- disease negative
†*PPV* positive predictive value ‡*NPV* negative predictive value
§the term *non-disease* is used because this only applies to one specific disease and does not imply health
‖*LRP* positive likelihood ratio ¶*LRN* negative likelihood ratio

Test accuracy is influenced by the prevalence of disease in the tested population, as well as the threshold for determining disease. Sensitivity and specificity, and measures derived from them, are often said to be independent of prevalence [152], and PPV and NPV are said to be dependent on prevalence [152]. Sensitivity is derived from $\frac{TP}{TP+FN}$ and specificity from $\frac{TN}{TN+FP}$. If D represents the set of people that have disease and $D\prime$ the set of people who do not have disease[7], then prevalence is given by $\frac{D}{D+D\prime}$. Continuing with this notation, the denominators for sensitivity and specificity are respectively $\frac{}{D}$ and $\frac{}{D\prime}$; so that the ratio between D and $D\prime$ is not a component of the derivation of these indices. In contrast PPV and NPV are respectively derived from $\frac{TP}{TP+FP}$ and $\frac{TN}{TN+FN}$. Using the same notation the denominator for PPV and NPV is respectively $\frac{}{D+D\prime}$ and $\frac{}{D\prime+D}$; so that the ratio between D and $D\prime$ has an impact on these measures.

Notwithstanding the mathematical independence of sensitivity and specificity from the ratio of D and $D\prime$, these measures of accuracy still remain dependent on prevalence to the extent that prevalence is a function of the threshold that is used to define disease. In different populations the spectrum and prevalence of disease vary, which has an impact on the PPV and NPV, along with sensitivity and specificity [148, 153, 154]. Test accuracy varies with prevalence but not in a consistent pattern: sometimes the sensitivity will remain constant and the specificity will fall, other-times the sensitivity will decrease and the specificity will increase [148]; other patterns may be observed. Therefore measures of test accuracy are best understood as a function of the prevalence of disease and the threshold for test positivity, with a trend for lower specificity to be associated with higher disease prevalence [154] and lower sensitivity to be associated with lower disease prevalence [148].

[7] To facilitate interpretation of the denominators, which have + signs, D and $D\prime$ are used rather than D+ and D-.

Figure 1.7: Distribution of test results between diseased and non-diseased



1.1.2.4 Thresholds in diagnostic testing

Clinical scenarios often encompass a broad range of possible presentations, manifesting in overlap between health and disease. For any given test there are usually a range of potential results and there is a spectrum of normal to abnormal. For example, the normal (non-diseased) range for many blood tests is derived from the normal (Gaussian) distribution, implying by definition that the most extreme 5% of the normal population will have an abnormal result if the normal population is defined as the middle 95%. Even for findings which may be considered to be dichotomous, such as the presence or absence of a cardiovascular murmur on auscultation, intra and interobserver variability imply that the abnormal range encompasses values or results that in some contexts might be judged as being normal by some observers [153, 155]. Therefore, when assessing the accuracy of a single test, particularly those with continuous results, it is often necessary to determine a threshold for normality, with values outside of this range being considered abnormal. For example, the threshold on the mini mental state examination [156], a brief test of cognitive function with scores from 0 (all wrong) to 30 (no errors; better cognition), is usually set at 24 with scores of 23 and below being an abnormal result [157]. The threshold for abnormality has consequences for test accuracy, with more extreme thresholds encompassing fewer people who are healthy and a more extreme set of people with disease [150]. Figure 1.7 shows a hypothetical distribution of a set of results on the Mini Mental state examination, if scores of 23 and below are taken to indicate abnormal then most of the abnormal results are in people who have disease.

For measures which describe the accuracy at a single threshold, such as those described in table 1.6 it is important to consider the paired measures, rather than single items, because they are related to one another and to the threshold for defining abnormal. Reconsider figure 1.7. If instead of 23 and below indicating abnormal the threshold were scores of 14 and below then only people with the

most extreme test results would be identified which may be more likely to identify those with more severe disease. Conversely the lower threshold of 14 instead of 23 would mean fewer people with non-disease would be incorrectly identified as having a problem. The impact on the test accuracy characteristics results in lower sensitivity, higher specificity; higher PPV, lower NPV. Conversely, if the threshold were taken to be scores of 28 and below, then a larger number of people with non-disease would have an abnormal test result, the impact on the test accuracy characteristics would be higher sensitivity, lower specificity; lower PPV, higher NPV. The threshold at which a test is judged as being abnormal depends on the clinical context, especially the setting (high or low prevalence) and purpose of the test, whether the test being used to rule in, rule out, triage, or refine a set of people who may have a particular condition [148]. The clinical consequences of test results also influence the desired test characteristics, for instance when seeking to identify blood borne viruses in blood donated for transfusions it would be important to use a test with very high sensitivity, because the consequences of not identifying a blood borne virus could be grave, whereas when selecting people for aggressive surgery or chemotherapy for cancer it would be important to minimise the number of false positive diagnoses because of the significant harms of unnecessary treatment or diagnostic labelling.

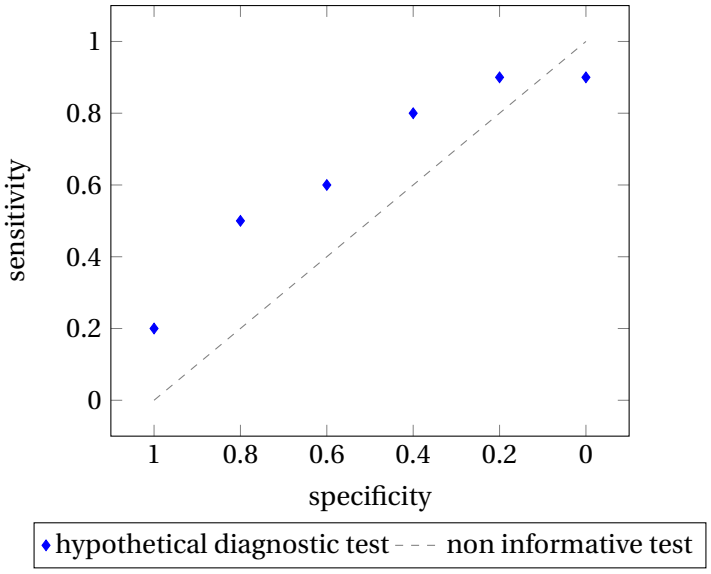
1.1.2.5 Discrimination and calibration

In clinical practice it is common to combine single tests to form a model which can be used to aid diagnosis or prognosis. For example, as described in Section 1.1.1.6, a model has been constructed to predict the risk of future dementia. To judge the value of a model, it is important to evaluate explained variation, discrimination, and calibration [158]. Explained variation can be quantified using the R^2 statistic or Briar score, which are respectively logarithmic and quadratic scoring rules to quantify the differences in observed and predicted outcomes [158].

Discrimination is a measure of the extent to which people with the target condition have higher risk predictions than those without the target condition [158]. Discrimination can be quantified using area under ROC curve (AUROC, Section 1.1.2.6), the discrimination slope (calculated as absolute difference in average prediction for those with and without the target condition), or visualised with box-plots or histograms [158]. For a binary outcome AUROC is equivalent [158] to the C statistic [159]. For survival analysis with censored data, for example with prediction models, quantifying discrimination with Uno's C or Royston's D may be less biased than AUROC. Uno's C is a modified C statistic that can be applied to censored data [160]. Royston's D statistic is an alternative measure of model discrimination which can be interpreted as *"the log hazard ratio resulting from dichotomizing a continuous prognostic index and fitting a Cox model with a single dummy variable to distinguish the groups ... a difference of at least 0.1 may be needed to see an important difference in the survival curves"* [161].

Calibration is a measure of how well model-predicted probability of a target condition approximates the observed values. Calibration can be quantified using the Hosmer-Lemeshow goodness-of-fit test [158] or by a plot of observed against predicted probabilities [158, 162]. When the intercept

Figure 1.8: Example ROC



of the calibration plot is close to zero, a slope of 1 can be interpreted as indicating good calibration [163].

1.1.2.6 Accuracy of tests over range of thresholds

For expressing the accuracy of a test over a range of thresholds a useful and commonly reported measure is the area under ROC curve (AUROC) [164]. Figure 1.8 is an example ROC plot where the a hypothetical test sensitivity is plotted against specificity, at six different thresholds. The plot of sensitivity against specificity is in line with recommendations [150], but often the plot is sensitivity against 1-specificity. A non-informative test is also plotted, which runs from (1,0) to (0,1), for this test the distributions of test results in people with and without disease coincide [150]. The interpretation of the AUROC is that it is the probability that for a randomly selected pair of diseased and non-diseased people in the study the result of the diseased person would be more abnormal than the result of the non-diseased person [165], and is equivalent to the Mann-Whitney-Wilcoxon test statistic [166]. AUROC is given by $\int_0^1 S_p S_e$, that is the integral of specificity with respect to sensitivity. An AUROC of 0.5 is uninformative and an AUROC of 1.0 is perfect, small changes in the AUROC may be associated with clinically important changes in the diagnostic accuracy. In contrast, the DOR is defined as the ratio of odds of positivity in subjects with disease relative to the odds in subjects without disease:

$$\frac{P(T+|D+) \times (P(T-|D-)}{P(T+|D-) \times P(T-|D+)} \tag{1.3}$$

The diagnostic odds ratio (DOR) can be used to express the accuracy of a test over a range of thresholds [165], or as a measure of test accuracy at a single threshold.

In clinical practice there can often be a preference to either identify disease correctly (TP, i.e. sensitivity) or identify those who do not have disease (TN i.e. specificity) and the relative importance of these two extremes depends on the role of the test and the setting. A disadvantage of both AUROC and DOR is that as single measures of test accuracy it can be difficult to understand how a particular result relates to the needs of the clinical scenario and whether the test is better at identifying people with (high sensitivity) or without (high specificity) disease. When using the AUROC only certain thresholds may be clinically appropriate so some investigators have advocated the use of partial AUROC for clinically relevant thresholds only, but this approach has the disadvantage that values just outside the interval are discarded [166]. The AUROC does not allow interpretation of the relative clinical consequences of FP and FN and indeed the mis-classification costs vary along the curve because they are dictated by its shape, which has been described as incoherent [166]. Prevalence also varies along the ROC curve as the gradient changes, which can be problematic when using the AUROC to compare tests [165].

Net benefit measures [167–169] are an alternative approach that summarise diagnostic performance in a single measure at a clinically relevant threshold [165]. The weighted comparison net benefit measure [169] weights differences in the sensitivity and specificity between two tests by the disease prevalence and the relative clinical (mis-classification) costs[8], which can be converted into a net benefit of x TP cases per n patients [165]. The weighted comparison parameter is given by:

$$WC = \delta sensitivity + [(\frac{1-prevalence}{prevalence}) \times \text{relative cost}(\frac{FP}{TP}) \times \delta specificity] \tag{1.4}$$

For example [165] consider two tests for the diagnosis of dementia AD8 sensitivity 0.91 specificity 0.91 [170] and IQCODE sensitivity 0.81 specificity 0.96 [171]. The difference in sensitivity is 0.1 and the difference in specificity is -0.05. In a population with estimated disease prevalence 13% and a 30 fold higher weighting for TP compared to FP the WC is given by:

$$WC = 0.1 + [\frac{0.87}{0.13} \times (0.3) \times -0.05] \tag{1.5}$$

Therefore WC = -0.0004; the negative value of the parameter indicates that test 2 (IQCODE) would be marginally preferable to AD8, whereas a positive value would indicate that AD8 was preferable to IQCODE. The weighted comparison can be framed as an increase in the TP per 100,000, if all the benefit is for TPs, by calculating WC × prevalence × 100,000; for the worked example this means 5.2 *fewer* TP people per 100,000 people would be identified using AD8 compared to IQCODE (-4 × 10^{-4} × 13% × 100,000 = -5.2).

1.1.2.7 Accuracy of tests in clinical context

Some authors have questioned the value of calculating the diagnostic accuracy of a single test in isolation, because in clinical practice it is common to use more than one test [172]. In practice

[8] not economic costs

a clinician typically weighs a number of possible alternative diagnoses for a clinical scenario. In contrast, in research the accuracy of tests is commonly evaluated with respect to a single diagnosis, termed the target condition. Criteria for the target condition should be specified explicitly because there is often more than one way to define the disease of interest, for instance a clinical syndrome, a particular investigation result, or post mortem.

One approach to address the limitations of evaluating the accuracy of single tests is to use a logistic regression equation, which can be shown to be equivalent to Bayes theorem [148], to analyse the accuracy of the combination of several tests. Alternative approaches to the analysis of tests in combination are outlined in Table 1.7 and include: simple tree building, classification and regression tree software, logistic regression analysis, manipulated logistic regression, neural networks, latent class analysis, Bayesian networks. Logistic regression is the most commonly used approach in the literature but may appear to imply the test is symmetrical, that is equally good for ruling in or ruling out disease, though this is rarely the case. Methods to overcome this limitation by transforming the data before analysis have been described [148, 173].

Table 1.7: Approaches to analysing combinations of tests [148]

Approach	**Advantages**	**Drawbacks**
Simple tree building	Straightforward Accounts for interactions More than two outcomes possible Thresholds can be chosen	Impossible with many tests Continuous tests must be categorised Excludes missing values No measure of imprecision Analysis not pre specified
Classification and regression tree (CART)	Sequence can be manipulated Interactions accounted for More than two outcomes possible Can handle missing data	Complex software needed May need large sample Variable results
Logistic regression	Continuous results stay continuous Interactions accounted for Usually only one outcome Multiple imputation possible	
Manipulated logistic regression	Tests are entered in a specified order Can be more clinically relevant	Variable results (subjective)
Neural networks	More than two outcomes possible Interactions and diagnostic asymmetry handled	Missing data problematic Difficult to interpret clinically
Latent class analysis	Helpful when no reference standard	Can be difficult to interpret Number of variables limited

1.1.2.8 Impact of tests on patient outcomes

Arguably, diagnosis is only useful or relevant to the extent that it addresses the patients questions about their illness [174]. The most important questions relating to diagnosis might often be questions about prognosis: what is likely to happen to the patient and what can be done to influence that [174, 175]. Determining outcomes of diagnostic processes that are relevant to patients and their kin may be one approach to reduce over-diagnosis. Overdiagnosis is the term given to the phenomenon where people undergoing medical evaluation are labelled as having pathology without evidence that the identification or treatment of this results in improved outcomes. Beyond reducing overdiagnosis, another advantage of the prognostic approach is that tests may be incorporated in modelling risk in the context of patient outcomes on a spectrum, rather than a dichotomous expression of disease or not-disease [175]. Finally, to the extent that prognosis (outcome) provides a more accurate classification about the impact and harm-balanced merits of interventions than diagnosis (disease state) a prognostic approach to clinical decision making may facilitate more individualised and stratified healthcare [175].

Test accuracy studies are often cross sectional in nature [148] and do not have extended follow-up. A disadvantage of the prognostic approach to clinical decision making and practice is the need to understand the important outcomes that are meaningful to a range of patients and their kin, and the requirement to follow people up for an extended period to ascertain these.

1.1.3 Clinical judgement

1.1.3.1 Reliability and observation

In clinical practice, many test results are subject to a degree of judgement. For instance, there is reported to be inter-observer variability about the interpretation of an Magnetic Resonance Image (MRI) of the brain in people with vascular dementia, that can be improved with the use of operational definitions by experienced readers [176]. Variability in clinical judgement on the result of a test may result in implicit, unmeasurable thresholds for tests, for instance the extent of micro-vascular infarcts that are normal or pathological [148].

Kappa is a measure of reliability [177], which is often regarded as a measure of agreement [178] for two or more observers of a categorical outcome, in contrast to weighted kappa and intraclass correlation coefficient which may respectively be applied to ordinal and continuous outcomes. These methods are less suitable for assessing diagnostic accuracy than the measures discussed in Section 1.1.2.3 because when evaluating diagnostic accuracy the reference standard is taken to represent truth, whereas in evaluation of reliability and agreement multiple observers are typically regarded as being equally likely to identify the true observation. To the extent that high observer variability increases random error in the test result the test accuracy may be adversely affected, but tests with high variability may still be useful for individual clinicians [179]. Observer bias may also influence the scoring of test results, either from subjective prior knowledge of the person under evaluation

which influences the implicit threshold, or from a clinician prejudging one test preferably to another and (perhaps subconsciously) making more effort to obtain what they perceive as an accurate result [148].

1.1.3.2 Diagnosis as categorisation

The diagnostic process can be scrutinised as a probabilistic, analytic process with a clear diagnostic category as the intended outcome. However, in clinical practice decisions are typically dichotomous [175], and reasoning is often non-analytic [180, 181], based on intuition which is derived from recognition [180]. Clinical judgement, also known as instinct or gestalt [182], is the holistic judgement about the diagnostic category. Understood in this framework, diagnosis is typically a system 1 cognitive process.

Table 1.8 compares the two processes that are described in the theory of the dual mode of cognition. System 1 is prone to a range of biases, such as neglecting ambiguity, ignoring absent evidence, overestimating low probabilities and using heuristics [181]. Interventions, especially guided reflection, can help to modify clinician behaviour and reduce diagnostic error [183]. In contrast, clinicians might also use a type 2 process in a restricted rule out to evaluate the probability of a limited set of diagnoses that are rare but critical not to miss. Decisions about the management of a

Table 1.8: Type 1 and Type 2 cognitive processes [181]

System 1	**System 2**
Unconscious	Conscious
Mostly emotional and involuntary	Mostly voluntary and unemotional
Implicit	Explicit
Low effort, high capacity, rapid	High effort, low capacity, slow

clinical scenario are based on judgements about the trade-offs of different options weighed against an implicit and perhaps subconscious threshold probability [143]. Decisions about treatment are simply conceptualised as being based on whether the patient exceeds the threshold probability for treatment (T), which is given by the ratio of the costs (C) and benefit (B) in the formula [184]

$$T = \frac{C}{B+C} \tag{1.6}$$

In general practice, many treatments, especially for acute problems such as sore throat, are prescription based and have relatively low cost and high benefit, so clinicians may have a low threshold probability for treatment.

In contrast, where an intervention has tangible risks for patients, such as neurosurgery, or requires long term treatment for chronic illness, such as anti-coagulation for atrial fibrillation, clinicians are more likely to have a higher threshold probability for treatment compared to acute minor illness. In

this situation, the risks of the investigation (i) directly influence the threshold for investigation (T_t), as determined by the formula [143]:

$$T_t = \frac{(P(FP) \times C) + i}{(P(FP) \times C) + (P(TP) \times B)} \tag{1.7}$$

therefore the threshold probability for investigation rather than excluding the diagnosis based on clinical features alone will fall as the risks of investigation fall and the accuracy of the test increases.

In contrast the threshold for treating rather than investigating, the test-treatment threshold [T_{ttrx}) is given by the formula [143]:

$$T_{ttrx} = \frac{(P(TN) \times C) - i}{(P(TN) \times C) + (P(FN) \times B)} \tag{1.8}$$

therefore the threshold for treating rather than investigating will tend to rise as the risks of investigation fall and the accuracy of the test increases.

If diagnosis represents a method of applied categorisation, it follows that clinicians will apply a classification framework that is fit for purpose in their clinical setting. In general practice, it is often difficult to place a person with a cluster of symptoms into a definite category, at least initially, and it can be easier to identify the categories that they do not fit in to. In one study, over 50% of cases in general practice could not be assigned to an definite initial diagnosis [142]. Consequently, GPs may assign people to categories based on the intended action (observe, providing "safety net" advice on the anticipated course of illness; investigate; treat; refer) rather than the specific diagnosis [142]. These categories reflect an implicit judgement based on the fit of the pattern of illness with illness scripts [185] that have been developed through the integration of knowledge and experience [186].

1.1.3.3 Low incidence serious disease

Clinicians often use a variety of heuristics as part of their diagnostic reasoning [180]. The availability heuristic is the concept that some diagnoses are more readily accessible to the cognitive processes of the clinician, and therefore more likely to be selected, than others. Clinicians encounter common conditions frequently, and so have many exemplars encompassing a diverse range of clinical phenotypes with which to compare a new case with, whereas less commonly encountered conditions are conceptualised using the textbook description which may be a poor fit for the patients presentation [187]. Some rare conditions may be readily available to clinicians for consideration if the cognitive representation is especially vivid such as if a condition has serious consequences, or is associated with a particularly significant past encounter such as a complaint or death. Alternatively clinicians may experience temporary increases in the cognitive availability of rare diagnosis if there is some reason for the condition to be more in their mind, such as if they have recently received education.

In 2018 the average number of patients aged over 60 years that were registered with each GP practice was 3,691 [188]. With a meta-analytic incidence of dementia in this age group of 17 per 1000 person years [14] each GP practice could expect to see approximately 64 new cases of dementia per year, that is around 21 per GP in a three-GP practice. In contrast the average GP working 4 days a

week has perhaps 6,300 consultations per year based on 35 contacts per day [189]. Therefore, in general practice, dementia is a low incidence, serious disease. The serious consequences of dementia, such as incontinence, immobility, and dependency in activities of daily living typically develop over years, whereas meningitis, which is also a rare but serious condition, is rapidly fatal. The impact of the varying time course is that while dementia is a serious condition, the stakes of not identifying it at any one encounter are relatively low, in contrast to meningitis which has high stakes consequences in the short term if missed and so is likely to be relatively cognitively available even though it is a rare disease [190]. Consequently many clinicians are likely to experience relatively low cognitive availability for their personal dementia exemplar: if the diagnosis of dementia is delayed by a week, or even a month, the consequences for the doctor of the delayed diagnosis are relatively minimal. Clinicians are likely to have particularly poor cognitive availability for dementia that is not due to Alzheimer's disease because the clinical exemplar probably includes memory loss [191, 192].

General practitioners can have important issues with dealing with low incidence serious disease [193]. When evaluating a patient with symptoms that could represent serious illness, but where the incidence in general practice is low, there is an inherent difficulty in identifying all the people who are ill without also unnecessarily intervening and raising anxiety in the much larger number of people whose symptoms, though similar, are attributable to a non-serious aetiology [193]. At a population level it may cause more harm to investigate many people to identify the one person with disease [194]. GP's clinical judgement can accurately identify people with serious illness [195], including cancer [196]. However, in general practice there will often be uncertainty in managing people with symptoms that indicate a non-zero risk of a serious pathology. GPs use a range of approaches to manage this uncertainty to refine the diagnosis (Section 1.1.2.1), including interrogating their "gut feeling" about a scenario, applying diagnostic algorithms explicitly or implicitly, arranging investigations and safety netting [193]. Clinician gut feeling is a sense that something is wrong, which can be central to diagnostic discrimination [193]. GPs may also consider their gut feeling when formulating a decision about prognosis and treatment, even if there is currently no specific diagnostic label [189].

1.2 Scientific justification for the work

Very little empirical work has investigated the accuracy of tests for diagnosing dementia in general practice. Even fewer studies have explored how accurate the judgement of general practitioners is compared to arguably more objective measures such as brief cognitive tests. Because test accuracy can vary between settings it is important to investigate the accuracy of tests in a general practice setting, where the prevalence of dementia in people with cognitive symptoms can be anticipated to be lower than in a group of people that have been referred to a specialist memory clinic. Demographic trends suggest that the workload for health systems associated with the diagnosis of dementia will increase, and the political agenda appears to be inclining towards a desire to explore an enhanced role for general practice in the diagnosis of cognitive disorders. Additional empirical

research, based in general practice, to understand the accuracy of tests for the diagnosis of dementia and the acceptability to patients of a GP based diagnosis is therefore important to inform future policy developments.

1.3 Aims and objectives of the thesis

The aim of this thesis is to investigate approaches to diagnosing dementia syndrome in general practice. The objectives are

- to determine the prevalence of dementia and mild cognitive impairment in a group of people presenting with symptoms of dementia to their general practitioner
- to measure the accuracy of a range of tests for diagnosing dementia syndrome in symptomatic people
- to identify if a combination of tests can have a high positive predictive value for identifying people with dementia and normal cognition
- to investigate the acceptability to patients of GPs diagnosing dementia

.

Chapter Summary and thesis outline

This Chapter has presented an introduction to important topics for the thesis. Chapter 2 gives the introduction, methods, results, and discussion of an empirical systematic review of the test accuracy of general practitioners clinical judgement for the diagnosis of dementia. Chapter 3 presents details of the methods and Chapter 4 the results of an empirical diagnostic test accuracy study investigating the accuracy of a range of tests for diagnosing dementia in general practice. Chapter 5 presents the rationale, methods, results and discussion of an empirical qualitative investigation to explore the acceptability to patients of GPs diagnosing dementia. Chapter 6 reviews the empirical findings in the context of the existing literature, discusses the strengths and limitations and considers the implications for research and clinical practice.

CHAPTER 2

SYSTEMATIC REVIEW

This Chapter gives details of the systematic review of accuracy of clinical judgement of primary care physicians for the diagnosis of dementia and cognitive impairment. Text from this Chapter has been published as a Cochrane Systematic review protocol [197]. The results and discussion were not published or submitted for review prior to submission of the thesis. Section one outlines the objectives and Section two the background. Section three describes the method, detailing the types of studies, participants, index test, target condition, reference standard, searches, selection of studies, assessment of risk of bias, and analysis. Section four describes the results, including the results of the search, a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [198], the characteristics of the included studies and their methodological quality using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [199], the findings relating to diagnostic accuracy and heterogeneity, and the sensitivity analyses. Section five summarises the main findings, discusses the strengths and limitations of the included studies and the review process, relates the findings to the literature, and identifies conclusions. A PRISMA-DTA checklist is provided in Appendix L.

2.1 Background

Details on dementia, diagnostic accuracy, and clinical judgement are provided in Chapter 1. Doctors use a variety of processes to make a diagnosis and decision, including non-analytical reasoning processes such as pattern recognition. Some people with dementia unfortunately have sufficiently advanced disease at the time of presentation that no formal evaluation is required to make the diagnosis of dementia syndrome and indeed GPs often report using their clinical judgement to make a diagnosis of dementia, rather than a standard instrument [200, 201].

2.2 Objective of the systematic review

To determine the accuracy of general practitioners' overall clinical judgement for diagnosing cognitive impairment and dementia in people presenting to primary care.

2.3 Method of the systematic review

2.3.1 Type of study

Both cross-sectional studies and cohort studies were included in the review. Cross-sectional studies are potentially at higher risk of incorporation bias than cohort studies and this was accounted for when assessing studies for risk of bias. The risk of incorporation bias may be higher in cross sectional studies than cohort studies because index tests and reference standards may be done by the same examiner, or the results of one examination may be communicated to the patient who may in turn communicate them to the other examiner. Conversely cross sectional studies are at lower risk of bias due to participant flow causing partial verification of the reference standard. The approach of excluding cross-sectional studies was judged to be too restrictive.

Case-control studies were excluded because they are at high risk of bias. Furthermore, by definition, any participants would have been recruited on the basis of disease state (dementia, cognitive impairment or normal) which would typically prevent GPs from making a blinded clinical judgement about the diagnosis.

2.3.2 Participants

Studies were only included if they recruited participants from primary care. Primary care was defined as first-contact health care provided by a non-specialist clinician in a continuing-care office setting. Studies were excluded if a consultation with a non-specialist occurred in hospital (including outpatients or emergency departments) because this was judged as unlikely to represent primary care in the sense that was relevant to the review objective.

Studies were included if GPs made a clinical judgement about the presence of cognitive impairment or dementia. Studies were excluded if GPs were required to make a judgement about the presence of cognitive impairment or dementia in all people attending primary care, regardless of age as this was judged to be akin to screening. To avoid heterogeneity in studies by professional role, included articles exclusively had GPs as the primary care provider, rather than other allied professions (e.g. advanced nurse practitioners, physician assistant etc.), because the training requirements for these different roles vary.

2.3.3 Index test

Clinical judgement was defined as being unaided by any additional test, investigation or inquiry beyond that which is immediately available to the clinician [202–204]. The review investigated a

single index test (clinical judgement) for two target conditions (cognitive impairment composite or dementia) (see Section 2.3.4). In everyday practice a clinical judgement is formed by a GP after an encounter with a patient, in which the GP typically accesses and reviews the medical record as part of their consultation. In contrast, in diagnostic accuracy research in general practice there are potentially three ways that clinical judgement may be defined.

Firstly, there may be a documented diagnosis of cognitive impairment or dementia in the medical records, but this is likely to reflect the process of documentation rather than clinical judgement (*documented* approach). In one study the number of patients registered as having dementia on GP surgery records increased by 9% when diagnostic coding was systematically audited [108]. Documented diagnoses of dementia in the medical record may not reflect the clinical judgement of a GP about the presence or absence of the target condition, and indeed the increase in the prevalence with consistent coding suggests that documentation may be systematically under-coded in the electronic medical record in routine practice.

Secondly, clinical judgement may be defined as an opinion based on knowledge of the patient and review of the medical notes, but not relating to a specific encounter with the patient (*retrospective* approach). However, this approach is likely to be affected by differences in consulting behaviour of people with cognitive impairment or dementia compared to those without these problems. An electronic record of a dementia diagnosis (which as indicated is likely to be systematically under-recorded compared to actual diagnoses [108]) has been associated with an increase in primary care attendances both around the time of recording on the electronic record [205] and in the five years preceding this [206]; in contrast other investigators have reported that people with established moderate-severe dementia have *fewer* consultations per year with a GP compared to those with no or mild dementia [207]. Because cognitive status is associated with consultation behaviour it may also be associated with the implicit (i.e. not formally recorded) knowledge that a GP has about an individual when forming a retrospective clinical judgement.

Thirdly, clinical judgement in diagnostic accuracy research may be defined as the impression formed by a clinician after consulting with a patient who has presented to a specific encounter with the doctor (*prospective* approach). The patient may be consulting with symptoms suggestive of possible dementia, but not necessarily, because it may be the consulting GP, or indeed a third party, who raises the possibility of cognitive problems. This was considered as the definition of clinical judgement that is most relevant to practice for this review. This definition of clinical judgement was judged to be most applicable to clinical practice, being contemporaneous with a specific encounter, and least likely to be subject to systematic bias in coding in the medical record.

Studies were included that used the third *prospective* definition. To avoid an empty review, studies that used the second *retrospective* definition were also included so long as the GP's determination about cognitive status had taken place before any definitive diagnosis. Studies that used the *documented* approach were excluded as this was considered to not reflect clinical judgement.

2.3.4 Target condition

There were two target conditions for the review: all-cause dementia, and cognitive impairment due to dementia or Mild Cognitive Impairment (MCI) [3].

Experience in clinical general practice is that when there are concerns about impaired cognition these are focused primarily on the possibility of dementia rather than MCI, but inevitably some people who are evaluated for possible dementia will be diagnosed with MCI. The second target condition includes both dementia and MCI because it would be unusual for a GP to diagnose MCI, especially on the basis of clinical judgement alone, because neuropsychological evaluation is often required. If clinical judgement was sensitive for any cognitive impairment then if the GP assessed the person as being cognitively normal it would be likely that the person had neither dementia nor MCI.

Studies were excluded if they investigated the accuracy of clinical judgement for risk prediction of future dementia. There was no restriction to a particular stage or clinical severity of dementia.

2.3.5 Reference Standard

2.3.5.1 Dementia

Studies were included if they used a definition from DSM-III-R, DSM-IV-TR, or ICD-10. Studies were also included if they used Automated Geriatric Examination for Computer Assisted Taxonomy (AGECAT) [208], Cambridge Mental Disorders of the Elderly Examination (CAMDEX) [209], CDR [73] or structured interview for the diagnosis of dementia of the Alzheimer type, multi-infarct dementia and dementias of other aetiology according to ICD-10 and DSM-III-R (SIDAM) [210] because these are well validated methods of applying the aforementioned diagnostic criteria. Studies were included if they used expert specialist clinical judgement as the reference standard; a specialist was defined as a clinician with particular expertise in diagnosing and managing dementia, practising in a hospital or secondary care environment, with the professional status of geriatrician, psychiatrist or neurologist. Studies were eligible for inclusion if they used longitudinal confirmation of the diagnosis of all-cause dementia in primary care, because it was anticipated that in some studies a specialist assessment would only be offered to some participants. Longitudinal confirmation of the diagnosis in primary care was operationalised as case record review occurring at least three months after the index test diagnosis of dementia where no other alternative diagnosis is identified. While recognising that many people who could be correctly diagnosed as having dementia by unaided clinical judgement (true positives) would have a fairly advanced stage of disease, stage of disease did not form part of the target condition definition.

Although the target condition was all-cause dementia, studies that used an aetiological sub-type definition were also eligible, these were: for Alzheimer disease the NINCDS-ADRDA criteria [62, 63]; for vascular dementia the NINDS-AIREN criteria [65]; for Lewy body dementia the Dementia with Lewy Body Consortium criteria [70, 71]; and for frontotemporal dementia the consensus criteria [66].

2.3.5.2 Cognitive impairment

Cognitive impairment was a composite target condition encompassing dementia (as defined above) and MCI. Any recognised definition of MCI was eligible e.g. the original or revised criteria of Petersen [93, 211], Winblad [94], or the NINCDS-ADRDA criteria [63]. Whilst acknowledging that causes of cognitive impairment extend beyond dementia and MCI (e.g. head injury, delirium, and neoplasm, Section 1.1.1.5) these were not part of the target condition for the review. Therefore, if (for example) the index tests indicated cognitive impairment or dementia, and further evaluation demonstrated that the clinical problem was neoplasm instead, the test would be false positive.

2.3.6 Searches

The search strategy was constructed in discussion and collaboration with Anna Noel-Storr, Information Specialist for the Cochrane Dementia and Cognitive Improvement Group. The search strategy and search report is provided in Appendix A.

The following electronic databases were searched by Anna Noel-Storr from inception to 29 April 2019: MEDLINE (OvidSP); Embase (OvidSP); Web of Science Core Collection, including the Science Citation Index and the Conference Proceedings Citation Index (Thomson Reuters Web of Science); PsycINFO (OvidSP), and LILACS (BIREME). Where appropriate controlled vocabulary was used, such as MeSH terms (in MEDLINE) and EMTREE (in Embase) and other controlled vocabulary in other databases. The reference lists of all included papers was also reviewed for additional studies. Search filters designed to retrieve diagnostic test accuracy studies were not used because available filters have not yet proved sensitive enough for systematic review searches [212]. No language restrictions were applied to the electronic searches.

2.3.7 Selection of studies

Two investigators independently screened the retrieved citations at the title and abstract stage, using Covidence software [213] to classify each citation as relevant, possibly relevant, or not relevant. Conflicts in classification were resolved by discussion until consensus was reached. Full text articles were obtained for all citations classified as either possibly relevant or relevant. Two investigators independently reviewed the full text articles using Covidence and made a final judgement about the relevance of the citation. All conflicts in classification were resolved by discussion. Articles that were excluded at the full text stage were given a reason for exclusion using the following hierarchy:

1. Inappropriate participants: not primary care
2. Inappropriate reference standard
3. Inappropriate index standard
    - Not GP

- Not clinical judgement

4. Inappropriate target condition

5. Inappropriate study design (i.e. not a diagnostic test accuracy study e.g. a study reporting qualitative data, descriptive epidemiology, randomised trial or survey)

Articles were classified under the highest order reason for exclusion, so that if a study was not set in primary care then it was excluded at level one, whereas if it was set in primary care but did not use an appropriate index test then it was excluded at level three. Only one reason was needed for a study to be excluded. Authors of included studies were contacted by email when necessary to obtain paired data on sensitivity and specificity if these were not clearly reported in the original article.

2.3.8 Assessment of risk of bias

Review Manager [214] was used to extract data to assess the risk of bias for each study, which was appraised separately by two investigators using the QUADAS-2 checklist [199]. Any disagreements were resolved by discussion.

2.3.9 Statistical analysis and data synthesis

Paired data on sensitivity and specificity were extracted from included studies separately by two investigators into Review Manager. Discrepancies were resolved by discussion informed by further review of the original manuscript until consensus was reached. Studies that did not report paired sensitivity and specificity were described but not included in the meta-analysis. Data were used to calculate the accuracy of the index test for diagnosing the two target conditions: cognitive impairment (including both MCI and all-cause dementia), and all-cause dementia. Diagnostic accuracy was calculated with 95% confidence intervals separately for each target condition, in all studies with available data.

Meta-analyses was performed on pairs of sensitivity and specificity in Stata version 13 [215] using `metandi` and `xtmelogit`. Since `metandi` does not allow analysis with fewer than four studies, and `xtmelogit` does not produce output parameters for plotting a summary ROC curve, summary ROC curves were generated using MetaDTA: Diagnostic Test Accuracy Meta-Analysis v1.45 [216]. The main meta-analysis was done in the studies at lowest risk of bias (one or fewer QUADAS-2 domains at high risk of bias). The bivariate random-effects model approach was used to estimate a summary point for sensitivity and specificity, assuming this is appropriate to do, and the hierarchical SROC (HSROC) model was used to estimate a summary ROC curve [150, 217–219]. The bivariate model and the HSROC model are mathematically equivalent when no covariates are fitted but differ in their parameterisation to (respectively) estimate an average sensitivity and specificity (bivariate approach) and average ROC curve (HSROC approach) [150, 219]. Because the bivariate approach estimates a single point along the HSROC curve it does not fully reflect the heterogeneity in the data and may

be of limited value. In contrast using the HSROC approach to plot a summary ROC curve allows investigators to observe where study points are plotted, and how close they lie in relation to the summary curve, which allows a clear depiction of heterogeneity [150]. Different all-cause dementia diagnostic criteria were analysed together. Meta-analyses by aetiological subtype of dementia was not done, because it is unlikely that GPs would make an aetiological subtype diagnosis. The analysis using the bivariate approach to estimate a summary point assumes that there is a constant threshold for the diagnosis of the target condition according to clinical judgement across studies; Section 2.5.2 discusses this in detail.

2.3.10 Investigations of heterogeneity

Three sources of heterogeneity were designated in advance of the analyses as being important to investigate: the definition used to define the reference standard (ICD-10, DSM-III-R or DSM-IV-TR); whether the clinical judgement of the GP was based on a prospective opinion (having just seen the patient) or a retrospective opinion (being asked to consider the case at some arbitrary date); and whether the GP had access to use medical records when giving their opinion. A further investigation of heterogeneity was defined when the characteristics of the included studies were known. The post-hoc analysis investigated heterogeneity by risk of bias in the flow and timing QUADAS-2 domain (high risk or not). Heterogeneity was initially investigated through visual examination of paired diagnostic accuracy data in forest plots, and the ROC plot of the raw data. These components of study design were extracted as covariates and added to the analytical model with likelihood ratio tests being used to compare model fit [150].

The length of vocational training programme for GP participants were specifically not examined as sources of heterogeneity because these were anticipated in advance to be poorly reported (or not reported) in original studies. Furthermore, it is recommended to only consider possible sources of heterogeneity which vary at the study level [220], which is unlikely to be the case for these domains.

Caution is advised when interpreting investigations of heterogeneity, especially exploratory post-hoc analyses [220].

2.3.11 Sensitivity analyses

A pre-specified sensitivity analysis investigated how the estimates of diagnostic accuracy were modified when including studies that were judged to be at high risk of bias in more than two QUADAS-2 domains.

2.3.12 Assessment of reporting bias

Quantitative methods for exploring reporting bias are not well established for studies of DTA [220] and so this was not examined.

2.4 Results

2.4.1 Results of the search

Figure 2.1 shows that the search yielded 12,681 citations and 8,118 remained after de-duplication. Review of the full text for 56 records led to 16 being included, referring to 10 studies, of which 9 were included in the meta-analysis. One study [221] was included in the review but not in the meta-analysis because it was not possible to obtain paired data on diagnostic accuracy either from the original paper or after correspondence with the authors. The study is included for transparent reporting because based on the study design paired diagnostic accuracy data should be available.

Of the 40 papers that were excluded at the full text review stage, nine were excluded because they were not based in primary care. Instead, these papers were set in the community [9, 191, 222–224], outpatients department [225–227] or hospital [228]. Seven papers were excluded because they used an ineligible reference standard; instead of using a reference standard from the pre-specified list (Section 2.3.5), the excluded papers used methods which would be at high risk of incorrectly classifying the target condition. Of the seven papers excluded because of an ineligible reference standard, four used cognitive tests as the reference standard, of which three [229–231] used the MMSE [156] and one [232] used the Blessed Dementia Scale, [233] two [234, 235] did not have dementia as a target condition and used a screening test for MCI which was not further detailed [234] or the World Health Organisation Composite International Diagnostic Interview (CIDI) [235], and one aimed to validate a new measure and cross-validated the new tool against other cognitive tests but not a diagnosis [236].

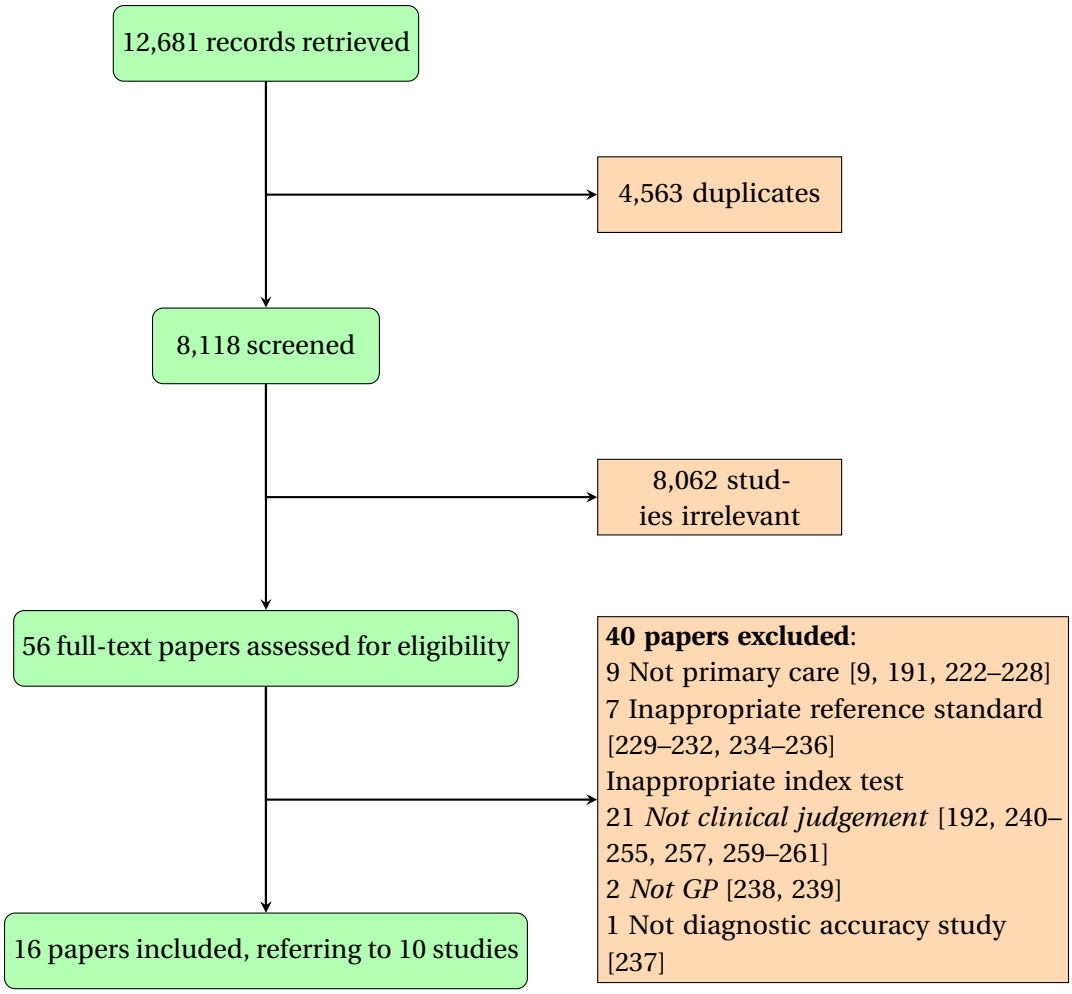
One study, which was a letter, was excluded because it was not a diagnostic test accuracy study [237].

Of the 23 papers that were excluded because they investigated an ineligible index test, two papers were not investigating the accuracy of GPs [238, 239]. Of the remaining 21 papers which were excluded because of an ineligible index test, five [192, 240–243] investigated the documentation of a diagnosis in the medical record which was specified in advance as an ineligible index test (Section 2.3.3). Of the remaining 16 papers that had an ineligible index test, eight papers [244–251] referring to one study were excluded because GPs were asked to use the Dutch dementia guidelines to make a diagnosis rather than their unaided clinical judgement, three papers [252–254] were excluded because GPs were asked to use the MMSE [156], one study [255] used the ADMP[1] scale [256], one study [257] used the 6CIT [258], one study [259] used the CDR [75], one study [260] used a list of warning signs of dementia from the Alzheimer's Association, and one study [261] used the standardized physicians' manual issued by the Ministry of Health, Labor and Welfare of Japan.

Two previous systematic reviews were identified [262, 263]. Table 2.1 compares the papers that were included in the two previous reviews with the current review and indicates that there were four studies that were included in all three reviews [221, 264–266]; two studies that were included only the

[1] ADMP is not an abbreviation

Figure 2.1: PRISMA flowchart



current review [267, 268]; four studies that were included in both the current review and Mitchell et al., [269–273]; two studies that were included in both Mitchell et al., and Van den Dungen et al., but not in the current review [192, 243]; and 12 studies that were only included by Mitchell et al., [191, 226, 227, 232, 247, 274–280].

Of the 14 studies that were included in previous reviews but not in the current review, 12 were excluded from the current review because the index test did not meet the inclusion criteria (details in Table 2.1), and two were excluded because the reference standard did not meet the inclusion criteria [232, 276]. Of the 12 studies that were excluded from the current review because the index test did not meet the inclusion criteria one study [247] required GPs to use the dutch dementia guidelines to make a diagnosis of dementia and the other 11 studies defined clinical judgement as a documented diagnosis of dementia in the medical record, neither of which met the criteria for clinical judgement as defined in the current review.

Table 2.2 summarises the characteristics of included studies regarding sampling, index test,

reference standard, participant flow, target condition, definition, and access to medical records. Table 2.3 provides a further overview of the participant selection, the characteristics of participants, verification of the target condition, and the proportion classified by GP judgement and the reference standard. The characteristics of Hawaii [221] are described but because there was no usable 2x2 data on diagnostic accuracy this study was not included in the meta analysis. Seven studies sampled consecutive consulting patients or the entire registered GP list (Cambridge City, Mannheim, Sydney, Zwolle, Hawaii, Antwerp, France) and three took a sample of patients from GP surgery lists (Cambridge, Amsterdam, AgeCoDe). Four studies (Cambridge City, Cambridge, Amsterdam, Zwolle) used a retrospective diagnosis, and six (Mannheim, Sydney, Hawaii, Antwerp, AgeCoDe, France) used a prospective opinion. Six studies (Cambridge City, Cambridge, Mannheim, Amsterdam, Zwolle, Antwerp) used CAMDEX as the reference standard, one used the Canberra interview for the Elderly (Sydney) or the SIDAM (AgeCoDe) and two (Hawaii, France) used expert opinion. Five studies (Cambridge City, Cambridge, Amsterdam, Hawaii, AgeCoDe) appeared to have complete verification of the index test, with the reference standard being administered to all participants, but only two of these (Cambridge and AgeCoDe) did not screen people in some way prior to the index test (see Table 2.2); the other studies had some form of partial verification. All studies investigated dementia as a target condition and four studies (Mannheim, Sydney, Zwolle, AgeCoDe) additionally investigated MCI as a target condition. Five studies (Cambridge City, Cambridge, Mannheim, Amsterdam, Sydney) used ICD-10 as the definition, two studies (Antwerp, AgeCoDe) used DSM-IV-TR and one study used each of DSM-III-R (Zwolle), CDR (Hawaii), and NINCDS-ADRDA (France). Access to the medical records was not available in two studies (Cambridge, Mannheim), was unclear in one (Cambridge City) and was available in the remainder.

Table 2.1: Comparison of included citations in systematic reviews

Citation	Index test	Reference standard	Mitchell 2011 [262]	Van den Dungen 2012 [263]	Current review	Study ID in review	Why excluded from this review ∗
Boise 2004 [274]	Medical records	CERAD †	Yes	No	No	-	Index test
Borson 2006 [191]	Medical records	CERAD	Yes	No	No	-	Index test
Boustani 2005 [275]	Medical records	CERAD	Yes	No	No	-	Index test
Bowers 1990 [276]	Prospective questionnaire	MMSE ‡	Yes	No	No	-	Reference standard
Brayne 1990 [267]	Retrospective rating	CAMDEX‖	No	No	Yes	Cambridge	-
Callahan 1995 [277]	Medical records	SPMSQ ¶	Yes	No	No	-	Reference standard; Index test
Chodosh 2004 [278]	Medical records	TDQ §	Yes	No	No	-	Reference standard; Index test
Cooper 1992 [281]	Prospective rating	CAMDEX	Yes	Yes	Yes	Mannheim	-
De Lepeleire 2004 [268]	Prospective rating	DSM-IV-TR	No	No	Yes	Antwerp	-
Eefsting 1996 [265]	Retrospective rating	CAMDEX	Yes	Yes	Yes	Zwolle	-
Ganguli 2004 [279]	Medical records	MMSE	Yes	No	No	-	Reference standard; Index test
Iliffe 1990 [280]	Medical records	MMSE	Yes	No	No	-	Reference standard; Index test
Jacinto 2009 [226]	Medical records	Expert consensus	Yes	No	No	-	Reference standard; Index test
Kaduszkiewicz 2010 [269]	Prospective rating	MCI [94]	Yes	No	Yes	AgeCoDe	-
Löppönen 2003 [192]	Medical records	DSM-IV-TR	Yes	Yes	No	-	Index test
Mant 1988 [232]	Doctors opinion	MMSE	Yes	No	No	-	Reference standard
O'Connor 1988 [270]	Retrospective rating	CAMDEX	Yes	No	Yes	Cambridge	-
Ollafsdottir 2000 [243]	Medical records	DSM-III-R	Yes	Yes	No	-	Index test
Pentzek 2009 [271]	Prospective rating	DSM-IV-TR	Yes	No	Yes	AgeCoDe	-
Pond 1994 [266]	Prospective rating	DSM-III-R	Yes	Yes	Yes	Sydney	-
Rondeau 2008 [272]	Retrospective rating	DSM-IV-TR	Yes	No	Yes	France	-
Valcour [221]	Prospective rating	CDR	Yes	Yes	Yes	Hawaii	-
Van Hout 2000 [247]	Prospective, applying Dutch guidelines	CAMDEX	Yes	No	No	-	Index test
Wilkins 2007 [227]	Medical records	CERAD	Yes	No	No	-	Index test
Wind 1994 [273]	Retrospective rating	CAMDEX	Yes	No	Yes	Amsterdam	-

∗ Design aspect which did not meet criteria for current review † Consortium to Establish a Registry for Alzheimer's Disease protocol (CERAD) [282]
‡ *MMSE* Mini Mental State Examination [156] ‖ *CAMDEX* Cambridge Mental Disorders of the Elderly Examination [209]
¶ *SPMSQ* Short Portable Mental Status Questionnaire [283] § *SPMSQ* Telephone Dementia Questionnaire [284]
DSM-IV-TR Diagnostic and Statistical Manual of Mental Disorders (4th ed., Text Revision) *MCI* Mild Cognitive Impairment
DSM-III-R Diagnostic and Statistical Manual of Mental Disorders (3rd ed., Revised) *CDR* Clinical Dementia Rating scale

Table 2.2: Characteristics of included studies

Citation *Country*	**Sampling**	**Index test**	**Reference standard**	**Flow**	**Target condition**	**Definition**	**Notes available**
Cambridge City [270] *UK*	All people aged 75 years and over from six GP lists in Cambridge and *1 in 3* ‡ from a 7th surgery	Retrospective	CAMDEX ∗	Timing not specified Partial verification: MMSE§ score <24: All score 24 or 25: *1 in 3*‡	Dementia	ICD-10 [5] †	Unclear
Cambridge [267] *UK*	Randomly selected women aged 70-74 years, and all women aged 75-79 years from rural GP surgery list	Retrospective	CAMDEX	Timing not specified Verification complete	Dementia	ICD-10	No
Mannheim [281] *Germany*	Consulting patients over 65 years seen in 24 GP surgeries over four weeks	Prospective	CAMDEX	Timing not specified Partial verification: random sample stratified by GP opinion	MCI Dementia	ICD-10 ∗∗	No
Amsterdam [273] *Netherlands*	Age stratified random sample from 30 GP surgery lists	Retrospective	CAMDEX	Timing not specified Partial verification: MMSE score <22: all score >22: random sample	Dementia	ICD-10	Yes
Sydney [266] *Australia*	Consulting patients in a retirement complex seen by GPs over four weeks	Prospective	Canberra Interview for the Elderly	Timing not specified Partial verification: 50% random sample	MCI Dementia	ICD-10 ∗∗	Yes
Zwolle [265] *Netherlands*	All patients aged 65 years and over on lists of eight GPs	Retrospective	CAMDEX	Timing not specified Partial verification: MMSE score <18: All score 18-23 random 2 in 3 score >23 random 1 in 3 score >27 none	MCI Dementia	DSM-III-R ∗∗	Yes

Citation *Country*	**Sampling**	**Index test**	**Reference standard**	**Flow**	**Target condition**	**Definition**	**Notes available**
Hawaii [221] *USA*	Consecutive patients aged 65 years or more at one GP surgery over six weeks	Prospective	Expert	Timing not specified Flow unclear	Dementia	CDR	Yes
Antwerp [268] *Belgium*	Consecutive patients aged 65 years or more with possible dementia were sought	Prospective	CAMDEX	Tests within one month Partial verification: IADL ‖score 4: all seen score <4 none seen	Dementia	DSM-IV-TR	Yes
AgeCoDe [269] *Germany*	Random sample of people aged 75-89 years registered with GP & postal invitation to participate	Prospective	SIDAM	Reference test 1.5 years & 3 years after index test Complete verification	MCI Dementia	Winblad [94] DSM-IV-TR	Yes
France [272] *France*	Consecutive patients in a trial to train GPs. Only control patients included.	Prospective	Expert	Timing not specified Partial verification: 222 of 375 diagnosed "dementia" by GP 38 of 711 diagnosed "not dementia" by GP 125 of 311 diagnosed "unsure" by GP	Dementia	NINCDS-ADRDA [63]	Yes

∗ *CAMDEX* Cambridge Mental Disorders of the Elderly Examination † *ICD-10* The International Classification of Diseases, Tenth Revision
‡ further detail not reported § *MMSE* Mini Mental State Examination (MMSE) *CDR* Clinical Dementia Rating scale (CDR)
‖IADL Lawton instrumental activities of daily living scale ∗∗ Same reference standard applies to both target conditions
Cambridge [267] included the minimal cases on CAMDEX in with mild cases. minimal dementia on CAMDEX approximates MCI
SIDAM structured interview for the diagnosis of dementia of the Alzheimer type, multi-infarct dementia and dementias
of other aetiology according to ICD-10 and DSM-III-R *UK* United Kingdom *USA* United States of America

Table 2.3: Comparison of diagnostic accuracy studies

Design aspect	**Mannheim [281]**	**Sydney [266]**	**Hawaii [221]**	**Antwerp [268]**	**AgeCoDe [269]**	**France [267]**	**Cambridge 1990 [267]**	**Zwolle [265]**	**Amsterdam [273]**	**Cambridge city [270]**
Index Test ‡	P	P	P	P	P	P	B	B	B	B
Participant selection										
Series ∗	C	C	C	C	R	C	R	C	R	C
Symptomatic	No	No †	No	No	No	Yes	No	No	No	No
Characteristics of participants										
Number (index test)	3721	200	303	1003	3242	1453	365	2536	475	444
Mean age (years)	76	83	75	75	80	81	-	73 §	75	-
% Female	70	86	63	63	66	71	100	-	62	-
% with dementia ∗	29	31	9	2	2	50	8	19	10	56
Verification with reference standard										
Verified N	407	105	303	10	2294	385	365	375	475	444
Verified (%) ∗	(11)	(53)	(100)	(1)	(70)	(26)	(100)	(15)	(100)	(100)
GP judgement (%)										
Not impaired	36	-	33	-	94	48	90	90	76	45
CIND	41	-	-	-	-	-	7	8	62	20
Dementia	23	27	33	-	6	26	3	2	50	18
Uncertain	-	-	33	-	-	22	-	-	-	17
Diagnostic accuracy of GP judgement										
Sensitivity (%)	91	42	-	100	51	73	34	39	52	58
Specificity (%)	76	89	-	100	96	58	94	99	94	78

∗ C consecutive R random D dementia (see Table 2.2 for definitions)

Symptomatic: symptoms *required* for participation

% verified = $\frac{\text{number underwent reference test}}{\text{number underwent index test}}$ % with dementia = $\frac{\text{number with dementia}}{\text{number verified}}$

† participants were not presenting with symptoms but GPs were asked to maximise the inclusion of people with suspected dementia

-: not reported § median ‡ P prospective B retrospective

2.4.2 Funding of included studies

Cambridge City was funded by the Charles Wolfson Charitable Trust. **Cambridge** was funded by the Medical Research Council and the Mental Health Foundation. **Mannheim** was funded by the Federal Ministry of Science and Technology, Bonn. **Amsterdam** was funded by grants from The Netherlands Health Research Promotion Programme and The Netherlands Foundation of Mental Health. **Sydney** was funded by the National Health and Medical Research Council. **Zwolle** was funded by National Fund for Mental Health, the Protestant Association for the Care of Chronic Patients, the Protestant Foundation for the Care of Chronic Patients and the Foundation for Health Care Research. **Hawaii** was funded by HMSA Foundation, Honolulu, Hawaii; and the John A. Hartford Center of Excellence in Geriatric Medicine, University of Hawaii. **Antwerp** was funded by Pfizer. **AgeCoDe** was funded by the German Federal Ministry of Education and Research. **France** was funded by Eisai and Pfizer.

2.4.3 Methodological quality of included studies

QUADAS-2 was used to judge the risk of bias in each study. Figure 2.2 shows that overall there was low risk of bias in the included studies. Specifically most studies were at low risk of bias in the patient selection domain, with the remainder being at unclear risk of bias [2]; and this domain had concern about applicability. All studies were at low risk of bias in the index test domain and most studies had low concern about applicability. Most studies were at low risk of bias in the reference standard domain and all had low concern about applicability, though two were at high risk of bias. Flow and timing was the domain where there was most risk of bias, Section 2.5.2.1 discusses this in detail.

Figure 2.2: Summary of methodological quality in included studies

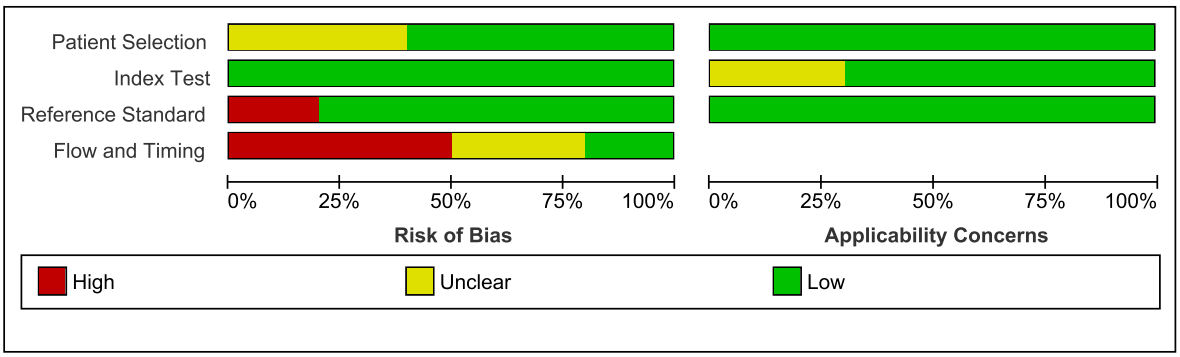


Figure 2.3 presents the summary risk of bias and applicability assessment for individual studies. Four studies were at unclear risk of bias in the patient selection domain. AgeCoDe [269] required potential participants to have had least one GP contact in the past 12 months and excluded people who were housebound, which may have resulted in the systematic exclusion of people who lived

[2] QUADAS-2 categorises risk of bias as high risk, low risk, or unclear risk, and applicability as high concern, low concern, or unclear concern

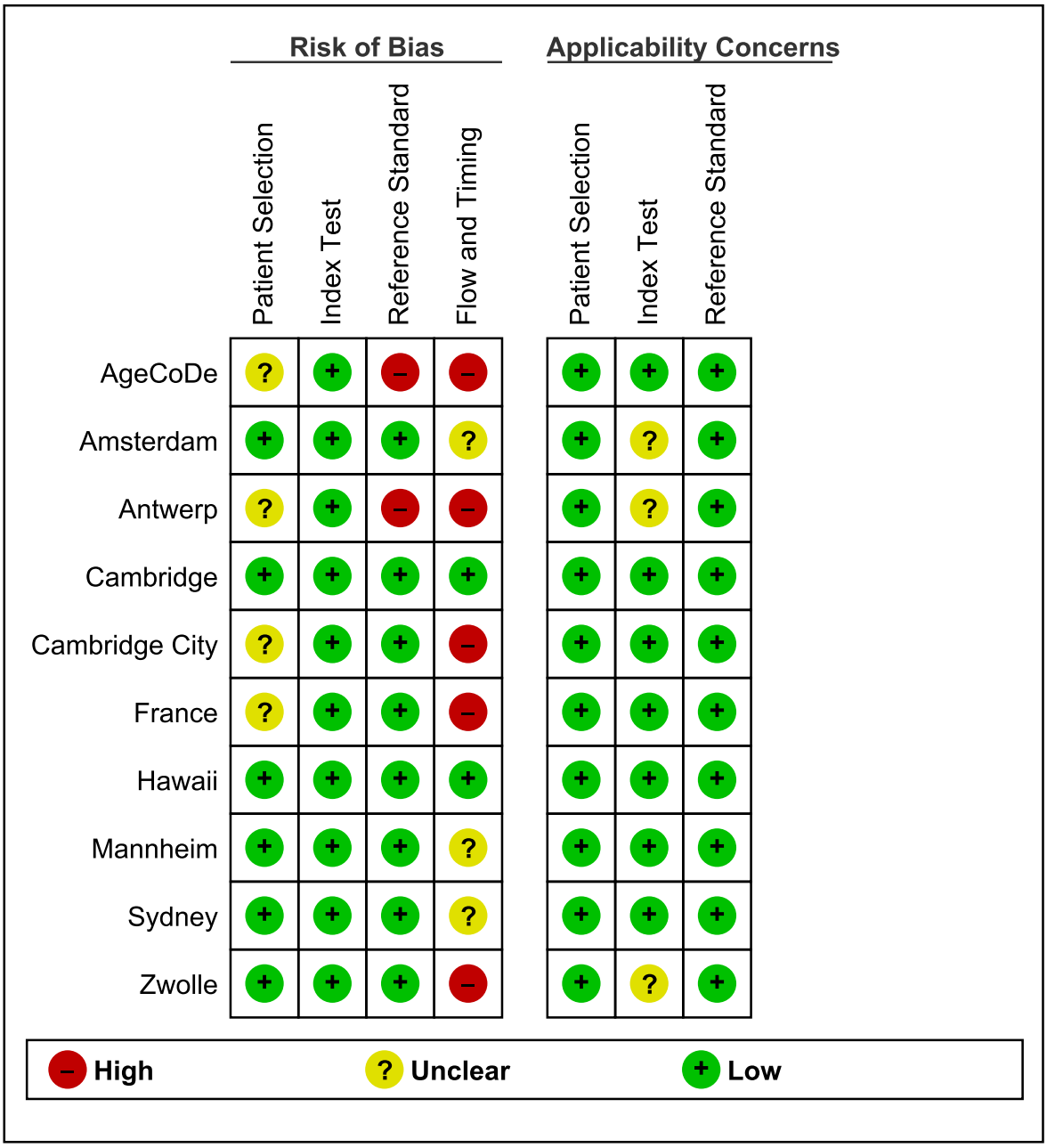
alone or did not seek help, or with the most severe impairment. Similarly Antwerp [268] specifically excluded people who lived in a residential home for the elderly (regardless of whether home visits were needed) and Cambridge City [270] excluded people who lived in long stay hospitals, and also those who were diagnosed as having "minimal dementia", which was not defined. France [272] stated that consecutive patients were enrolled, but it was not clear that these patients were genuinely consecutive because five patients per GP were recruited over two years, which appears low. There were no concerns about the applicability of patient selection.

All studies were judged to be at low risk of bias in the index test domain. However, three studies had unclear applicability of the index test. Amsterdam [273] and Zwolle [265] explained the reference standard criteria (but not the diagnosis) to participating GPs before they made their judgement. Antwerp [268] asked GPs to give their opinion after they had asked four questions on a four-item test of instrumental activities of daily living [285], though did not specifically state the GPs had to use the results of this instrument.

Two studies were judged to be at high risk of bias in the reference standard domain. Both AgeCoDe [269] and Antwerp [268] were judged to be at high risk of bias in this domain because they incorporated the index test into the reference standard. No studies were at unclear risk of bias in the reference standard domain and there were no concerns about the applicability of the reference standard.

Five studies were judged to be at high risk of bias in the flow and timing domain. AgeCoDe [269] applied the reference standard at 1.5 years and 3 years after the index test assessment. Cambridge City [270] did not provide any information on timing, but was at risk of partial verification because the reference test CAMDEX was used to evaluate people who scored 23 or less on the MMSE, together with sample of those who scored 24 or 25, but none of those who scored over 25. Zwolle [265] followed a similar procedure, performing the reference standard on all participants scoring 17 or below on the MMSE, together with a random 2/3 sample of those scoring 18-23, a random 1/3 sample of those scoring between 24-27 and none of those scoring 28 and above. France [272] was also at high risk of bias in this domain because of partial verification: 222 of 375 (59%) people diagnosed with dementia by GP were seen by a specialist, in contrast 38 of of 711 (5%) people not diagnosed with dementia by GP were seen by a specialist, and 125 of 311 (40%) people with an uncertain GP diagnosis were seen by a specialist. Antwerp [268] was at high risk of bias in this domain because it appeared that the reference standard was only done on 10 people of 1003 who were evaluated by a GP. Mannheim [281] was at unclear risk of bias because there was partial verification but a stratified random sample was taken for verification, containing equal numbers in each of the four categories of GP assigned impairment. Similarly, Sydney [266] took a random sample for verification by the reference standard and was at unclear risk of bias. Amsterdam [273] was at unclear risk of bias because although there appeared to be full verification, this was derived from people who had been screened with the MMSE prior to the index test, including a sample of those who scored up to 30 (see Table 2.2), and there was no information on timing.

Figure 2.3: QUADAS-2 summary for each study



Two studies, AgeCoDe [269] and Antwerp [268] were judged to be at high risk of bias in two QUADAS-2 domains (reference standard, flow and timing).

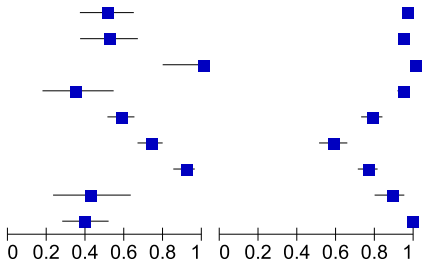
2.4.4 Findings

2.4.4.1 Target condition: Dementia

Two studies reported the prevalence of dementia, Zwolle [265] reported a prevalence of 7% and Cambridge City [270] reported a prevalence of 11%. In both cases the reported prevalence takes account of weighting in the sampling for the reference test and so differs from the raw calculation of $\frac{TP+FN}{TP+FN+FP+TN}$.

Figure 2.4: Forest plot of GPs clinical judgement for diagnosis of dementia

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
AgeCoDe	28	92	27	2147	0.51 [0.37, 0.65]	0.96 [0.95, 0.97]
Amsterdam	25	25	23	402	0.52 [0.37, 0.67]	0.94 [0.91, 0.96]
Antwerp	16	0	0	987	1.00 [0.79, 1.00]	1.00 [1.00, 1.00]
Cambridge	10	19	19	311	0.34 [0.18, 0.54]	0.94 [0.91, 0.96]
Cambridge City	121	51	87	185	0.58 [0.51, 0.65]	0.78 [0.73, 0.83]
France	143	79	52	111	0.73 [0.67, 0.79]	0.58 [0.51, 0.66]
Mannheim	107	69	10	221	0.91 [0.85, 0.96]	0.76 [0.71, 0.81]
Sydney	11	9	15	70	0.42 [0.23, 0.63]	0.89 [0.79, 0.95]
Zwolle	28	4	43	300	0.39 [0.28, 0.52]	0.99 [0.97, 1.00]

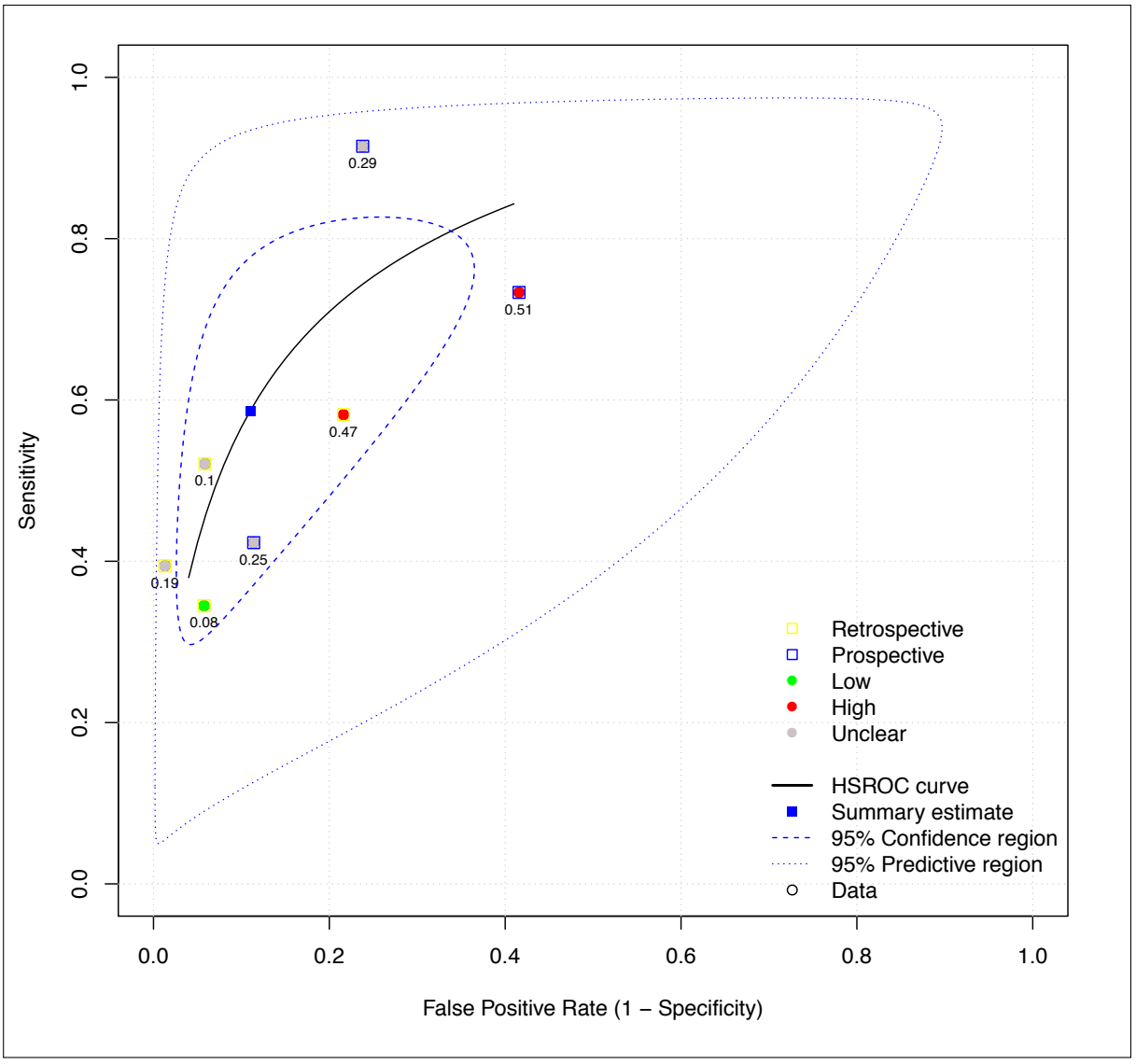


TP true positive **FP** false positive **FN** false negative **TN** true negative **95% CI** 95% confidence interval

Figure 2.4 is a forest plot of the accuracy of clinical judgement of GPs for the diagnosis of dementia in the nine studies that had paired data on sensitivity and specificity. Excluding Antwerp, which reported a sensitivity and specificity of 100% but was also one of two studies that were at high risk of bias in two QUADAS-2 domains, in individual studies sensitivity ranged from 34% (95% CI 18% to 54%) in Cambridge to 91% (95% CI 85% to 96%) in Mannheim, which was one of only two studies (the other being France) that reported higher sensitivity than specificity. Specificity was generally higher than sensitivity and ranged from 58% (95% CI 51% to 66%) in France to 99% (95% CI 97% to 100%) in Zwolle.

Figure 2.5 is a summary plot of the accuracy of clinical judgement of GPs for the diagnosis of dementia. Recall that as described in Section 2.3.9 and Section 2.3.11 the main analysis was restricted to studies at lowest risk of bias, and therefore AgeCoDe [269] and Antwerp [268] were excluded from the main meta-analysis. The individual study points are shown with the calculated prevalence ($\frac{\mathrm{TP+FN}}{\mathrm{n}}$), together with an indication of whether the index test was prospective (blue square surround) or retrospective (yellow circle surround), and are coloured to indicate the risk of bias in the flow and timing domain (green low risk of bias, grey unclear risk of bias, red high risk of bias). The blue single filled dot indicates the summary point estimate of diagnostic accuracy, and is displayed on a summary ROC curve. The dashed bubble indicates the 95% confidence interval around the summary point and the larger dotted bubble indicates the 95% prediction region. The 95% confidence interval indicates the region where in 95 samples out of 100 the calculated 95% confidence interval will contain the true mean value, based on the included data [286]. The 95% prediction region indicates

Figure 2.5: Summary plot of GPs clinical judgement for diagnosis of dementia



Colours of dots: indicate risk of bias in flow and timing domain, all are at unclear risk of bias in this domain
Surround indicates prospective retrospective ; decimals indicate prevalence of target condition

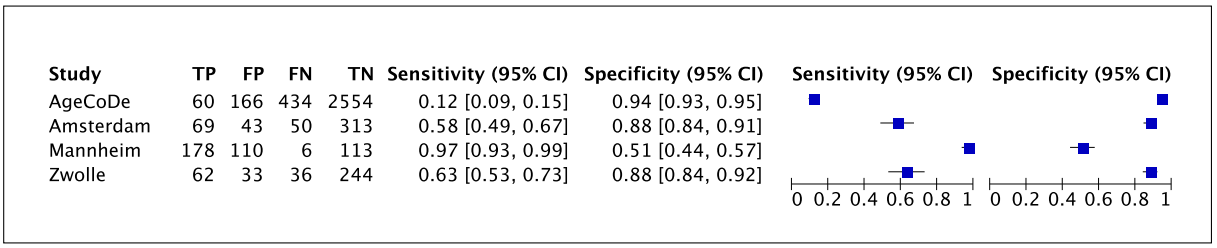
the area where the results of a future study could be expected to lie, based on the analysed data [286].

In the meta analysis for dementia as the target condition, the summary diagnostic accuracy of clinical judgement of general practitioners was sensitivity 59% (95% CI 41% to 74%), specificity 89% (95% CI 77% to 95%), diagnostic odds ratio (DOR) 11 (95% CI 6 to 22), positive likelihood ratio 5.3 (95% CI 2.7 to 10.7) and negative likelihood ratio 0.46 (95% CI 0.33 to 0.66). As shown in Figure 2.5 the summary point is an average of the studies in the meta-analysis. The single studies with the most comparable diagnostic accuracy to the summary point on either sensitivity or specificity had *lower* values for the other measure: the single study with a sensitivity closest to the summary point (see

Figure 2.4) was Cambridge City (sensitivity 58% specificity 78%) and the single study with a specificity closest to the summary point was Sydney (specificity 89%, sensitivity 42%). Furthermore, only four of the seven studies in the meta-analysis were included in the 95% confidence region displayed on the summary ROC curve, with one lying just outside (Zwolle), and two lying further outside (Mannheim, France). Therefore the summary point in isolation is over-simplistic as a representation of the diagnostic accuracy of clinical judgement for dementia and does not reflect the heterogeneity in the data.

2.4.4.2 Target condition: Cognitive impairment

Figure 2.6: Forest plot of GPs clinical judgement for diagnosis of cognitive impairment



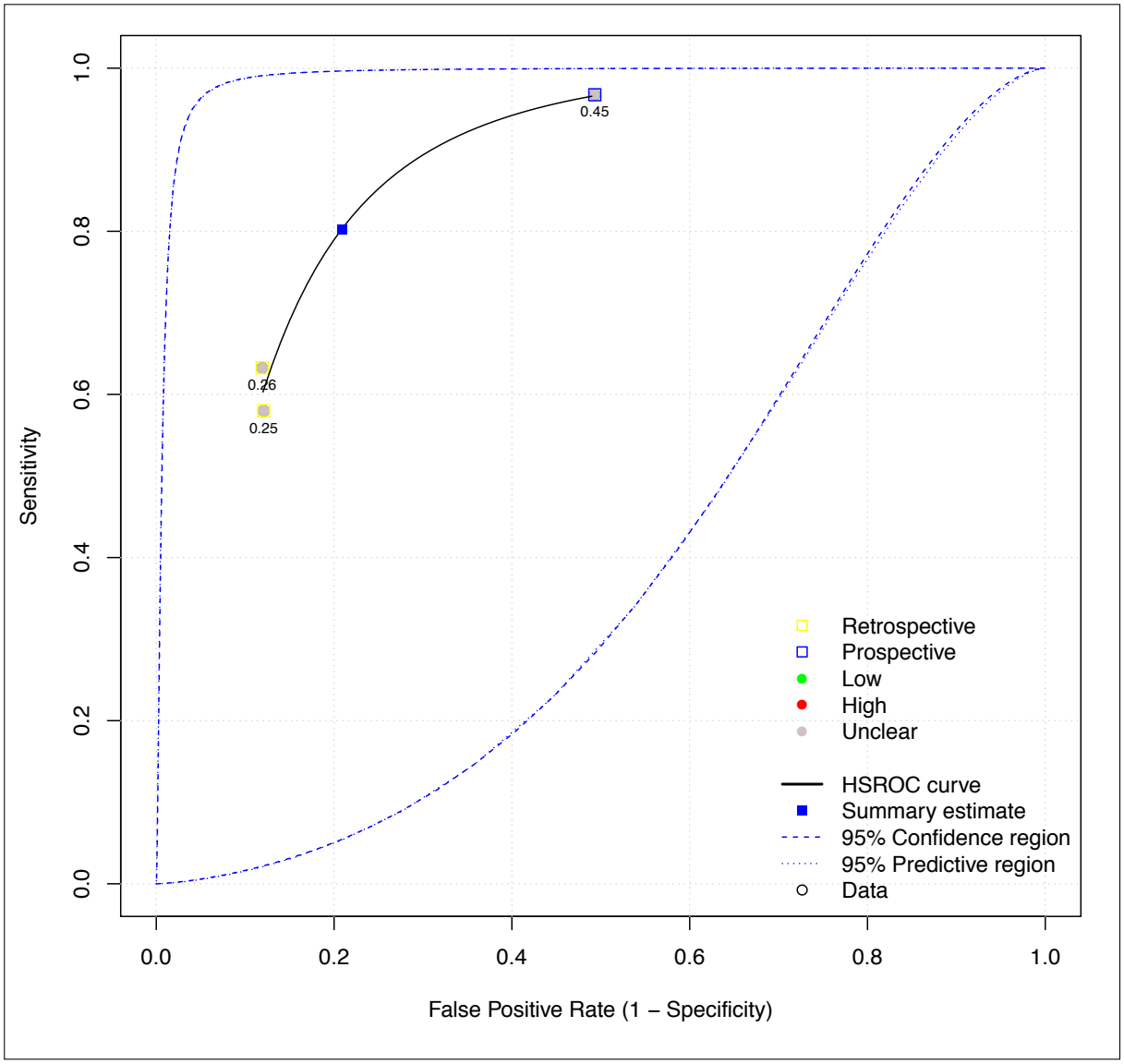
TP true positive **FP** false positive **FN** false negative **TN** true negative **95% CI** 95% confidence interval

Figure 2.6 is a forest plot of the accuracy of clinical judgement of GPs for the diagnosis of cognitive impairment in the four studies that had paired data on sensitivity and specificity. Excluding AgeCoDe, which was at high risk of bias in two QUADAS-2 domains and reported sensitivity 97% (95% CI 93% to 99%) specificity 94% (95% CI 93% to 95%), in individual studies sensitivity ranged from 58% (95% CI 49% to 67%) in Amsterdam to 97% (95% CI 93% to 99%) in Mannheim, which was the only study that reported higher sensitivity than specificity. Specificity ranged from 51% (95% CI 44% to 57%) in Mannheim to 88% in both Amsterdam (95% CI 84% to 91%) and Zwolle (95% CI 84% to 92%).

Figure 2.7 is a summary plot of the accuracy of clinical judgement of GPs for the diagnosis of cognitive impairment. The individual study points are shown with the calculated prevalence, together with an indication of whether the index test was prospective (blue square surround) or retrospective (yellow circle surround), and are coloured to indicate the risk of bias in the flow and timing domain (green low risk of bias, grey unclear risk of bias, red high risk of bias). The blue single filled dot indicates the summary point estimate of diagnostic accuracy, and is displayed on a summary ROC curve. The dashed bubble indicates the 95% confidence interval around the summary point and the larger dotted bubble indicates the 95% prediction region.

In the meta analysis for cognitive impairment (including dementia) as the target condition, the summary diagnostic accuracy of clinical judgement of general practitioners was sensitivity 80% (95% CI 45% to 95%), specificity 79% (95% CI 57% to 92%), DOR 15 (95% CI 8 to 29), positive likelihood ratio 3.8 (95% CI 2.6 to 5.8) and negative likelihood ratio 0.25 (95% CI 0.09 to 0.72). No single study had a comparable diagnostic accuracy to the summary point on either sensitivity or specificity. Furthermore, the 95% confidence region was the same size as the 95% prediction region which covered a large amount of ROC space. Therefore the summary point in isolation is over-simplistic as a representation of the diagnostic accuracy of clinical judgement for dementia and does not reflect the heterogeneity in the data.

Figure 2.7: Summary plot of GPs clinical judgement for diagnosis of cognitive impairment



Colours of dots: indicate risk of bias in flow and timing domain, all are at unclear risk of bias in this domain
Surround indicates prospective retrospective ; decimals indicate prevalence of target condition

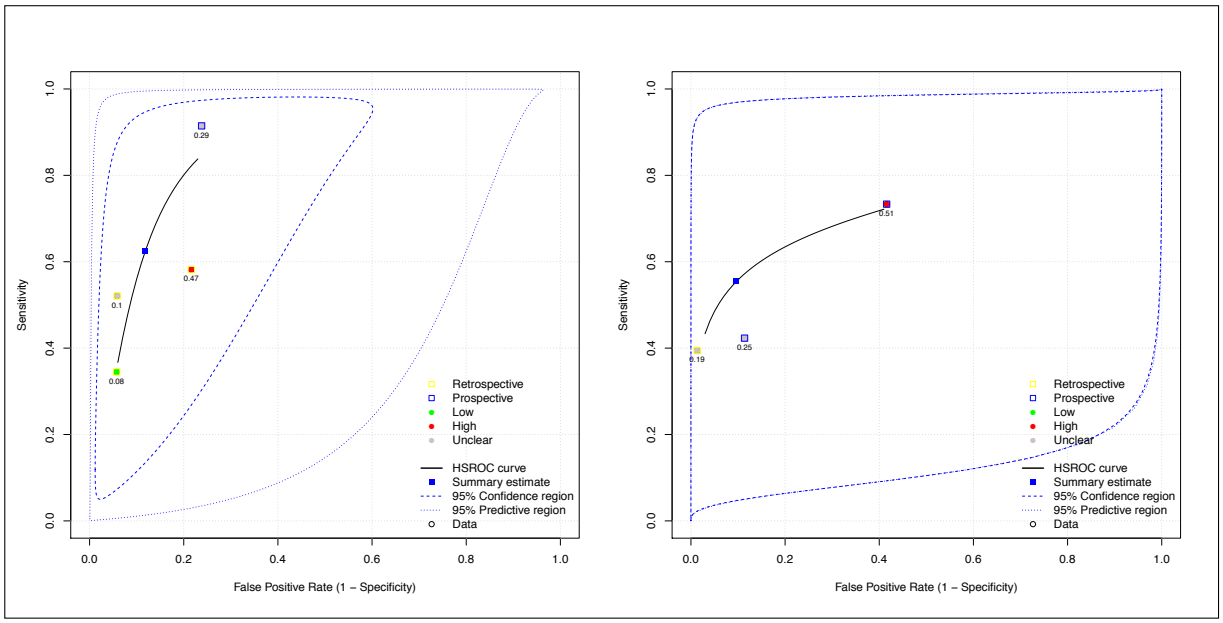
2.4.4.3 Heterogeneity

Methods for analysis are described in Section 2.3.10. Unfortunately due to technical limitations described in Section 2.3.9 it is not possible to plot all of the data with two curves on a single ROC space.

Definition Figure 2.8 shows the ROC plots for the studies by definition, with the studies that used the ICD-10 definition on the right and studies that used either DSM-III-R or DSM-IV-TR on the left. For both plots, especially the DSM plot there is significant uncertainty in the 95% confidence interval and 95% prediction region; for the DSM plot this fills almost the entire ROC space. The summary points are plotted but not reported in numbers because the uncertainty means these are a poor representation of the data. There does not appear to be any strong visual evidence to support the possibility of heterogeneity in sensitivity between ICD-10 and DSM definitions, because the ICD-10 data cover such a large amount of ROC space for sensitivity. It is possible there may be some heterogeneity in specificity by definition, with studies that used the ICD-10 or DSM-III-R definition appearing to report a higher specificity, but this could be due to chance. The study that used the DSM-IV-TR definition (France) had a higher sensitivity than the two studies that used a DSM-III-R definition (Sydney and Zwolle) at a cost of lower specificity, but this finding could be due to chance. Alternatively, the specificity in France could be an outlier, possibly related to the high risk of bias in the flow and timing domain. The shape of the ROC curves is broadly comparable, i.e. concave, with the slope gradient being steeper in the studies that used an ICD-10 definition than those that used a DSM definition. In a meta-regression model using `xtmelogit` the model could not converge.

Index test Figure 2.9 shows the ROC plots for the studies by index test with the studies that used a prospective clinical judgement on the right and a retrospective clinical judgement on the left. For both plots, especially the prospective index test plot there is significant uncertainty in the 95% confidence interval and 95% prediction region; for the prospective index test plot this fills the entire ROC space. The summary points are plotted but not reported in numbers because the uncertainty means these are a poor representation of the data. There does not appear to be any strong visual evidence to support the possibility of heterogeneity in specificity between prospective and retrospective judgement, but it is possible that a retrospective definition may be more specific than a prospective definition. Similarly, there does not appear to be any strong visual evidence to support the possibility of heterogeneity in sensitivity between prospective and retrospective judgement, but it is possible that a prospective judgement has a higher sensitivity than a retrospective judgement, but this could be due to chance. The shape of the ROC curves is broadly comparable, with the slope gradient being steeper in the studies that used a prospective judgement than those that used a retrospective judgement. In a meta-regression model using `xtmelogit` the model could not converge.

Figure 2.8: Summary plot showing clinical judgement of GPs for diagnosis of dementia according to ICD-10 (right) and DSM-III-R & DSM-IV-TR definition (left)

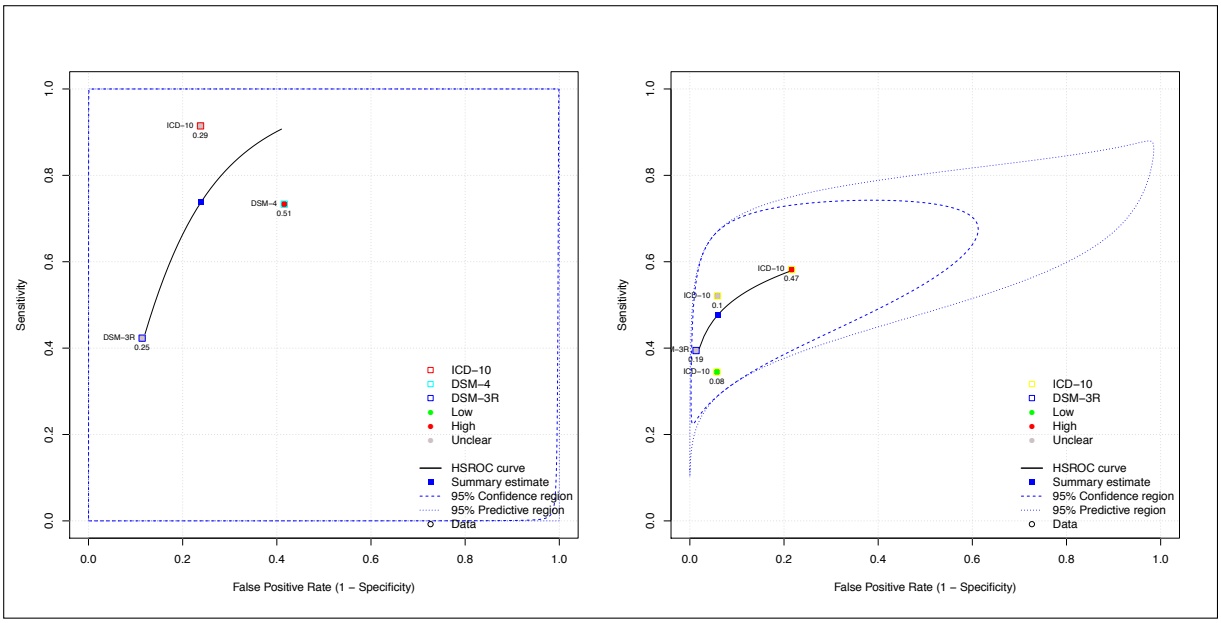


Colours of dots: indicate risk of bias in flow and timing domain, all are at unclear risk of bias in this domain
Surround indicates prospective retrospective ; decimals indicate prevalence of target condition in the left figure **France** - high risk in flow & timing - used DSM-IV-TR

Medical records Figure 2.10 shows the ROC plots for the studies by access to the medical records with the studies that allowed access to the medical records on the right and those that did not definitely allow access on the left. There is significant uncertainty in the 95% confidence interval and 95% prediction region; for the plot of studies with access to the medical records the 95% confidence interval contains a large amount of ROC space but the prediction region contains almost the whole plot, whereas for the plot with records either not available or uncertain the uncertainty extends over the whole plot. The summary points are plotted but not reported in numbers because the uncertainty means these are a poor representation of the data. It is plausible that studies that reported access to the medical records was available have higher sensitivity than those where access was not available or uncertain, however, there are two studies where access was available where sensitivity was low (Sydney and Zwolle) so this finding could be due to chance, especially as one of the two studies with higher sensitivity (France) was at high risk of bias in the flow and timing QUADAS-2 domain. There does not appear to be any strong visual evidence to support the possibility of heterogeneity in specificity by access to the medical records. The shape of the ROC curves is broadly comparable, i.e. concave, with a similar gradient to the slopes. In a meta-regression model using `xtmelogit` the model could not converge.

Flow and Timing Figure 2.11 shows the ROC plots for the studies by risk of bias in the QUADAS-2 flow and timing domain, with studies at low and unclear risk of bias on the right and studies at

Figure 2.9: Summary plot showing clinical judgement of GPs for diagnosis of dementia according to prospective index test (right) and retrospective index test (left)

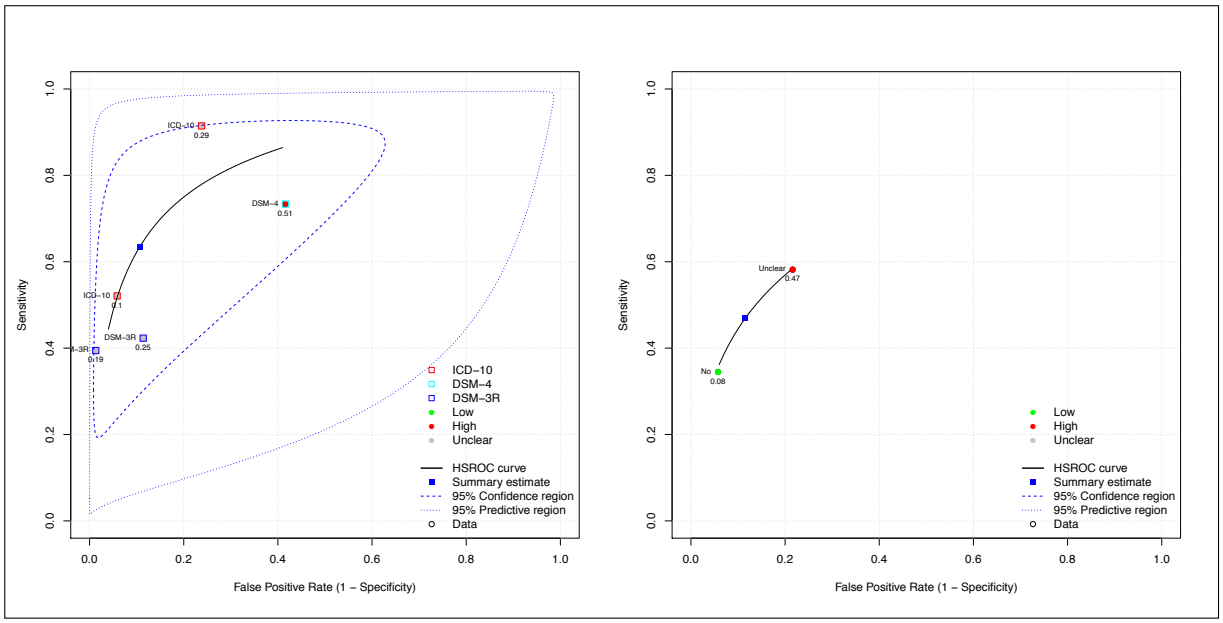


The coloured dots indicate the risk of bias in the flow and timing QUADAS-2 domain
decimals indicate prevalence of target condition

high risk of bias on the left. There is significant uncertainty in the 95% confidence interval and 95% prediction region especially for the left plot which encompasses the whole plot. For the plot on the right side, compared to the 95% confidence interval in the main analysis, shown in Figure 2.5, there is less uncertainty indicating that removing the studies at high risk of bias in the flow and timing QUADAS-2 domain has reduced the uncertainty. However, the 95% prediction interval still contains a large amount of ROC space. The summary points are plotted but not reported in numbers because the uncertainty means these are a poor representation of the data. It is plausible that studies that were at high risk of bias in the flow and timing domain had lower specificity than studies that were at low or unclear risk of bias in this domain. However, this finding could also be due to chance, because only one study was at low risk of bias in the flow and timing domain (Cambridge). The shape of the ROC curves is broadly comparable, i.e. concave, with the gradient slope being steeper in the plot of studies at low and unclear risk of bias. In a meta-regression model using `xtmelogit` the model could not converge.

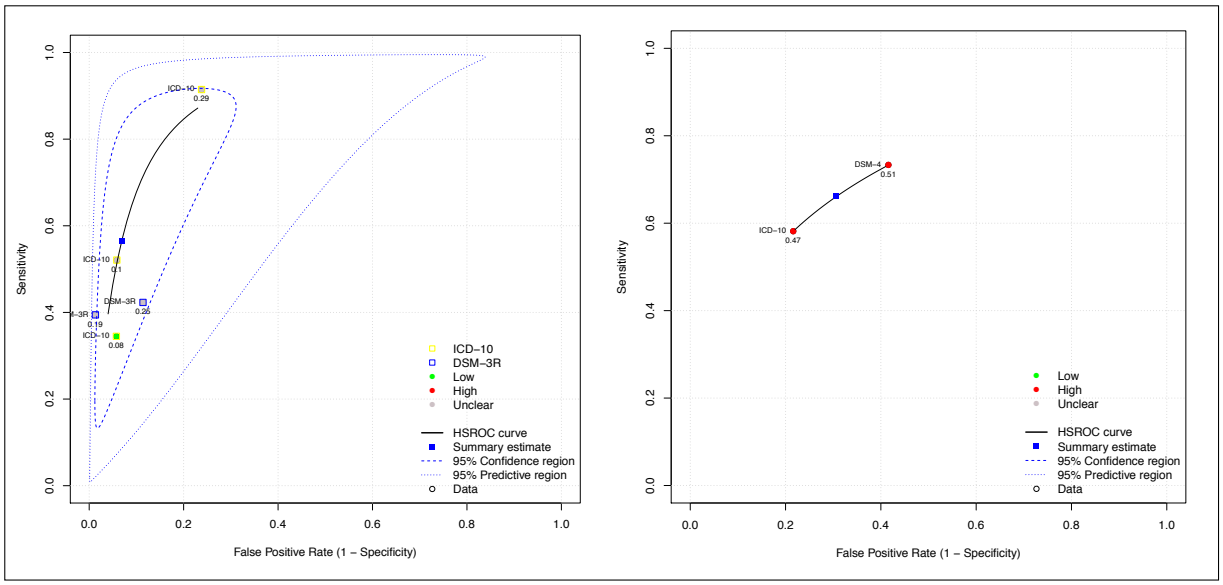
Cognitive impairment target condition There were three studies in the meta-analysis with cognitive impairment as the target condition, which restricted the opportunity to explore heterogeneity. For definition, two studies used the ICD-10 definition and one used the DSM-III-R definition. For index test, two studies used a retrospective judgement and one used a prospective judgement. For access to the medical records, all studies allowed access. For risk of bias in the flow and timing

Figure 2.10: Summary plot showing clinical judgement of GPs for diagnosis of dementia by access to medical records available (right) and not available or uncertain (left)



The coloured dots indicate the risk of bias in the flow and timing QUADAS-2 domain
ICD-10 was the definition used for both studies on the left plot
decimals indicate prevalence of target condition

Figure 2.11: Summary plot showing clinical judgement of GPs for diagnosis of dementia by risk of bias in QUADAS-2 flow and timing domain: low or unclear (right) and high (left)



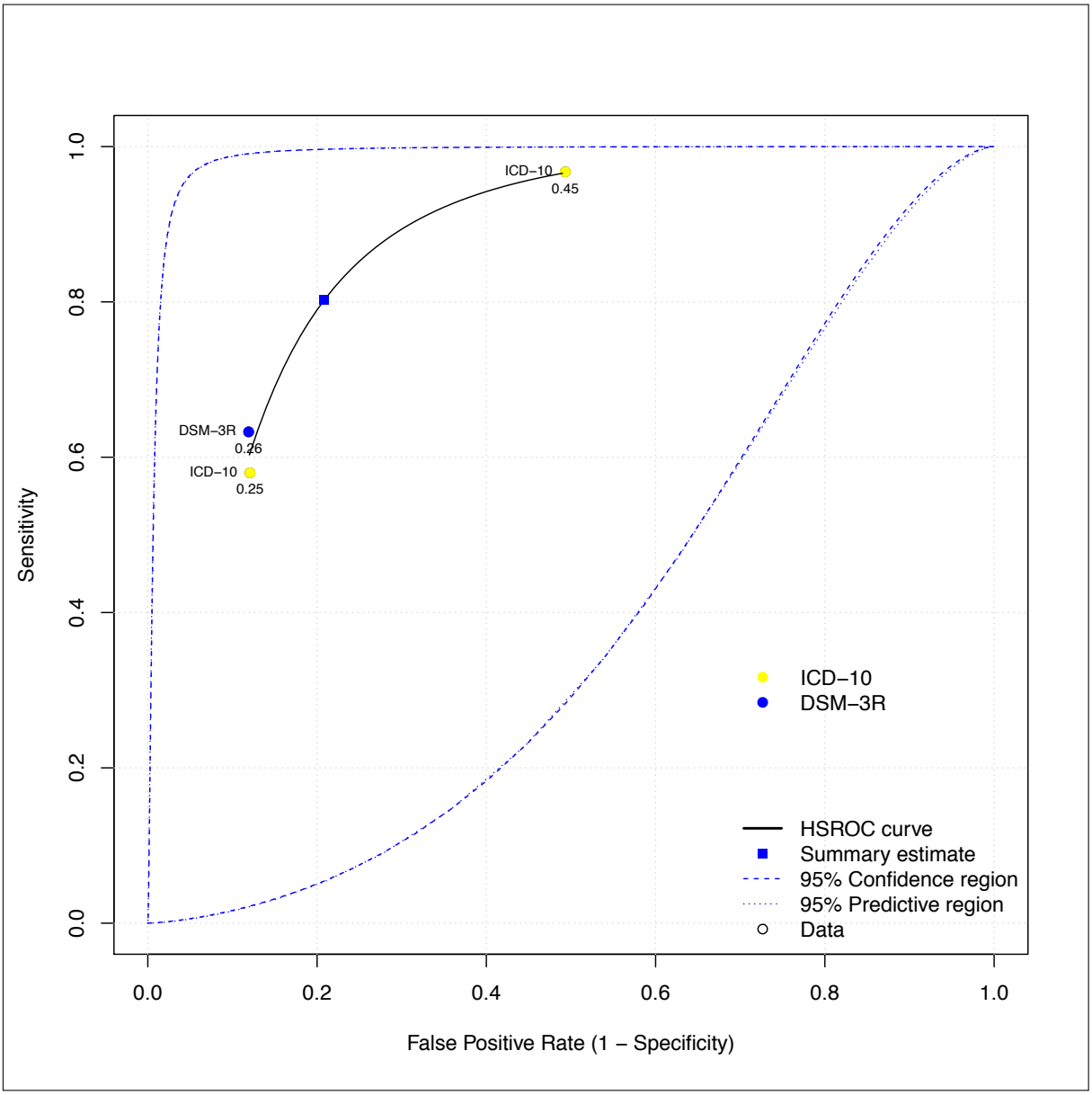
The coloured dots indicate the risk of bias in the flow and timing QUADAS-2 domain
decimals indicate prevalence of target condition

QUADAS-2 domain all three studies were at unclear risk of bias.

Figure 2.12 plots the data for the target condition cognitive impairment. One study used a prospective index test, and used the ICD-10 definition for the target condition (Mannheim); whereas two studies used a retrospective index test, of which one used the ICD-10 definition for the target condition (Amsterdam) and one used the DSM-III-R definition (Zwolle). Figure 2.12 shows that there is no evidence of heterogeneity by definition for the cognitive impairment target condition, because two studies that used different definitions (Zwolle, DSM-III-R; Amsterdam, ICD-10) had similar sensitivity and specificity. However, one of the two studies (Mannheim) that used the ICD-10 definition had higher sensitivity and lower specificity than Zwolle so it is impossible to draw firm conclusions. Because of the small number of studies there is a possibility of a type 2 error, i.e. that there may be heterogeneity which is not shown. Figure 2.13 (right) shows that compared to the analysis of all three studies including the one that used the DSM-III-R definition, in the analysis of the two studies that used the ICD-10 definition the ROC curve is the same shape, fills a similar ROC space, but has a wider 95% confidence and prediction interval (so wide that it fills the plot).

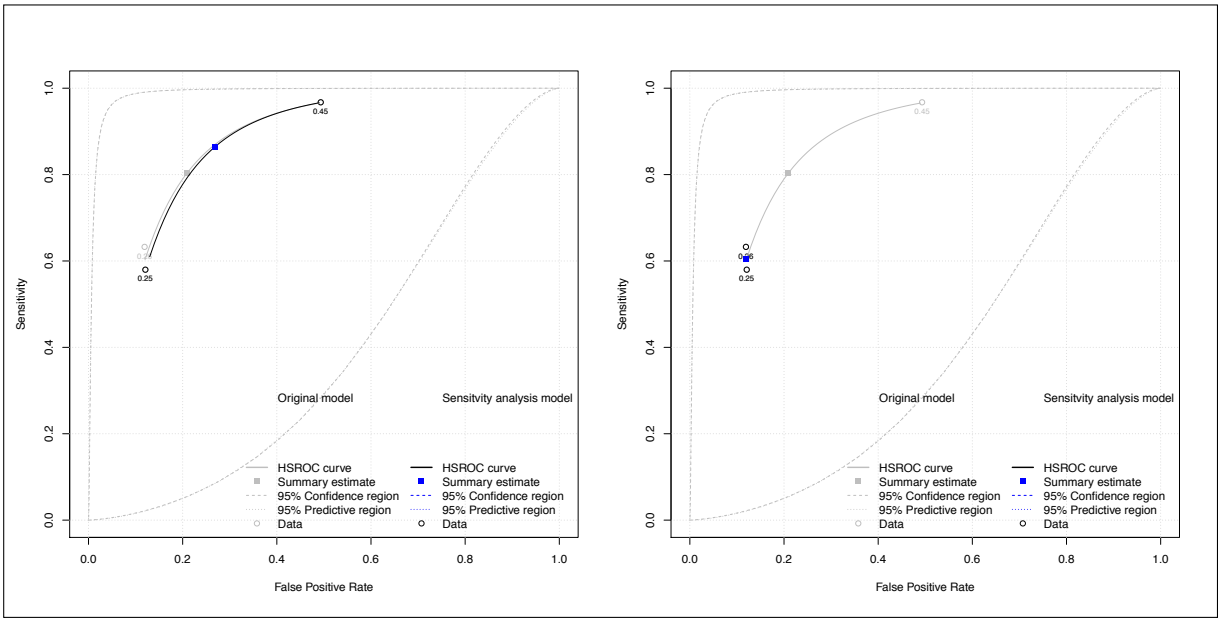
Figure 2.13 (left) shows a plot of studies that used retrospective judgement compared to the one that used prospective judgement. There is a possibility that prospective clinical judgement is more sensitive and less specific than retrospective clinical judgement but there is significant uncertainty and no firm conclusions can be drawn. There were insufficient studies to perform meta-regression.

Figure 2.12: Summary plot showing clinical judgement of GPs for diagnosis of cognitive impairment target condition showing definition



Amsterdam: ICD-10, retrospective; **Mannheim**: ICD-10, prospective, highest sensitivity; **Zwolle**: DSM-III-R, retrospective
decimals indicate prevalence of target condition

Figure 2.13: Summary plot for clinical judgement for target condition cognitive impairment, showing the summary curve for studies that used the ICD-10 definition (right) and retrospective judgement (left)



Amsterdam: ICD-10, retrospective; **Mannheim**: ICD-10, prospective, highest sensitivity; **Zwolle**: DSM-III-R, retrospective
The sensitivity analysis model: is the model with either just ICD-10 studies (right) or just retrospective judgement (left)
The original model: is the model with all three studies; the labelling ("sensitivity analysis") cannot be changed
decimals indicate prevalence of target condition

2.4.4.4 Sensitivity analyses

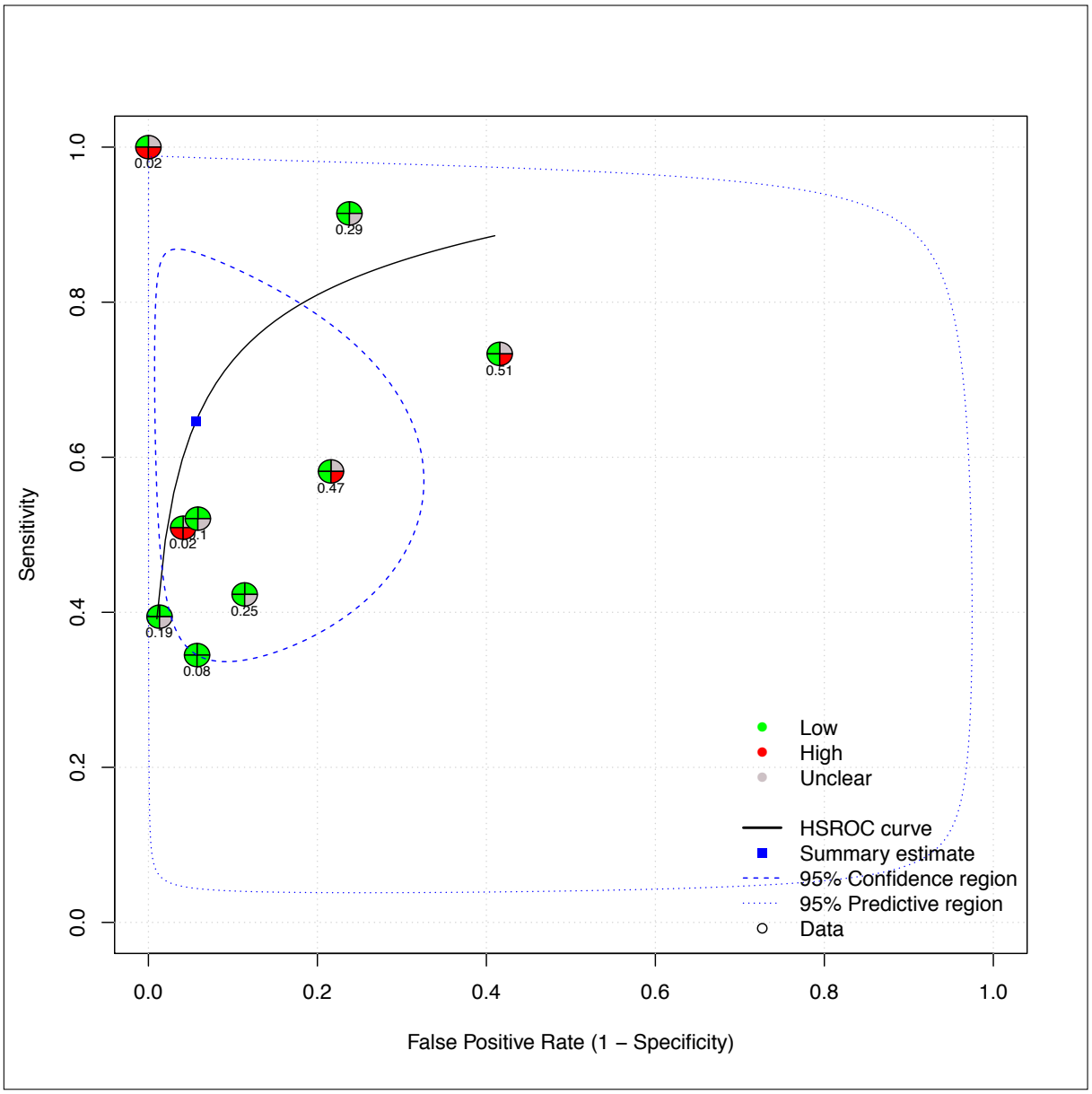
Dementia target condition Figure 2.14 shows the summary ROC plot for the pre-specified sensitivity analysis which included the two studies (AgeCoDe and Antwerp) that were judged to be at high risk of bias in two QUADAS-2 domains. Compared to the main analysis there is greater uncertainty and more heterogeneity in the data. One of the studies that was excluded from the main analysis (Antwerp) is an outlier, with very high sensitivity and high specificity; the other study at high risk of bias in two QUADAS-2 domains (AgeCoDe) had more comparable sensitivity and specificity to the other eight studies. The summary estimate was sensitivity 65% (95% CI 44% to 80%) specificity 94% (95% CI 82% to 98%). This compares to sensitivity 59% (95% CI 41% to 74%) specificity 89% (95% CI 77% to 95%) in the main analysis. Therefore when including the two studies that were judged to be at high risk of bias in more than two QUADAS-2 domains the sensitivity was 6 percentage points higher and the specificity was 5 percentage points higher; both were within the 95% CI for the main analysis. However, the uncertainty in the estimates was increased substantially as shown by the wider 95% confidence and prediction regions in the summary ROC plot and because of heterogeneity the point estimates are not a good representation of the data.

Cognitive impairment target condition Figure 2.15 shows the summary ROC plot for the pre-specified sensitivity analysis for the target condition cognitive impairment. Compared to the main analysis there is greater uncertainty and more heterogeneity in the data. The study that was excluded from the main analysis (AgeCoDe) is an outlier, with very low sensitivity compared to the other three studies. The summary estimate was sensitivity 72% (95% CI 33% to 93%) specificity 79% (95% CI 57% to 91%). This compares to sensitivity 80% (95% CI 45% to 95%) specificity 79% (95% CI 57% to 92%) in the main analysis. Therefore when including the study that was judged to be at high risk of bias in more than two QUADAS-2 domains the sensitivity was 8 percentage points lower and the specificity was the same; both were within the 95% CI for the main analysis. However, the uncertainty in the estimates was increased substantially as shown by the wider 95% confidence and prediction regions in the summary ROC plot and because of heterogeneity the point estimates are not a good representation of the data.

Re-coded target condition By re-coding the TN as TP it is possible to meta-analyse the accuracy of GPs for the target condition normal, which is sensitivity 79% (95% CI 57% to 92%) specificity 80% (95% CI 45% to 95%). [3]

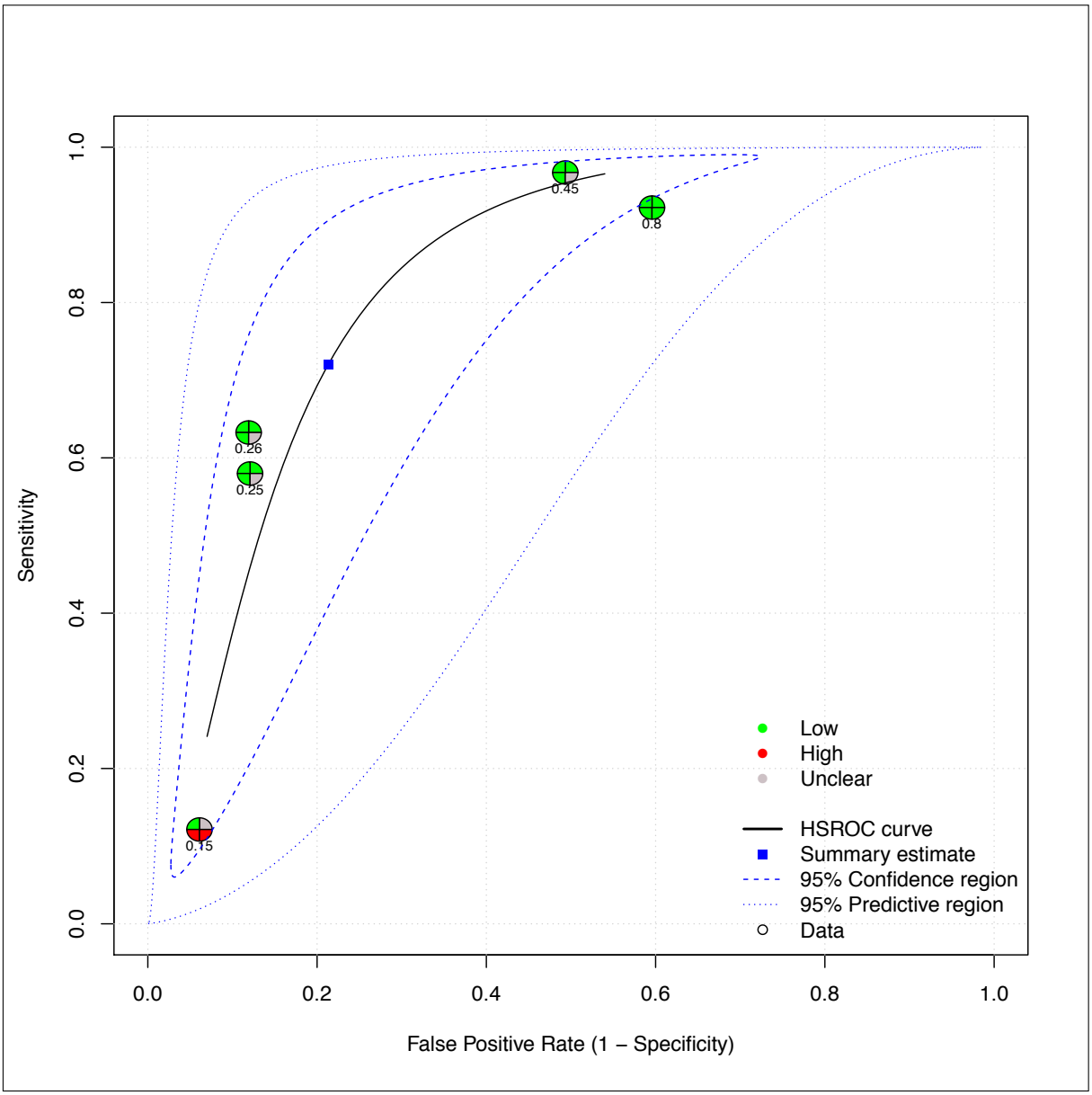
[3] If normal were the target condition instead of cognitive impairment then people who are true negative ($Target^{-}$ $Test^{-}$) for cognitive impairment as the target condition become true positive ($Target^{+}$ $Test^{+}$) for normal as the target condition. For example: a score of <24 on the MMSE indicates cognitive impairment; a candidate (who is normal) scores 25. For the target condition cognitive impairment they are $Test^{-}$ $Target^{-}$ so they are a TN. For the target condition normal they are $Test^{+}$ $Target^{+}$ so they are a TP. By analogy, FN become FP, FP become FN, and TP become TN.

Figure 2.14: Summary plot for sensitivity analysis of clinical judgement for target condition dementia



Amsterdam: ICD-10, retrospective; **Mannheim**: ICD-10, prospective, highest sensitivity; **Zwolle**: DSM-III-R, retrospective
decimals indicate prevalence of target condition

Figure 2.15: Summary plot for sensitivity analysis of clinical judgement for target condition cognitive impairment



Amsterdam: ICD-10, retrospective; **Mannheim**: ICD-10, prospective, highest sensitivity; **Zwolle**: DSM-III-R, retrospective
decimals indicate prevalence of target condition

2.5 Discussion of the systematic review

2.5.1 Summary of main findings

Five electronic databases were searched for studies investigating the accuracy of clinical judgement of general practitioners for the diagnosis of two target conditions, dementia, and cognitive impairment including dementia. From 12,681 records retrieved, sixteen articles referring to 10 studies were included in the review but one study did not report paired data on test accuracy. Overall the included studies were judged to be at low risk of bias and to have low concern about applicability. However, five of the seven studies included in the meta analysis were at high risk of bias in the flow and timing QUADAS-2 domain and two were also at high risk of bias in the reference standard domain.

In the seven studies in the meta analysis for dementia as the target condition the test accuracy ranged from sensitivity 34% (95% CI 18% to 54%) for Cambridge to 91% (95% CI 85% to 96%) for Mannheim; and specificity ranged from 58% (95% CI 51% to 66%) for France to 99% (95% CI 97% to 100%) for Zwolle. In the three studies in the meta analysis for cognitive impairment as the target condition the test accuracy ranged from sensitivity 58% (95% CI 49% to 67%) for Amsterdam to 97% (95% CI 93% to 99%) for Mannheim; and specificity ranged from 51% (95% CI 44% to 57%) for Mannheim to 88% (95% CI 84% to 91%) for Amsterdam and 88% (95% CI 84% to 92%) for Zwolle. Because of the wide 95% confidence intervals and prediction intervals there is significant uncertainty in the findings and they should not be regarded as definitive.

There was heterogeneity in the data between studies, as illustrated in Figures 2.5 and 2.7. Diagnostic accuracy reviews often find greater heterogeneity in sensitivity than specificity because primary diagnostic accuracy studies typically contain fewer people with disease than without disease [220]. It was difficult to draw firm conclusions about heterogeneity but the data were compatible with studies that used ICD-10, or applied retrospective judgement, having higher specificity compared to studies with DSM definitions or using prospective judgement. For sensitivity there was no evidence of heterogeneity by definition, studies that used a prospective index test may have had higher sensitivity than studies that used a retrospective index test, and studies that allowed access to the medical records may have had higher sensitivity than studies that did not or where this was unclear. Studies at high risk of bias in the flow and timing domain appeared to have lower specificity and sensitivity than studies at unclear or low risk of bias in this domain. However, all of these findings could have been due to chance and there was significant uncertainty.

2.5.2 Strengths and limitations

2.5.2.1 Strengths and weaknesses of the included studies

One limitation of the included studies, which is inherent in the general practice setting, is the low prevalence of the target condition, which consequently leads to higher uncertainty in the estimates of diagnostic test accuracy and contributes to the heterogeneity between studies. The seven studies in

the meta analysis incorporated a total of 445 people who were true positives and 249 false negatives for the target condition dementia, totalling 694 people who were disease positive out of 2550, for a prevalence of 27% (95% CI 26% to 29%). In contrast the average prevalence of dementia in the included studies, calculated for each study as $\frac{truepositives+falsenegatives}{total}$ was 21%, and ranged from 2% in AgeCoDe and Antwerp to 47% in Cambridge City and 51% in France. Participant flow in studies, and specifically incomplete verification of the target condition with the reference standard means that calculated prevalence may not accurately reflect the true prevalence of the target condition. However, only two studies reported a figure for prevalence that accounted for the design of participant flow: Zwolle [265] reported a prevalence of 7% and Cambridge City [270] reported a prevalence of 11%. In contrast prevalence of dementia in the community is reported to be around 6% [239, 287].

The strength of evidence in this review is restricted by limitations in the primary studies regarding participant flow, and heterogeneity in the data. Only one study in the meta analysis [267] was judged to be at low risk of bias in the flow and timing QUADAS-2 domain. This may in part reflect the practical difficulties of investigating a disease in a low prevalence setting: large numbers of patients require evaluation to identify the people with disease; evaluating large numbers of people with a reference standard is resource intensive (i.e. expensive) and arguably burdensome for people who are unlikely to have a cognitive disorder. Partial verification may have led to over optimistic estimates of diagnostic accuracy in some studies. In contrast all studies were judged at low risk of bias in the index test QUADAS-2 domain. However, despite this there was still considerable heterogeneity in the data, which is probably inherent in the nature of the index test. Background literature relating to clinical judgement is discussed in detail in Section 1.1.3 and Section 2.5.2.2 discusses the review findings in context.

In contrast, the evidence in this review is supported by methodological strengths of the included studies, such as participant sampling and the conduct of the index test and reference standard. Five of the seven studies in the meta analysis were at judged at low risk of bias in the patient selection QUADAS-2 domain and the remainder to be at unclear risk of bias; all studies reported that participants were sampled either consecutively or randomly. All seven studies were at low risk of bias in the reference standard QUADAS-2 domain.

The estimates of accuracy varied between studies but despite this there were consistent trends. In general specificity was higher than sensitivity, specificity was relatively high (at least 0.78 in eight studies), whereas sensitivity was modest (at best 0.91 in eight studies). The findings were robust to a sensitivity analysis. Both prospective and retrospective clinical judgement generally had higher specificity than sensitivity. These findings suggest that clinical judgement of GPs is generally better at identifying healthy people as disease free than recognising disease in affected people. Section 2.5.4 and Figure 2.16 explore the implications of this for practice.

There was low concern about the applicability of studies to the review question. Only three studies had unclear concern[4] about the applicability of the index test but otherwise all studies were

[4]QUADAS-2 categorises risk of bias as high risk, low risk, or unclear risk, and applicability as high concern, low concern,

of low concern in all QUADAS-2 applicability domains.

Overall the characteristics of studies, quantity, numbers of disease positive, and consistency of the findings are supportive factors for the strength of evidence in the review. In contrast the strength of evidence is diluted by the heterogeneity in the data. However, heterogeneity is present in all diagnostic test accuracy reviews to some extent, and the nature of clinical judgement is that it will tend to be more heterogeneous than a machine read test. Given the nature of the index test, substantial heterogeneity in the data is likely to be unavoidable.

2.5.2.2 Strengths and weaknesses of the review process

The search was comprehensive and systematic. The search strategy was written in consultation with a highly experienced systematic reviewer with substantial expertise in diagnostic test accuracy studies investigating cognitive disorders and dementia. A number of databases were searched and the search yielded more results that previous reviews, notwithstanding the later search date. All studies that were identified by previous systematic reviews were identified by the current search strategy, though not all were included in the review. The comprehensiveness of the search makes it unlikely that the findings would be undermined by some unidentified study, but this cannot be fully excluded. In particular unpublished studies are hard to identify, and methods to identify reporting bias in diagnostic test accuracy reviews are not yet well established [220].

Studies were excluded that used a documented diagnosis of dementia in the medical record as the index test. This meant that some studies that had been included by previous reviews were not included in this review. Furthermore, it is not possible to infer the accuracy of documented diagnosis in the GP medical record for the diagnosis of dementia from this review. However, this exclusion means that the findings of this review are highly applicable to the review question which is arguably more focused on clinical practice than previous reviews. Understanding the accuracy of a documented diagnosis of dementia in the medical record would be helpful if (for instance) investigators were seeking to ascertain people with dementia diagnosed by a GP in a database of routinely collected data from primary care. Other investigators have reported inconsistency in ascertaining cases of dementia from routinely collected data [288].

The process for including and excluding studies was robust, as was data extraction, being done separately in duplicate by two reviewers and disagreements resolved by consensus. Authors of original studies were contacted when necessary to obtain data on diagnostic accuracy, but this was not available. The approach to the statistical analysis used standard techniques that are recommended by the Cochrane diagnostic test accuracy review methods group. Recently reported approaches such as imputation of data [289] were not used, but are less widely recognised and are not yet mentioned in the Cochrane handbook [150, 220] .

An important limitation is the difficulty in understanding how the accuracy of clinical judgement for the diagnosis of dementia or cognitive impairment is related to the stage of the condition. De-

or unclear concern

mentia is a progressive disease, and it is possible that GPs do not make a diagnosis of dementia until the condition has advanced to a stage where the diagnosis is relatively apparent, perhaps even to a non-medical person. If GP clinical judgement is only accurate in advanced stages of the disease then arguably it contributes little beyond the opinion of the patients family and friends. However, attempting to analyse how the accuracy of clinical judgement varied over different stages of disease would have led to small numbers in the analysis, and would have been hampered by reporting in primary studies. The review did however include two target conditions: dementia, and cognitive impairment including dementia. The cognitive impairment target condition includes a wider spectrum of people including those with milder degrees of cognitive impairment was well as those with florid dementia. There was no strong evidence of difference in the sensitivity and specificity of clinical judgement for the two target conditions, with overlapping confidence intervals and the findings limited by uncertainty, however when comparing Figure 2.5 and Figure 2.7 it is possible that the sensitivity of clinical judgement for cognitive impairment may be higher than for dementia, without loss of specificity.

An important aspect of the interpretation of the results is the threshold of clinical judgement for the diagnosis of the target conditions. It is only sensible to estimate average sensitivity and specificity at a common test threshold [220]. The philosophy in this review is that the common threshold for the diagnosis of clinical judgement is the diagnostic label of dementia or cognitive impairment. A clinician makes a clinical judgement about the presence of dementia (or indeed any diagnosis) when they judge that the patient fits better, on balance, *in* that group of people than *outside* of the group, especially with regards to prognosis and response to treatment [140, 175]. While different clinicians may formulate differing conclusions about whether or not a condition is present in a particular patient, their threshold for clinical judgement in decision making is likely to be a function of factors such as the implications of the disease (regarding prognosis or treatment), their familiarity with the patient, and the urgency of the decision. This review takes the stance that if a GP participating in one study (for instance Antwerp) had instead been participating in a different study (for instance Amsterdam) evaluating the same patient under the the same circumstances would lead to the same decision about the target condition, because the *threshold* for the target condition in both studies was consistently a diagnostic label of dementia. That is, that there is low intra-observer variability in the classification of a person as having dementia. Unfortunately, there is very little evidence on the intra-observer variability of the diagnosis of dementia in general practice. Available data indicate good int*er*-observer agreement (kappa 0.63-0.90) for the categorisation of dementia | no dementia [290–292] but these studies are based on specialists applying standardised criteria and no similar studies investigating the reliability of clinical judgement of general practitioners could be identified. However, the alternate view is that there is no common threshold for clinical judgement because it is a subjective test. This view is considered less plausible because this position would mean that despite identical circumstances, clinician and patient a different classification could be reached regarding the target condition (due to varying test threshold), and (though the play of chance is acknowledged)

this is considered unlikely. If correct, the view of no common index test threshold would imply that it is not appropriate to perform bivariate meta-analysis of the data on clinical judgement. As discussed in Section 1.1.3 many tests (perhaps all other than those which are read by a machine) are subject to a degree of clinical judgement, user variability and implicit thresholds.

There were four differences between the methods as described in this Chapter and those in the published protocol [197]. Firstly, studies were not restricted to people with cognitive symptoms for because (a) this restriction was judged on reflection to be incompatible with the retrospective definition of clinical judgement (which was included in the published protocol) and (b) this approach would have been overly restrictive. Secondly, all abstracts were screened by two reviewers rather than one. Thirdly, information on covariates was not extracted for stage of dementia, experience of GPs, proportion of male and female doctors, or type of practice because these factors were found to be poorly reported and judged on reflection to be of less relevance to the review question. Fourthly, based on discussion with expert statistical advisers, the main analysis was in studies at lowest risk of bias and a sensitivity analysis was done when studies at high risk of bias in two or more QUADAS-2 domains were included.

Table 2.4: Diagnostic accuracy of clinical judgement for diagnosis of dementia in context

Test	**Number of studies**	**Total participants**	**Threshold for normal**	**Prevalence of dementia**	**Sensitivity** (95% CI)	**Specificity** (95% CI)	**Diagnostic odds ratio**
Mini-Cog [293]	4		Max = 4				
[294]		142	≤ 2	35%	1.00 (0.93 to 1.00)	0.40 (0.30 to 0.50)	∞
[295]		423	≤ 2	5%	1.00 (0.84 to 1.00)	0.85 (0.81 to 0.89)	∞
[296]		383	≤ 2	6%	0.76 (0.53 to 0.92)	0.73 (0.68 to 0.77)	9 (95% CI 3 to 30)
[297]		569	≤ 3	90%	0.84 (0.81 to 0.87)	0.27 (0.16 to 0.41)	2 (95% CI 1 to 4)
26 item IQCODE [298] ∗	1	262	max = 5	7%			
[171]			3.2		1.0 †	0.76	∞
			3.3		1.0	0.82	∞
			3.4		1.0	0.87	∞
			3.5		0.88	0.91	72 (95% CI 14 to 670)
			3.6		0.81	0.96	99 (95% CI 21 to 593)
			3.7		0.75	0.98	158 (95% CI 29 to 931)
AD8 [299]	1	309	max = 8	14%			
[170]			3		0.91 (0.78 to 0.97)	0.91 (0.87 to 0.94)	100 (95% CI 32 to 406)
MMSE [156] ‡	6		max = 30				
[294]		360	24	21%	1.00 (0.95 to 1.0)	0.46 (0.40 to 0.52)	∞
[300]		303	24	26%	0.81 (0.70 to 0.89)	0.65 (0.59 to 0.72)	8 (95% CI 4 to 16)
[301]		176	25	47%	0.80 (0.70 to 0.88)	0.76 (0.66 to 0.84)	13 (95% CI 6 to 28)
[302]		160	19	9%	0.80 (0.52 to 0.96)	0.85 (0.80 to 0.91)	25 (95% CI 6 to 145)
[303]		314	23	9%	0.68 (0.48 to 0.84)	0.92 (0.88 to 0.95)	24 (95% CI 9 to 67)
[266]		368	24	16%	0.37 (0.24 to 0.51)	0.95 (0.92 to 0.97)	11 (95% CI 5 to 24)
Clinical judgement							
For dementia	7	2550	-	27%	0.59 (0.41 to 0.74)	0.89 (0.77 to 0.95)	11 (95% CI 6 to 22)
For cognitive impairment	3	1257	-	32%	0.80 (0.45 to 0.95)	0.79 (0.57 to 0.92)	15 (95% CI 8 to 29)

∗ *IQCODE* Short Form of the Informant Questionnaire on Cognitive Decline in the Elderly. Typically the short 16 item version [304] is used in practice
† confidence intervals not available. ‡ *MMSE* Mini Mental State Examination

2.5.3 Relation to literature

Table 2.4 presents the results in the context of the existing literature, based on findings of a series of Cochrane reviews of diagnostic test accuracy that reported on the accuracy of brief cognitive tests for the diagnosis of dementia in primary care. Comparisons between tests are difficult because of variations in study design such as sampling, reference standard and especially participant flow. Additionally different sensitivity and specificity combinations may result in the same diagnostic odds ratio [220]. Furthermore, the summary estimates in this review do not reflect the heterogeneity in the data. Comparisons between different studies requires great care because the comparisons are indirect and are very susceptible to confounding by factors other than the test of interest. However, in general in comparison to the tests that are outlined in Table 2.4 clinical judgement has lower sensitivity but higher specificity. To some extent this is expected. Recall that specificity = $\frac{TN}{TN+FP}$. As discussed in the previous paragraph, GPs in studies that contributed to this review may have understood their clinical judgement *as equivalent to* the diagnosis, with potentially important implications for their patients and so they may have been erring on the side of minimising the number of false positives, which would tend to lead to a higher specificity. From Figure 2.5 and Figure 2.7 it is plausible that studies with higher sensitivity are also those with a higher prevalence of the target condition, which may suggest that the implicit threshold for clinical judgement of dementia varies with the prevalence of the target condition. However, it is difficult to draw any firm conclusions about because of limitations in the data, for example in Figure 2.5 two studies with similar prevalence have very different sensitivity (Sydney prevalence 25% sensitivity 42%; Mannheim prevalence 29% sensitivity 91%) that are both at unclear risk of bias in the flow and timing QUADAS-2 domain. In contrast, many of the brief cognitive tests that are outlined in Table 2.4 have the purpose of improving the identification of people with cognitive disorders and so would tend to minimise the number of false negatives, optimising sensitivity.

None of the Cochrane systematic reviews of brief cognitive tests were able to perform meta analysis, either because of insufficient studies or because of heterogeneity in the data, particularly regarding varying test threshold. For the mini-cog in four studies the sensitivity for the diagnosis of dementia varied between 76% to 100% and the specificity varied between 27% to 85%; the study at the lowest risk of bias, at low risk on all QUADAS-2 domains, [296] reported a sensitivity of 76% and specificity 73% for a DOR of 9. For the IQCODE only one study, which was at high risk of bias in all domains, was identified in primary care, which used the long form (26 item) rather than the more commonly used (and recommended [305]) 16 item version. At a threshold of 3.2 the IQCODE for the diagnosis of dementia had sensitivity 100% and specificity 76% whereas at a threshold of 3.7 the sensitivity was 75%, specificity 98% [171]. For the AD8 in one study in primary care, which was at high risk of bias in the flow and timing QUADAS-2 domain and unclear risk of bias in the selection and index test QUADAS-2 domains, the sensitivity and specificity were both 91% [170]. For the MMSE in six studies in primary care the accuracy for the diagnosis of dementia ranged from sensitivity 37% specificity 95% at a threshold of 24 [266] to sensitivity 100% specificity 46% at a threshold of 24 [294];

none of the studies were at high risk of bias in more than one QUADAS-2 domain.

Overall in the context of the existing literature the present findings suggest that clinical judgement of GPs is comparable to brief cognitive tests for the diagnosis of dementia. Clinical judgement for the diagnosis of dementia is a comparatively high specificity, moderate sensitivity test. The accuracy of clinical judgement for the diagnosis of dementia DOR 11 (95% CI 6 to 22) is comparable to reported figures for the accuracy of the MMSE at a threshold of 24, DOR 11 (95% CI 5 to 24), and the accuracy of clinical judgement for the diagnosis of cognitive impairment DOR 15 (95% CI 8 to 29) is comparable to reported figures for the accuracy of the MMSE at a threshold of 25, DOR 13 (95% CI 6 to 28).

Clinical pathway Clinical judgement, as described in Section 1.1.3 is a categorisation process which is (as defined in Section 2.3.3) *unaided by any additional test, investigation or inquiry beyond that which is immediately available to the clinician.* Arguably, clinical judgement, as detailed in Section 1.1.3, is typically formulated without conscious direction. Clinicians will typically reach a judgement about problems which they are presented with, however the problem arises: for instance when performing chest auscultation a GP may observe an abnormal looking mole and decide the patient requires referral for this, even if the patient did not know the mole was there. Similarly, a GP may observe that a patient is having cognitive problems, for example being repetitive or vague in their presentation of the clinical problem, even if the patient is not presenting about these things directly. The extent to which the GP identifies these coincidental problems probably depends on the extent to which the non-explicitly presented problems are consciously available to the GP (the availability heuristic; Section 1.1.3.3)

The Author suggests that typically the presentation of cognitive problems is broadly speaking one of the following scenarios:

1. Cognitive problems, noted by the patient or someone else, are the main reason for the encounter in a routine (planned) encounter;

2. Cognitive problems are not the main reason for the encounter but are opportunistically noted by the clinician;

3. A crisis has occurred. This may be a physical crisis (such as chest infection) with associated cognitive problems such as delirium, or a cognitive crisis such as getting lost, wandering, leaving the stove on, or a driving accident.

It is likely that the prevalence of underlying cognitive problems is different in each scenario, for example, cognitive problems may be more likely in scenarios 1 and 3 than in scenario 2, but this is a question for future research. The findings of this review suggest that clinical judgement of dementia and cognitive impairment is a comparatively high specificity moderate sensitivity test. Using clinical judgement without any additional test, especially when the patient presents with subjective cognitive problems, would be at risk of missing cases of dementia; this is discussed further in Paragraph *Natural frequencies for tests and judgement* below. The role of clinical judgement in the clinical pathway

depends partly on the clinical scenario. For example, in scenario 1 above, clinical judgement may be an add-on test to other measures because it is likely that some form of objective testing will be needed to help meet patients' ideas about what they expect from an encounter which is focused on cognition. In contrast, in scenario 2 and 3 clinical judgement is unavoidable (recall from Section 1.1.3.2 that clinical judgement is a system 1 cognitive process) and is *indispensable* in recognising the possibility of underlying cognitive impairment: if the clinician does not recognise impairment as being possible then no further tests will be done.

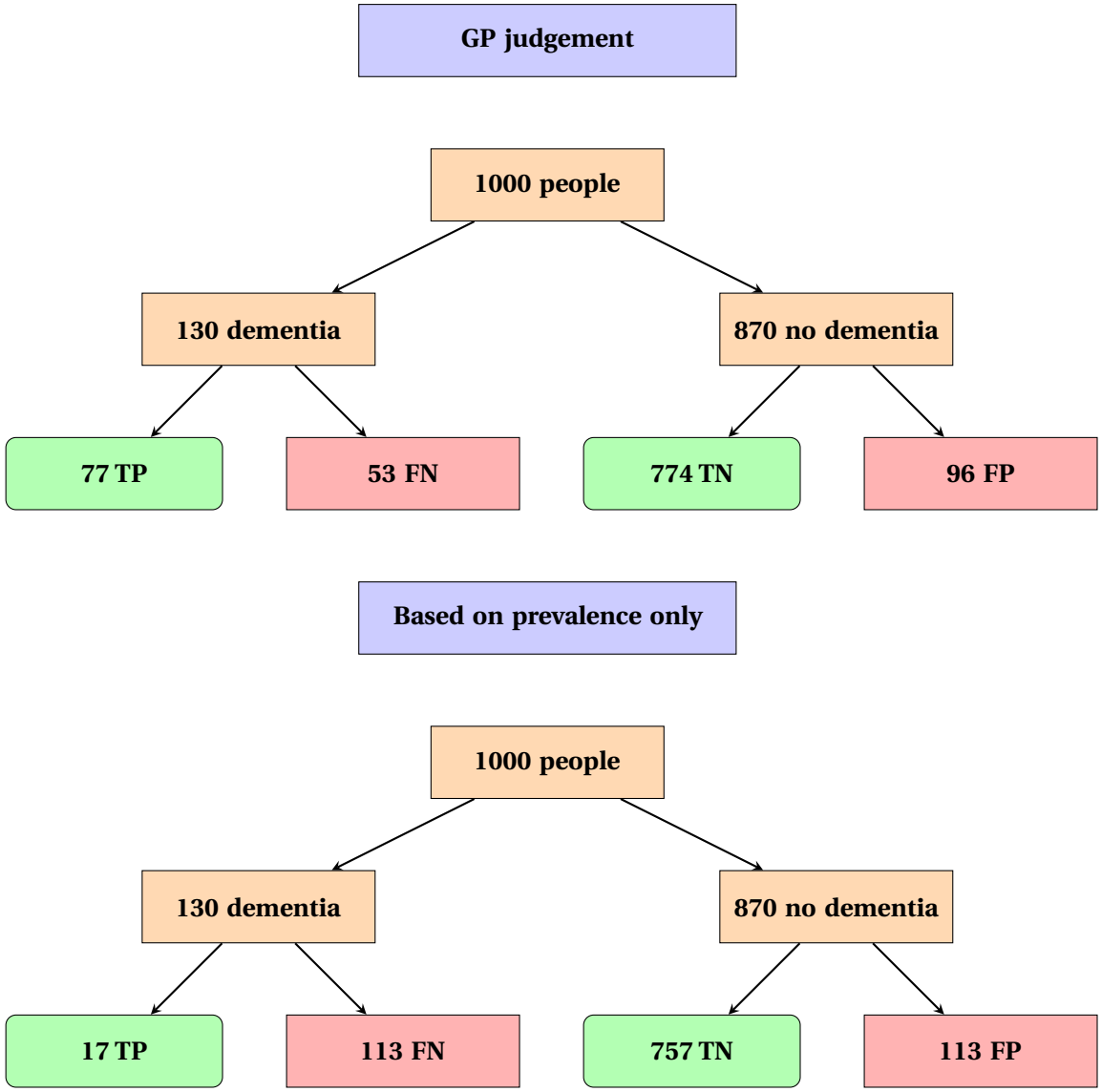
2.5.4 Conclusion and implications

One interpretation of these findings is that the clinical judgement of a GP is of limited practical value because in a low prevalence setting there is a high probability that person without disease will be labelled as disease free if people are sorted randomly into categories based on the prior probability of disease (prevalence). Figure 2.16 shows that if disease state was allocated randomly based on prevalence then in a group of 1000 people of whom 130 had dementia (see Figure 1.2), 77 *additional* people would be given an incorrect diagnosis, of whom 60 would be additional false negatives (to those mis-assigned by GP clinical judgement) and 17 would be additional false positives. Clinical judgement could be expected to correctly identify 60 more people of the 130 with disease and 17 more people of the 870 with no disease than categorisation based on prevalence alone. It should be noted that these figures are calculated on the basis of the point estimates, which as emphasised previously do not reflect the heterogeneity in the data, and which may be over-simplistic as a summary of the diagnostic accuracy of clinical judgement.

Natural frequencies for tests and judgement Some investigators have argued that tests should be used to support the identification of people with dementia [306], though this is debated [307]. Using a brief cognitive test that is completed by the patient (such as the MMSE at a threshold of 25) would result in 104 TPs (27 more than clinical judgement), but at the cost of 209 FPs (113 more than clinical judgement); meaning that compared to clinical judgement alone, using the MMSE at a threshold of 25 would result in four additional disease free people undergoing further workup or being misdiagnosed for every one additional person with disease who was correctly identified ($\frac{113}{27}$). In contrast asking an informant to complete the IQCODE, at a threshold of 3.5, would result in 114 TPs and 78 FPs, resulting in two fewer diseased free people being identified for every 1 additional person who was identified with dementia ($\frac{37}{-18}$), compared to clinical judgement. Therefore if it is a goal to identify people with dementia who are missed by their GP it may be preferable to use informant measures rather than patient measures. It can be anticipated that this may cause some difficulties for people who are socially isolated.

From a public heath perspective an important implication is that in a population of 1000 people of whom 130 have dementia, a GP could be anticipated to classify 828 as not having dementia, and of these 94% will not have dementia. In contrast, if people were randomly assigned to a diagnosis on

Figure 2.16: Natural frequencies of clinical judgement and random allocation of disease state based on prevalence



Based on 13% prevalence of dementia ($\frac{765}{5847}$ people had dementia; Figure 2.4 see also Figure 1.2), clinical judgement 64% sensitivity 94% specificity. Prevalence only, e.g. 17 **TP** based on 13% *130

TP true positive **FN** false negative **TN** true negative **FP** false positive

the basis of the disease prevalence then 870 people would be classified as free of dementia, of whom 87% would actually not have dementia; of the 130 people who were classed as having dementia only 17 would have the disease.

A question for health systems that deserves consideration is what is the threshold probability of dementia at which specialist evaluation is warranted. In the United Kingdom GPs are encouraged to refer for evaluation to exclude cancer if the probability of cancer exceeds 3% in an adult and 1% in a child. Compared to cancer, dementia is also a serious condition with important implications.

Implications for future studies An important unanswered question from this review is what is the accuracy of GP clinical judgement for the diagnosis of dementia in people with symptoms compared to those without. Most studies in this review did not distinguish between people attending their GP with symptoms of dementia and those attending their GP for other reasons. This is especially important because the prevalence of dementia in people attending their GP with symptoms can be anticipated to be higher than the 13% calculated in this review, as well as also altering the spectrum of disease in people who do have dementia (being more severe in people who are presenting with symptoms). Overall the estimates of diagnostic accuracy in this review are likely an underestimate of diagnostic accuracy in people with symptoms.

There are least two further important unanswered question from this review. Firstly, what is the comparative accuracy of clinical judgement compared to brief cognitive tests [308] for diagnosing dementia. Direct comparison of test accuracy in a single study is preferable to indirect comparisons but requires additional methodological considerations to ensure robust study design. Secondly, what is the evidence for heterogeneity in the aspects of study design that were investigated in this review? To investigate heterogeneity in test accuracy further, future studies could attempt to apply more than one definition, could elicit retrospective judgement in a randomly selected sample of people who had already had prospective GP clinical judgement and who were intended to receive the reference standard assessment (but had not yet), could blind some (randomly assigned) participating GPs to using the medical records for the retrospective judgement and others to not, and could take steps to attempt full verification of the reference standard and provide extended follow-up of the medical records for those people who were not verified.

Dementia is an uncommon disease in general practice and it has been projected that most GPs could expect one to two new cases each year, per physician [309]. If a GP thinks a patient has dementia then there may be merit in using a brief cognitive test to measure objective cognitive impairment because the probability of dementia based on GP clinical judgement alone is not high enough to confirm dementia. It is also important to objectively quantify cognitive functioning in someone with *symptoms* of dementia who is judged by their GP to be disease free, because the sensitivity of clinical judgement is too low to definitively exclude cognitive problems.

Chapter Summary

This Chapter has presented the methods and results for a review of diagnostic test accuracy of GP clinical judgement for two target conditions: dementia and cognitive impairment. Clinical judgement was defined as *being unaided by any additional test, investigation or inquiry beyond that which is immediately available to the clinician.*

A search was conducted in five electronic databases by an information specialist without language restrictions. The search yielded 12,681 citations, of which 8118 remained after de-duplication and were independently screened by two reviewers. Full text records were reviewed for 56 papers and of these 16 were included, referring to 10 studies of which 7 could be included in the meta-analysis. Data extraction and quality appraisal was done in duplicate and disagreements were resolved by discussion. Studies were classified by whether the GP judgement of the target condition was prospective: done after seeing a patient, or retrospective: made in hindsight by reflecting on their past encounters with the patient.

QUADAS-2 was used to appraise the quality of included studies, which were generally at low risk of bias, with the exception of the flow and timing domain. For dementia as the target condition the sensitivity of GP judgement ranged from 34% (95% CI 18% to 54%) for Cambridge to 91% (95% CI 85% to 96%) for Mannheim; and specificity ranged from 58% (95% CI 51% to 66%) for France to 99% (95% CI 97% to 100%) for Zwolle.

There was was substantial heterogeneity between studies. Uncertainty in the estimates prevented firm conclusions about how differences in study design affected heterogeneity in diagnostic accuracy.

CHAPTER 3

METHODS

This Chapter details the methods for an empirical quantitative study to investigate the accuracy of tests for diagnosing dementia in people presenting to GPs with cognitive symptoms. The first Section defines the target conditions and the research questions. The second Section describes the recruitment of participants, including exclusion criteria, the setting, context, sampling and referral to the study. The third Section describes administrative arrangements including details of how referrals to the study were processed, how research clinic appointments were arranged, and other standard operating procedures. The fourth Section describes the test methods and data collection for tests, including the rationale for index test selection, an overview of the tests in the index battery, the specialist evaluation, and the collection of follow-up data after the research clinic. The fifth Section provides details of the reference standard and describes how the medical record extract was reviewed. The sixth Section describes data collection, including locations of research clinics, and revisions to the case report form, and data management. The seventh Section describes the statistical methods. The eighth Section reviews the methods and considers the risk of bias using the QUADAS-2 tool.

3.1 Target conditions and research questions

3.1.1 Target condition definition

The two target conditions for evaluating diagnostic accuracy were dementia (of any aetiology) and normal. Dementia was chosen as a target condition because tests with high accuracy for diagnosing this condition would allow GPs to make a positive confirmatory diagnosis of dementia in people with manifest impairment. If GPs were able to make a positive diagnosis of dementia this might avoid referral to specialist services unless there were features which required expert input to manage, or

there were other factors complicating the presentation such as young age, atypical symptoms, or rapid onset and progression. Normal cognition was chosen as a target condition because tests with a high accuracy for diagnosing this would allow GPs to identify people with *no* current impairment, who could potentially be reassured. People who GPs could not make a positive diagnosis of either dementia or normal cognition would likely require a individualised approach including options as outlined in Section 1.1.2.1, such as test of time, further tests, or referral to specialist; these people may have MCI, dementia or normal cognition.

These target conditions were chosen as being most relevant to general practice because general practitioners are unlikely to have sufficient expertise to differentiate specific aetiologies and because most patients with cognitive impairment in primary care will have a composite aetiology (Section 1.1.1.3). Dementia was defined according to ICD-10 criteria.

3.1.2 Quantitative research questions

The research questions for the quantitative study were:

- What is the accuracy of clinical judgement of general practitioners for diagnosing the target conditions in people with cognitive symptoms?
- What is the accuracy of a range of tests for diagnosing the target conditions in people presenting to general practice with cognitive symptoms?
- Which combinations of tests have the greatest net benefit to diagnose the target conditions?

3.2 Participants

Participants were people with symptoms of possible dementia, who were aged at least 70 years, and had been referred by their GP surgery during the recruitment period. People who had already been diagnosed with dementia were not eligible. In contrast, people who had been previously evaluated for cognitive symptoms in the past were eligible so long as there was a persisting genuine concern about the possibility of dementia. Symptoms of dementia are detailed in Section 1.1.1.4. The concern about symptoms of dementia was permitted to be from the person themselves, their kin, a health professional including their GP, or another person. There was no eligibility threshold for the extent of concern about dementia, but people were required to have had symptoms for at least six months and be able to attend with an informant, because these aspects of the history facilitated a robust reference standard. People with symptoms that were progressing every week, or those with neurological symptoms that were co-incident with the cognitive impairment were not eligible, because these were considered to be clinical red flags for rare but important neurodegenerative disease such as prion disorder. Table 3.1 shows the medical comorbidities that were exclusion criteria to avoid including people with complicated presentations that were judged likely to require specialist evaluation in all cases. People with very severe dementia, operationalised as inability to consent to participate, were

excluded. People with very severe dementia were judged, in consultation with a lay advisory panel, to find participating in research burdensome, and have difficulty in completing formal evaluations of cognition. Transport was provided free of charge to participants on request, either with a taxi or with a wheelchair adapted minibus. People were not excluded if they were resident in a nursing home, older persons home, or supported housing, so long as they were fit enough to travel out to an appointment. Translation services were offered to all participants and arranged on request, so that inability to communicate in English did not exclude people.

Table 3.1: Medical comorbidities that were exclusion criteria

Prior diagnosis of a parkinsonian condition ∗
Multiple sclerosis
Learning disability
Motor neuron disease
Huntington's disease
Registered blind
Severe hearing impairment †
∗ including Parkinson's disease
† operationalised as unable to use telephone

3.2.1 Setting

Participants were recruited from GP practices in the Bristol, North Somerset, and South Gloucestershire area (BNSSG) area. BNSSG is an diverse geographical area within approximately 15 miles of the City of Bristol, that had a total population of 900,000 people across 82 general practices. Practices in Bristol were predominantly in an urban catchment area, whereas those in South Gloucestershire and Somerset included patients in suburban and semi-rural areas. Practices in the West of England Clinical Research Network were invited to participate through direct contact, regular circulars and networking events. All practices who expressed interest before the deadline were included in the study but the practice start dates for recruitment were staggered for operational and logistical reasons. Practices that were identified by the Clinical Research Network as having particular expertise in recruiting to studies were selected to an early start date. Table 3.2 shows the details of recruiting practices. The median list size of practices was 11,333 (interquartile range 8004 to 14,654) and list size ranged from 4,224 (Stoke Bishop (L81622)) to 21,588 (Mendip Vale (L81086)). The practice with the highest proportion of patients in the eligible age group out of the total list size was Clevedon (L81102), in a relatively affluent town outside of Bristol, which had 1,383 of 7,229 patients eligible (19%). The practice with the lowest proportion of patients in the eligible age group was Redfield (L81061), in a relatively deprived inner city ward, with 524 of 8431 (6%) being eligible. Recruitment began in March 2015 and is detailed further in Section 3.2.4.

Table 3.2: Details of recruiting Practices [310]

Location ∗ **(code †)**	**List size**	**List at least 70 years**	**Set up date**	**Region ‡**
Backwell (L81060)	13,500	2,378	December 2015	N.S
Bedminster (L81053)	8,004	1,160	February 2015	Bristol
Clevedon (L81102)	7,229	1,383	September 2016	N.S
Clevedon (L81040)	16,094	2,701	April 2015	N.S
Clifton (L81081)	12,965	1,042	March 2016	Bristol
Close Farm (L81050)	7,102	803	December 2015	S.G
Fishponds (L81087)	10,644	1,024	April 2015	S.G
Frampton Cotterell (L81014)	14,654	2,553	February 2016	S.G
Hanham (L81079)	21,327	3,343	February 2015	S.G
Henleaze (L81131)	9,120	1,501	October 2015	Bristol
Horfield (L81022)	15,474	1,133	March 2015	Bristol
Kingswood (L81063)	12,062	1,632	May 2015	S.G
Long Ashton (L81056)	6,900	939	May 2016	N.S
Mendip Vale (L81086)	21,588	3,486	April 2015	N.S
Portishead (L81004)	18,285	3,235	February 2016	N.S
Redfield (L81061)	8,431	524	February 2016	Bristol
Shirehampton (L81008)	11,333	1,228	March 2015	Bristol
Southmead (L81098)	7,423	951	December 2015	Bristol
Stoke Bishop (L81622)	4,224	366	April 2015	Bristol
Westbury on Trym (L81017)	9,772	1,375	March 2015	Bristol
Yate (L81047)	13,320	2,199	January 2016	S.G
Total	249,556	34,956		

∗ Location is given rather than practice name because practice names change with mergers and change of doctors
† Practice code is a unique identifier held by NHS England
‡ *N.S* North Somerset *S.G* South Gloucestershire

3.2.2 Context

GP practices in the UK were funded through a complex set of arrangements typically including an annually negotiated national contract, the Quality and Outcomes Framework, public health programmes, and additional local or direct (i.e. national) enhanced services to support activity that was particularly important to the government of the day, along with other NHS (National Health Service) and non-NHS activities [311]. Total NHS funding received per patient was typically in the order of £150 per patient per year [312]. Between April 2013 and April 2016 many GP practices (81% at one point, [313]) participated in a direct enhanced service under which they received £0.37 per patient on their list and a further payment based on the number of assessments the practice had performed out of the national total, in return they were asked to *"a) identify patients at clinical risk of dementia; b) offer an assessment to detect for possible signs of dementia in those at risk; c) offer a*

referral for diagnosis where dementia is suspected; and, d) support the health and well-being of carers for patients diagnosed with dementia." [314]. In addition, since July 2013 GP practices in Bristol, but not South Gloucestershire or North Somerset, were able to participate in a separate programme under which they were paid £500 to cover set up costs, plus £164 for each new diagnosis of dementia, in return they were asked to [315]:

- *"Adopt the shared care pathway including management of people stable on dementia medication.*
- *Undertake a diagnosis of uncomplicated dementia (Alzheimer's Disease or Vascular Dementia) within a Primary Care setting (using the agreed template) and provide appropriate post diagnostic support and signposting information*
- *Carry out enhanced reviews of people with dementia and their family/carer (using the agreed template) that delivers review of all medication including cholinesterase inhibitors, memantine and anti- psychotic medication.*
- *Undertake investigations [as detailed] and investigate any abnormalities to exclude potentially treatable causes*
- *Undertake a diagnosis of dementia and initiate medication in line with guidance*
- *Complete a plan for the patient that includes relevant information including where to go for further support and signposting*
- *Note the diagnosis of dementia, if made in secondary care and record accordingly with relevant read code.*
- *Review stable cases of people with dementia, currently seen in secondary care (payment £40 per review)*
- *Review every person diagnosed with dementia at least once a year (6 monthly if on dementia related medication, 3 monthly if on anti- psychotic medication), following the review template provided*
- *Continue the prescribing of Cholinesterase Inhibitor or memantine*
- *Notify the Memory Nurse of any adverse drug reactions, deterioration in condition or any other clinical concerns regarding the person's health that can not be managed in Primary Care*
- *In order to qualify for payment the practice must complete the work detailed above. The memory nurse for the locality will be able to provide support, advice and guidance. The memory nurse will be able to carry out joint home visits with the GP/nurse. If the memory nurse carries out a home visit on their own, the practice will not be eligible to claim payment.*

- *Practices will receive a bonus payment of £200 if they achieve a 5% increase in the number of people on the practice register with a diagnosis of dementia, in year, or if they achieve 65% of expected diagnosis against expected prevalence."*

The clinical commissioners were consulted about the programme of quantitative research, and advised that GPs would still be able to participate in the funded work outlined above, as well as the research study. The commissioner requested that all clinical decision making regarding diagnosis, explanation to the patient, investigations, onward referral, and interventions be done by the GP surgery rather than the research clinic; the research team concurred that this approach would be appropriate. The potential impact of these programmes of funded activity on the setting is explored in the Discussion.

3.2.3 Sampling

GPs were encouraged to refer a consecutive series of eligible patients with no need to perform any form of prior testing. Table 3.3 shows details of barriers to referral that were identified and mitigated. To encourage recruitment a computer prompt was used at the point of care to remind GPs about the study. GP practices in BNSSG all used the EMIS[1] Web [316] clinical records software, which stored details of patients medications, investigations, documents, appointments and consultations electronically. EMIS Web allowed for the entry of both free text and coded data. Software in participating GP practices was programmed so that when GPs entered a problem code regarding memory loss, confusion, or cognitive decline, or synonyms for these things in the consultation page, a prompt reminded them about the research study. Table 3.4 gives details of the full list of codes that triggered the electronic point of care prompt. Some clinical problems (e.g. confusion) have more than one code, so all variants were included in the list of trigger codes. To avoid GPs being prompted to refer people who had already been diagnosed with dementia, codes for dementia were specifically not included in the list of trigger codes.

The electronic prompt allowed the GP to directly open the study referral form. This was designed to facilitate discussion with the patient and make the referral process easy. Alternatively the GP could cancel the prompt and the reason for the prompt being cancelled was recorded. This process was tested in five GP practices and the feedback from GPs was very positive; there was only one change suggested which was to change the code that was used to record the use of the prompt.

3.2.4 Referral to the study

The referring GP checked eligibility, gained the consent of the symptomatic person to refer them to the research study, and completed a referral form providing details on the patient and their contact details. The referral form also asked who had concerns about cognitive symptoms (the patient, their kin, the GP, another professional, someone else); the GP's clinical judgement about the extent of

[1] EMIS originally stood for Egton Medical Information Systems, but EMIS is now the name of the company

Table 3.3: Methods to mitigate potential barriers to recruitment

Barrier	**Mitigation**
GP not remembering the study	Computer prompt
GP unsure about eligibility criteria	GPs encouraged to be inclusive, study staff recheck and confirm eligibility
Difficulty contacting participant to arrange appointment	Multiple call backs ∗
Patient difficulty in accessing research clinic	Offer transport, range of days and times
Acute illness in participants	Allow people to rearrange appointment
Participants forgetting appointment	Reminder telephone call in week of appointment
Uncertainty about participation	Allow people time to think and call back

∗ at least four attempted calls per person at different times and days

cognitive problems (operationalised as normal, CIND (cognitive impairment, but not dementia), dementia); the confidence of the GP in this opinion on a 10cm visual analogue scale; the rationale of the GP in forming their clinical judgement. The referral form reminded the GP about the local guidelines for investigating possible dementia, but the GP was not required to arrange any tests *exclusively* for the research study. GPs were encouraged to try to continue with their routine clinical practice and not let referral to the study influence their decisions; for instance GPs were not prevented from referring people to the NHS memory service simultaneously or subsequently to the person attending the research study.

3.3 Administrative arrangements

GPs were asked to send referrals to the research study team by secure nhs.net email, which is approved for patient identifiable communication within the NHS. The nhs.net email account was reviewed regularly to identify new referrals and issues relating to existing referrals. The study administration team followed a standard operating procedure (SOP) to review the referral form and contact potential participants. If there was any missing information on the referral form then the GP practice was sent a template letter which identified and requested the missing information, and the referral was added to a tracking sheet for follow-up. Missing data was followed up with at least three contacts to the practice for all referrals, regardless of whether the person was subsequently recruited to the study. If there were no contact details for the referred person then the practice was asked to supply these urgently. When contact details were provided the referred person was contacted on at least three occasions on at least two different parts of the day. If no contact had been made within a week of the referral the GP practice was advised that it had not been possible to contact the person.

The study administration team were given a SOP and script for the initial conversation with the

Table 3.4: Codes to trigger study prompt

Code	Clinical term
2841	Confused
1B1A-2	Memory loss symptom
1B1A-3	Memory disturbance
1B1A1	Short term memory loss
2841-1	Confusion
28E	Cognitive decline
28E0	Mild cognitive impairment
28E3	Cognitive impairment
311B	Cognitive assessment
38C15	Initial memory assessment
3A-1	Memory assessment
E030-2	Toxic confusional state
E042	Chronic confusional state
E2A10	Mild memory disturbance
E2A11	Mild memory disturbance
EMISCCO13	Cognition NOS
EMISCCO2	Cognition
EMISNQIM12	Impaired Cognition
R009	[D] confusion
R009-1	[D] Senile confusion

the clinical term may have more than one code

person about the study. This confirmed the details of the referred person (name, date of birth, address) and then confirmed their eligibility for the study. When the referred person had been confirmed to be eligible for the study, details of the study were provided and the administration team discussed details of options for appointments. Consent forms and patient information leaflets were sent to participants prior to their attendance at the clinic. The administration team offered to book transport or translation services for people, and provided the study email address and mobile phone details. A time was agreed for a reminder telephone call three days before the appointment.

The initial conversation SOP also provided details on the action to take if the referred person reported information that implied that they were not eligible for the study. The administration team were instructed to only resolve someone as ineligible if the statement could be confirmed, with consent, from a second party (e.g. the referral form, a spouse, or relative). For instance, if the person claimed to have no problem with their memory this was was not taken at face value and the person was only judged to be ineligible if this was verified by either the referral form or a second person. If the person wanted more time to decide whether to take part then a follow up telephone call was scheduled with the administrative team.

A set of SOP were provided in addition to the SOP for the initial telephone call. The SOP are provided in full in the Appendices. Administrative SOP before the research clinic appointment

are provided in Appendix B. These included how to deal with missing information in the referral form, how to book a research clinic, template emails and correspondence for arranging transport and communicating with GP surgeries, how to deal with issues regarding appointments, how to address the matter of there being apparently no cognitive concerns, and how to handle a cancelled appointment. SOP for use on the day of the research clinic are provided in Appendix C, these address a checklist of matters to be done on the day, assessing capacity and taking consent; how to administer the index tests; how to handle fire, safeguarding and serious adverse events; and how to deal with participant withdrawal. SOP for after the research clinic appointment are provided in Appendix D, these cover how to deal with requests for a copy of the notes taken at the research clinic and how to scan the case report forms securely.

3.3.1 Ethical approval

The empirical research received a favourable ethical opinion from the National Research Ethics Service Committee London - Bromley (reference 14/LO/2025) on 26 November 2014. An amendment, which allowed for the qualitative interviews to take place, received a favourable opinion on 04 May 2015.

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. Staff employed by the West of England Clinical Research Network had honorary NHS contracts and all other researchers (including the author and the specialists) accessing NHS data and patients had a substantive contract directly with the NHS for a clinical role. The University of Bristol was the Sponsor for the study and indemnified the empirical work.

3.4 Test methods

Each participant underwent a set of index tests (Section 3.4.1) and a specialist assessment (Section 3.4.2) on the same day, at the same appointment. Staff from the Clinical Research Network visited each practice six months after the research clinic to take an extract of the medical records prior and subsequent to the research clinic and this is described in Section 3.4.3.

At the research clinic two examiners undertook a separate blinded evaluation of each participant, one who performed the index test battery, and one who performed the specialist assessment. The examiners were not aware of any information about the participants other than that which was elicited during the assessment. Neither examiner was aware of the referring GPs opinion on the participants cognitive status, and furthermore the index test examiner was not aware of the details of the clinical history (symptoms, duration, who had reported problems) because these were not elicited in the encounter. Efforts were made to put participants at ease on their arrival at the research clinic. Signs were displayed prominently at the research site to indicate where the clinic was happening, and GP reception staff were briefed by the practice manager and by researchers. Participants were

welcomed as soon as possible after their arrival by the research team, and taken to the rooms that the research clinic was operating from. The researcher assessed the capacity of each participant by asking them their reason for coming to the practice, and checking that they were willing to be there. Once participants had been determined to have capacity they were offered the opportunity to discuss the study, and then asked to sign a form to record their consent to participate. Participants were randomly assigned to receive the index test battery first or the specialist assessment first, and were offered refreshments in an interval between the two assessments.

Participants received a short debrief at the end of the research clinic where they were advised to contact their GP in a few weeks to discuss the next steps but not informed of any diagnosis. There were two reasons for this, firstly there were concerns that test results that would influence a diagnosis would not be available at the research clinic due to delays in the clinical pathway or difficulties with information governance. Indeed, tests were uniformly not available to the specialist at the research clinic so that the specialist was blinded to investigation results, forcing them to make a decision on the basis of their clinical judgement. Secondly it was agreed with the host clinical commissioning groups that contractual requirements (Section 3.2.1) required the general practitioner to make and convey the diagnosis, accounting for the clinical judgement of the specialist assessment and diagnostic tests. If the research team identified any safeguarding concerns the participants were advised to discuss these the same day with the referring GP practice.

3.4.1 Index tests

Tests for the index test assessment were identified after a review of the literature and a guide for clinicians regarding tests that could be used in the evaluation of possible cognitive disorder [317]. Tests were selected for the index test battery on the basis of the following criteria:

1. Available to use for free – i.e. not copyright (therefore mini mental state examination [156] excluded);
2. Previously evaluated in a primary care setting in at least one study;
3. Not been evaluated in primary care before but conceptually of interest (e.g. Sniffin' sticks);
4. Good diagnostic accuracy, determined as Youden index [149] of greater than 0.75, or a sensitivity or specificity of greater than 0.85 at the optimal reported threshold;
5. The diagnostic accuracy was weighed with the burden and time needed to complete the test, with brief tests with high PPV being preferred.

Table 3.5 shows the tests that were selected for inclusion in the index assessment. A range of cognitive domains were examined. Where possible, items were not repeated, e.g. the 6CIT and the GPCOG both require the recall of a 5-item name and address, and to avoid burdening and potentially confusing participants this item was done once and then scored separately for each test.

Table 3.5: Tests in the index battery

Index test *Abbreviation* ∗	**Items**
Memory alteration test [318] *MAT*	40 items Encoding; Immediate Recall; Orientation; Semantic memory; Free and cued recall
Eurotest [319]	11 items Knowledge; Calculation; Verbal fluency; Delayed recall
Phototest [320]	14 items Name pictured items; Verbal fluency; Free and cued recall
Scenery Picture Memory Test [321] *SPMT*	23 items Free recall of 23 items in a picture; Cued recall of 3 items
Six Item Cognitive Impairment Test [258] *6CIT*	6 items Orientation; Delayed recall
General Practitioner Assessment of Cognition [301] *GPCOG*	5 items Orientation; Mark hands and time on clock; Free recall; Informant section
Mini-Cog [293]	2 items Clock draw and recall
Time and change (TAC) [322] *TAC*	2 items Telling time on a printed clock; Making change from coins
Timed up-and-go [323] *TAC*	3 trials Stand, walk 3 meters, turn and sit down
Extra pyramidal signs scale [324] *EPSS*	7 items Standardised evaluation for extra pyramidal signs
Sniffin' sticks [325, 326]	12 items Identify a smell in a standardised pen from a choice of 4
Pfeffer FAQ [327] *FAQ*	10 items Informant reported status of instrumental ADLs †
Lawton IADL [328] *IADL*	8 items Informant reported status of instrumental ADLs †
Katz ADL [329] *ADL*	6 items Informant reported status of independence in ADLs †
AD8 [299]	8 items Informant report of symptoms over last "several years"
IQCODE ‡ short version [304]	15 items Informant report of symptoms over last 10 years

∗ Abbreviation, where appropriate
† *ADLs* Activities of daily living
‡ *IQCODE* Short Form of the Informant Questionnaire on Cognitive Decline in the Elderly

During the process of selecting tests for inclusion in the index battery, a review was published [330] regarding the use of cognitive tests in primary care. Compared to the tests identified by that group the index battery did not include the memory impairment screen (MIS) [331] or the abbreviated mental test (AMT) [332]. The MIS had all rights reserved and was therefore not eligible for the index battery, but the index battery did include indicators which reflect similar aspects of cognitive testing. Under evaluation with MIS candidates are asked to read from a list of four words, then engaged in a distractor activity, and then scored on free and cued recall; in comparison when under examination by Phototest (which was available for use under a creative commons license, and included in the index battery) participants are asked to identify six photos, perform a distractor task, and then scored on free and cued recall. Of the 10 items in the AMT, six were included in the index battery, whereas age, recognition of two people, year of First World War, and name of present monarch were not included. Arguably these four items of the AMT which were not included in the index battery are less discriminative in people with mild impairment (age, recognition), or are culturally determined (war, monarch).

The Montreal Cognitive Assessment (MOCA [333]) was initially not included in the index battery as it was originally designed to diagnose or identify MCI, had been advocated for use in secondary care [317] and had not been investigated in primary care [334]. However, the protocol was revised in light of subsequent policy changes in 2015 that encouraged GPs to diagnose dementia in typical situations without referring to a specialist [335] using the MOCA as the preferred instrument. The Memory alteration test (MAT) was replaced with the MOCA because it was judged that including both the MOCA and the MAT would be overly burdensome for participants and have little added value. The MOCA is a 22 item test that is scored out of 30 and evaluates visuospatial skills, naming, memory, attention, language, abstraction, delayed recall and orientation.

Sniffin sticks had to be imported on special order and so were added at a later stage so as to avoid delaying the start of recruitment while waiting for this single test.

All index test assessments were conducted by the same examiner, the author: a male medical doctor who had obtained Membership of the Royal College of Physicians (UK) and who was completing post-graduate training in general practice. Excluding Sniffin Sticks, the order of the index tests in the battery was randomly assigned for each participant so as to avoid the effect of order influencing test accuracy, results of the randomisation process are provided in Appendix K. Each participant was offered the chance to undertake every test in the battery, but the examiner was responsive to the participants and if they appeared to be becoming tired or distressed then they were asked if they wanted to rest or indeed skip a question. Index tests were conducted as instructed by the original authors. The assessment was piloted to ensure it was not overly burdensome and took 25 minutes with a healthy person and 50 minutes in a person with cognitive impairment.

3.4.2 Specialist assessment

Each specialist assessment was conducted by the same examiner: a female medical doctor who had more than 20 years specialist expertise in the field of dementia. A standardised clinical evaluation was done for each participant, lasting approximately an hour and comprising clinical history, the Addenbrooke's Cognitive Examination III (ACE3) [336], Brief Assessment Schedule Depression Cards (BASDEC) [337] and the Bristol activities of daily living questionnaire (BADL) [338]. The specialist did not have access to any investigation results. If the specialist judged that further investigations and assessment were necessary to exclude a rare dementia aetiology the referring GP was advised to consider referring the person to the NHS memory service. The specialist was asked to reach a clinical judgement about the cognitive status of participants operationalised as normal, cognitive impairment, or dementia, as well as the most likely aetiology of the dementia based on the information available to them.

3.4.3 Follow-up

Staff from the West of England Clinical Research Network visited participating GP practices and extracted a limited set of data from the EMIS Web medical record of consenting participants. The data comprised consultation history, documents, and investigations, in the period six months before to six months after the research clinic appointment. Consultation history was exported to word processing software, electronically redacted, exported as a secure PDF, and emailed by secure nhs.net email to University staff in a separate file for each participant. Documents and investigations that could not be sent electronically were printed at the participating GP practice, labelled with the study identifier, redacted of identifying information, transported to secure storage at the University of Bristol in a locked folder, scanned onto secure University of Bristol servers, and then securely destroyed.

3.5 Reference standard

The original intention was that the reference standard would be a consensus diagnosis by three independent consultants: a neurologist, a geriatrician, and an old age psychiatrist. Procedures were developed so that the consultants would be able to remotely review the information from the research clinic that had been gathered during the specialist assessment. An electronic case report form was designed, piloted and refined to facilitate the data capture of the consensus review. Investigation results, including laboratory tests and neuro-imaging, were to be reviewed by the consensus review group.

Unfortunately, due to circumstances beyond the control of the investigators it was not possible to complete the consensus review prior to writing the thesis. Therefore the reference standard was based on the clinical opinion of the specialist clinician at the research clinic, who reached a decision about the clinical diagnosis of dementia syndrome according to ICD-10 criteria or MCI according to Peterson criteria [93]. The specialist clinician based their opinion on the information that was

available to them at the clinic, as detailed in Section 3.4.2. The specialist at the research clinic was not forced to make only one judgement about the level of impairment (normal, MCI, dementia), because discriminating borderline cases (MCI or mild dementia; MCI or normal) is especially difficult and the original plan had been that the specialist panel would adjudicate these cases with the benefit of follow up information and test results. Therefore, for each participant, there were two possible outcomes of the specialist evaluation at the research clinic with respect to the level of impairment:

1. **The specialist assigned one and only one category**. For these cases the process outlined in Section 3.5.1 was followed.

2. **The specialist assigned no category, or more than one category**. These cases were reviewed by a second independent specialist to allow a single impairment category to be assigned. The review was of all information pertinent to the case, excepting index test data, as had been originally planned, but this was done by a single specialist only rather than three. The specialist formulated their best judgement on the basis of the available information and this was taken as the reference standard.

3.5.1 Review of medical record extract

All available notes needed to be reviewed because the original intention was that these would be reviewed by the specialist panel and there was the possibility that the notes review might yield information that would contradict the specialist assessment. Therefore the intention of this process was to establish if there any evidence in the extract of the medical records that would contradict the impairment category assigned by the specialist evaluation, according to ICD-10 definition. The extract of the medical records was reviewed by the investigator and discussed with advisers. The following components were judged in advance to contradict the specialist opinion:

1. Alternative diagnosis assigned by a secondary care specialist, as an explanation for the symptoms,

2. Problem coded by a primary care specialist after the research clinic and adjudicated by two research General Practitioners as explaining the symptoms,

3. Abnormal CT imaging incompatible with diagnosis (i.e. not reported as changes associated with age, vascular damage, or typical processes associated with dementia). CT imaging findings that were possibly incompatible with the diagnosis were discussed with a specialist.

An alternative diagnosis (item 1) was only taken as contradicting the specialist opinion if it was an alternative disease process other than a dementia, e.g. cancer. A change to the impairment category, whether downgrading (i.e. from MCI to normal, dementia to MCI, or indeed dementia to normal) or upgrading (vice versa) was not accepted as contradicting the specialist opinion at the research clinic. The reason for this is that typically the grading of cognitive impairment categories hinges critically

on the judgement regarding the extent to which a person with cognitive symptoms is experiencing impairment in functioning in daily life. The view of the investigator, in discussion with advisers, was that the specialist clinician had formulated their judgement about the level of impairment on the basis of a detailed, systematic and structured evaluation, in the context of substantial experience and expertise. In contrast in BNSSG at the time of the research the NHS provided a range of evaluations for people with memory problems, including assessments by non-medical staff using the Mini-Addenbrooke's Cognitive Examination [339], assessments by GPs using a variety of brief cognitive tests [340], and evaluations by specialist doctors and psychologists in old age psychiatry or neurology clinics. Therefore there was substantial heterogeneity in the evaluation processes for people outside of the research setting, and arguably the determinations on the category of cognitive impairment were being made (in almost all cases) by a less expert person than at the research clinic.

The absence of items 1-3 when the medical records were reviewed was interpreted as indicating that there was no evidence to contradict the opinion of the specialist evaluation at the research clinic and therefore the specialist opinion at the research clinic was taken as the reference standard. Therefore there was one reference standard, expert specialist assessment according to ICD-10 criteria, based on two possible approaches:

1. Judgement by an expert based on the clinical assessment of a patient at a research clinic
2. For borderline cases, where it was not possible to reach a single determination about the cognitive impairment category at the research clinic, a single expert reviewed the information from the research clinic (excluding the index tests) together with the medical records extract from six months before the research clinic to six months after the research clinic.

3.6 Data collection and management

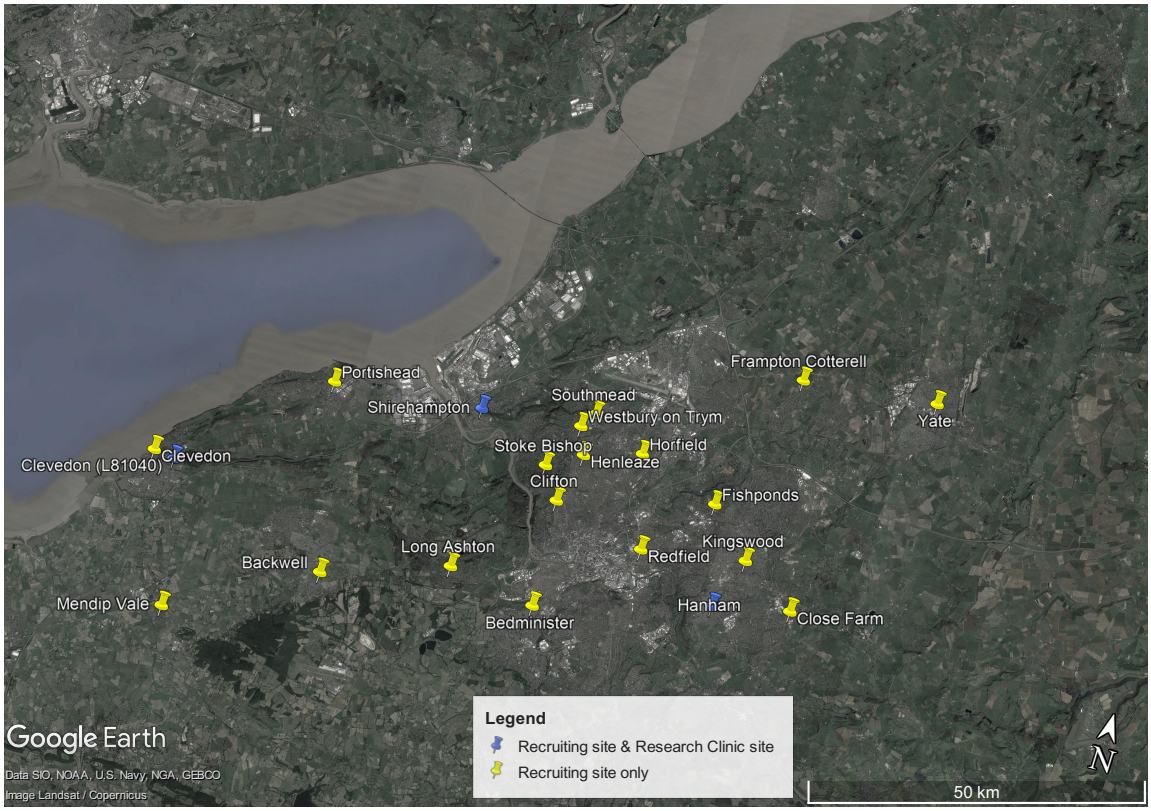
3.6.1 Research clinic locations

Research clinics were held at one of three participating GP practices: Clevedon (L81040), Hanham (L81079) or Shirehampton (L81008). Research clinic sites were chosen on the basis of accessibility to patients in recruiting practices across the Clinical Research Network, in practices that had adequate capacity of rooms for the clinic. Additionally a one-off research clinic was held at Frampton Cotterell (L81014) because this study site only agreed to participate on the basis that a research clinic would be held at that site. Figure 3.1 plots the location of the recruiting GP surgeries and the research clinic locations.

3.6.2 Case report form

A case report form was used to collect and record the data for each participant. The case report form was subject to four revisions because of changes to the set of index tests that were used during the course of the study. The revisions are described in Appendix E.

Figure 3.1: Map of recruiting sites and research clinics [341]



Recruiting sites: Clevedon (West), Shirehampton (Central), Hanham (East)

3.6.3 Data management

Study data were electronically entered and managed using REDCap (Research Electronic Data Capture [342, 343]) hosted at the University of Bristol. REDCap is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources. The electronic data capture form was initially tested on a mock case report form and then refined. The revisions are described in Appendix E.

Data were entered by four different data coders. Each coder was individually trained by the investigator with the session covering how to use REDCap and how to enter data; coders were encouraged to ask any questions and to ask the investigator how data should be handled on a case-by-case basis if they were unsure. All coders completed a test case report form which was checked, verified and discussed with the investigator before the coder entered real data. The number of cases

entered by each coder was determined by the duration of time that they were able to be committed and employed to enter study data. Table 3.6 shows that the number of cases entered by each coder ranged from 19 to 108.

Data were exported from REDCap, imported to Stata version 13 [215] and then tabulated and plotted to verify the quality of data entry. When data items were identified as missing or implausible the original case report forms were reviewed and values checked and re-entered in the REDCap database. To further quality assure the data entry process a 10% random sample of each case entered by a coder was selected for duplicate data entry. RStudio [344] was used to randomly select 10% of cases coded by each coder.

Table 3.6: Errors in coded cases, by coder

Coder	Coded cases (n)	Cases reviewed (n)	Number of coded items	Number of errors (%)
A	108	11	5423	5 (0.09%)
B	60	6	2958	1 (0.03%)
C	54	6	2886	7 (0.02%)
D	19	2	986	0 (0.00%)
Total	**241**	**25**	**12,253**	**13 (0.11%)**

For each of the cases that were selected for review, the original case report form was reviewed against the REDCap entry, data were manually checked and when incorrect, corrected and enumerated. The total case report form comprised 478 items (version 1), 479 items (version 2) and 493 items (version 3). Table 3.6 shows the number of items coded and number of errors by each coder in the 10% random sample.

3.7 Statistical methods

3.7.1 Characteristics of participants

Data were exported from REDCap to Stata and cleaned. Details on recruitment were tabulated by practice and list size. A summary was plotted by practice of people who were referred to the study by GPs and who participated, or who were ineligible, declined to participate, or were uncontactable. Characteristics of participants including age, sex, duration of symptoms, education, symptom pattern, and ACE3 score were tabulated by cognitive status according to the reference standard. Known characteristics of people who declined were tabulated against those who participated. Separate logistic regression analyses were used with declined as the dependent variable and GP judgement, age (in years), and female sex as the independent variables to test the hypothesis of no association between 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).

Separate linear regression analyses were used with cognition in categories (normal, MCI, dementia, other) as the independent variable and the dependent variables: age at clinic in years, total ACE3 score, ACE3 sub-domain scores (for attention, memory, fluency, language, and visuospatial domains) to test the hypothesis of no association between these variables. Total ACE3 score was plotted using box plots over the GP judgement, to illustrate the range of scores in each group, and compared to four exemplar tests, MOCA, AD8, Minicog and IQCODE (which were chosen because Cochrane reviews have been done on these tests). Standardised scores on each of the five ACE3 sub-domains were plotted using box plots over the cognitive categories, to explore the profile of performance between different groups and to illustrate the range of scores in each ACE3 sub-domain.

3.7.2 Characteristics of tests

The index tests were evaluated at the threshold indicated below, where •• indicates the threshold for dementia:

Clinician or informant completed	*Physical tests*	*Multi-domain tests*	*Brief tests*
GP judgement •• dementia	EPSS ••>1 [345]	Eurotest ••<21 [319]	TAC ••<2 [322]
FAQ ••>2 [346]	TUG ••>15 [255]	MAT ••<28 [318]	SPMT ••<10 [321]
ADL ••>1 [347]	Sniffin Sticks ••<11 [348]	MOCA ••<26 [333]	Phototest ••<27 [320]
IADL ••<5 [347]		GPCOG ∗	Minicog ••<3 [293]
AD8 ••>1 [349]		6CIT ••>7 [258]	
IQCODE ••>3.3 [350]			

∗ For GPCOG a two stage approach to scoring is used whereby scores >8 indicate normal and <5 indicate abnormal and scores 5-8 indicate GPCOGi needed, where scores <4 indicate abnormal [301]. All of the thresholds were pre-specified.

Characteristics of the index tests were tabulated by cognitive category according to the reference standard, with mean test scores and test duration compared to normal (with 95% 95% confidence intervals). Separate linear regression analyses with the number of tests that were declined by participants as the dependent variable were used to test the null hypothesis of no association between declined tests and the independent variables cognition (in categories: normal, MCI, dementia, other) and physical frailty (using TUG). Because the number of people who declined tests was small, logistic regression analyses with declined tests (none, ≥ 1) as the dependent variable and the independent variables cognition and physical frailty were also used.

Duration of each index test was summarised and tabulated, and plotted using box plots. Linear regression was used to explore the association between test duration as the dependent variable and cognition in categories as the independent variable to test the null hypothesis of no association between cognition and test duration. Some people may take much longer to complete tests than others, and operationally in practice it is important to know what length of appointment is needed for a person who is booked to take a test. Therefore average test duration was judged of less relevance

to clinicians and patients than the 95th centile of duration (C_{95}), the time that 95% of participants completed the test within. On this basis the tests were classified as follows; short tests (<5 minutes - or half of a typical GP consultation), medium tests (≥ 5 minutes but ≤ 10 minutes), and longer tests (> 10 minutes).

3.7.3 Informant tests

Five tests which asked questions of the informant of the person with cognitive problems were completed both as standalone index tests both by independent completion of a paper questionnaire and then by interview (by the GP researcher). These five tests were Functional Activities Questionnaire (FAQ [327]), Lawton instrumental activities of daily living scale (IADL [328]), Katz index of activities of daily living (ADL [329]), Galvin AD8 Dementia Screening Interview (AD8 [299]), and Short Form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE [304]). Weighted kappa was used to judge agreement between the different approaches to completing these five index tests. Weighted kappa was used to account for the fact that (for instance) a disagreement between categories 0 and 1 is less than the disagreement between categories 0 and 12. Weighted kappa was calculated for the whole instrument rather than individual items on the instrument because only the instrument as a whole would be used for diagnostic accuracy. The summary diagnostic accuracy of the informant measures was assessed using the DOR, and ROC curves were plotted to compare the accuracy of measures when completed by independent completion against those completed by interview.

3.7.4 Single test accuracy

For each of 17 index tests, scores were dichotomised at previously reported thresholds (Section 3.7.2), so that the results were either test positive or negative. For GP clinical judgement the threshold for the target condition dementia was a judgement of "dementia", whereas for the target condition normal the threshold was a judgement of not normal, including people GPs judged as having dementia and CIND (Section 3.2.4).

Measures of test accuracy are discussed in detail in Section 1.1.2.3. Accuracy (referred to hereafter as classification accuracy for clarity), defined as $\frac{TP+TN}{TP+TN+FP+FN}$ is dependent on prevalence of the disease. However, despite the association of classification accuracy with prevalence, when comparing tests under constant prevalence, as in this empirical study, relative classification accuracy of tests is of *some* relevance to patients and clinicians to the extent that it indicates the probability that a person will be correctly classified when that single test is used for diagnosis. This is of some importance clinically because in a low prevalence setting it is possible that a test could have perfect sensitivity but low classification accuracy, with a large proportion of people mis-classified and potentially needing unnecessary further testing. Classification accuracy was calculated for each index test.

Diagnostic accuracy of each single test was calculated using standard measures: sensitivity, specificity, PPV, NPV, LRP, LRN and DOR. The number of TP, TN, FP, and FN was plotted using

natural frequencies. Calibration and discrimination for the single tests were not considered, both because it would be usual for single tests to be done in isolation in clinical practice and also because when using each single test at the pre-specified threshold the test could take only two values (positive or negative).

3.7.4.1 Decision curve analysis and net benefit

Decision curve analysis [167] was used to show the net benefit of each index test at varying threshold probabilities. Net benefit is calculated as

$$\text{Net Benefit} = (sensitivity \times prevalence) - [(1 - specificity) \times prevalence \times w] \tag{3.1}$$

Where w is the odds at the threshold probability.

With an index test that takes values indicating the predicted probability of disease $\hat{p}$, p_t is the threshold probability, and sensitivity and specificity at a given threshold are calculated by defining test positive as $\hat{p} > p_t$ [351]. The threshold probability p_t indicates the preferred management approach for the clinician when dealing with a patient, a higher p_t indicates a stronger preference for avoiding unnecessary interventions. At a threshold probability of 80% (1:4 odds) the clinician is of the opinion that treating a well person as having dementia is four times as bad as missing a case. At a threshold probability of 10% (9:1 odds) the clinician is of the view that missing a case of dementia is nine times as bad as intervening in someone who doesn't have dementia [351]. Net benefit is expressed in units of true positives, typically does not consider confidence intervals, and there is no minimum specified difference in net benefit required to identify the optimal strategy [351]. Decision curves are plotted with a *treat all* line which indicates the net benefit of treating everybody as disease positive, and a *treat none* line which conversely indicates the net benefit of treating nobody.

Figure 3.2 shows an example decision curve for MMSE based on data from a systematic review [157]. Curves are plotted for the net benefit of MMSE for the target condition dementia, at test thresholds (referred to in this paragraph as "cut-points") of 17 (sensitivity 70.1% specificity 92.9%) and 24 (sensitivity 100% specificity 45.9%) indicating normal; 77 of 360 in the data-set had dementia (prevalence 21.39%). The treat none line and the treat all line are also plotted. The figure shows that at threshold probabilities of up to 15% the MMSE at a cut-point of 24 has the greatest net benefit. In contrast, the MMSE at a cut-point of 17 has greatest net benefit at threshold probabilities of 15% to 73%, and at threshold probabilities above 73% the treat none approach has the greatest net benefit. Recall Equation 3.1. At a threshold probability of 30%, MMSE at a cut-point of 17 has a net benefit of

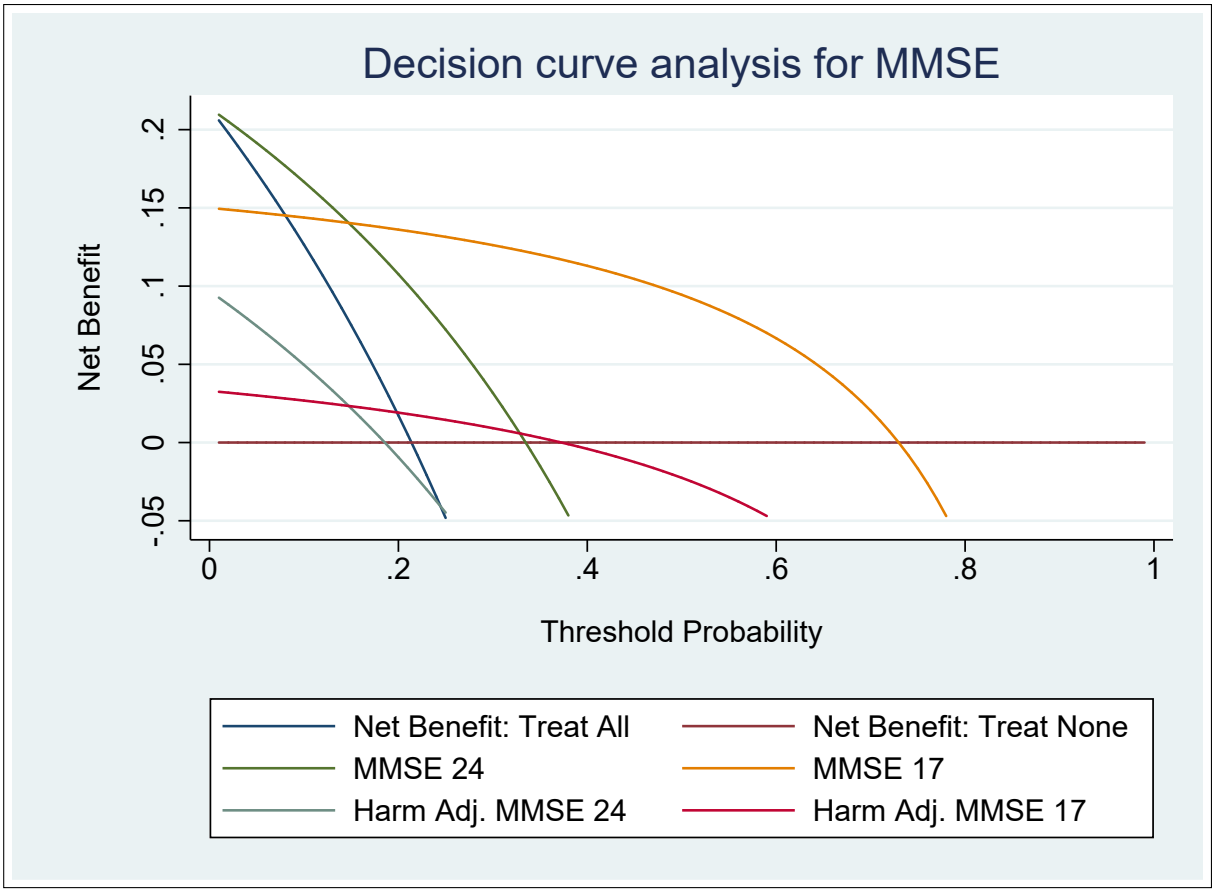
$$\text{Net Benefit} = (0.701 \times 0.2139) - \left[(1 - 0.929) \times 0.2139 \times \frac{0.3}{0.7}\right] \tag{3.2}$$

$= 0.126024 \approx 0.13$, meaning that there are 13 true positives for every 100 patients in the target population. In contrast MMSE at a cut-point of 24 has a net-benefit of

$$\text{Net Benefit} = (1 \times 0.2139) - \left[(1 - 0.459) \times 0.2139 \times \frac{0.3}{0.7}\right] \tag{3.3}$$

= 0.0316371. ≈ 0.03, meaning that there are 3 true positives for every 100 patients in the target population.

Figure 3.2: Example decision curves for MMSE at thresholds of 17 and 24



The curves plot the net benefit across a range of thresholds, along with the treat-all and treat-none lines.
Curves are smoothed, as recommended to minimise unstable results [168]
Based on data from a systematic review [157]

In decision curve analysis harm can be quantified as the reciprocal of *"the number of tests that would be done to identify one true case, **assuming the test is perfect**"* [351, 352]. All of the evaluated index tests are relatively simple tests (mostly pen-and-paper), and are not invasive. However, there may be harm from the time spent completing the test, and this is also the main determinant of the test cost. Therefore harm for each test was derived from the number of tests per hour, based on the median test duration:

$$harm = \frac{1}{\frac{60}{\text{median test duration}}} \tag{3.4}$$

Therefore harm is less when the test duration is less. To continue the example in Figure 3.2, MMSE is said to take five minutes for someone with normal cognition and seven minutes for someone with

dementia [157], so the harm can be calculated [2]:

$$\text{harm for MMSE} = \frac{1}{\frac{60}{7}} \tag{3.5}$$

Therefore harm for MMSE = 0.117, which is the reciprocal of the number of people tested per hour. Figure 3.2 shows that the harm adjusted net benefit curves for MMSE are parallel to the curves without adjustment for harm, and are 0.117 TP lower at each threshold.

For informant completed tests and GP judgement the harm was zero because these tests would not take any additional time in an additional consultation; GP judgement would be reached after the end of an index consultation and an informant completed test could be given to the informant to take away and return. Minicog was derived from items on GPCOG and so no time was available, so time was estimated from the time taken to complete GPCOG.

Decision curves were plotted for each of the individual index tests, for both target conditions, both with harm = 0 (naïve curves) and harm quantified as in Equation 3.5 (harm-adjusted curves). Curves for the single tests were plotted on the same axis as the test in combination with GP judgement (see Section 3.7.5.1) to facilitate interpretation, because it would be typical in clinical practice for the index tests to be done in combination with GP judgement; the plots are provided in Appendix H. A plot of naïve curves for each of the index tests combined with GP judgement, on a single axis, is provided in Section 4.6, and a separate axis shows all harm-adjusted curves.

As detailed in Section 1.1.3.2, decisions about treatment are based on probabilities which are derived from the ratio of the costs and benefits. At the present time, given the absence of a disease modifying treatment, especially in an older adult with multi-morbidity, the threshold for diagnosis of dementia may be relatively high, because false positives are arguably more harmful than false negatives. In contrast the threshold probability for normal cognition in someone with cognitive symptoms may be lower than the threshold probability for dementia, because interventions to aid cognition and independence are arguably of benefit regardless of whether cognition is normal or impaired. Therefore to facilitate interpretation, the plot of multiple index tests in Section 4.6 is shown only for threshold probabilities above 80% for the target condition dementia and above 60% for the target condition normal.

[2] in a sample of 240 people of whom 55% have dementia the median test time for MMSE would be seven minutes

3.7.5 Index tests in combination

3.7.5.1 Two test combination: GP combined tests

The accuracy of each index test in combination with GP clinical judgement was evaluated (termed GP combined tests). This approach was taken because in clinical practice GP clinical judgement would likely be formulated before any other tests took place. Based on clinical experience and reports of GP diagnosis of dementia in practice which indicated that up to four consultations were used to make the diagnosis [340], it was judged probable that a GP would use an initial encounter to formulate a clinical judgement about the cognition and then use subsequent encounters to conduct further tests based on their clinical judgement.

Logistic regression equation Logistic regression analyses were used to evaluate the accuracy of each index test in combination with GP clinical judgement (the GP combined tests) for the diagnosis of the two target conditions: dementia or normal. The dependent variable was the target condition (either dementia or normal, taking values: 0=target condition not present, 1=target condition present). The independent variables were GP clinical judgement and one of the other index tests listed in Section 3.7.2. GP clinical judgement was represented by the variable *gpd* and *gpn* respectively indicating a GP judgement of dementia or normal, and taking values 0 (test negative) or 1 (test positive). Similarly, for the other index tests the possible values were 0=test target condition not present, 1=test target condition present. The logistic regression equation took the form (for example) `logistic dementia gpd moca`. The `predict` command was used to predict the probability of the target condition based on the results of the logistic regression equation in the estimation sample. For each GP combined test, predicted probabilities took one of up to four categories of probability, because there were up to four permutations of values for the independent variables in the logistic regression equations (0|0; 0|1; 1|0; 1|1). The `diagt` command was used to calculate the diagnostic accuracy. `Diagt` collapsed the categories of predicted probabilities of disease into two categories, the highest predicted based on the logistic regression equation which defined test positives, and the other categories.

Evaluation of GP combined tests Standard measures of diagnostic accuracy (sensitivity, specificity, likelihood ratios, and predictive values) were calculated for the GP combined tests, together with 95% confidence intervals. As described in Section 3.7.4.1, decision curves were plotted for all tests, both as single tests and as GP combined tests, for both target conditions, both with harm = 0 (naïve curves) and harm quantified as in Equation 3.5 (harm-adjusted curves); these plots are provided in Appendix H.

The discrimination of the GP combined tests compared to the single test was assessed using visual inspections of ROC curves plotted using `roccomp`. Calibration for GP combined tests was assessed by using `pmcalplot` to plot observed probabilities of dementia based on the reference standard, against the predicted probabilities (based on paragraph *logistic regression equation* above).

The individual plots of calibration and discrimination are provided in Appendix I. GP combined tests were judged to have good discrimination if the AUROC was ≥ 0.75; calibration was assessed by visual inspection [353].

From the GP combined tests that were judged to have good discrimination and calibration four decision curves were plotted on the same axis, two **p**atient-completed ($t^{\mathbf{p}}$A and $t^{\mathbf{p}}$B) and two **i**nformant-completed ($t^{\mathbf{i}}$A and $t^{\mathbf{i}}$B) combined tests, selected on the basis of the highest LRP for the GP combined test. LRP was used as the indicator of diagnostic accuracy because an aim of the thesis was *to identify if a combination of tests can have a high positive predictive value for identifying people with dementia* and LRP determines the the informative value of a positive test result [148]. Patient- or informant- completion was considered because in clinical practice it would be preferable to have objective evidence of both impaired cognitive function and of concern from a knowledgeable informant (see definitions in Section 1.1.1.1).

The analysis of tests using decision curves on the same axis was limited to four (t^pA and t^pB;t^iA and t^iB [3]) for each target condition because plotting all of the possible naïve curves (let alone the harm-adjusted curves) would have resulted in 19 lines on the plot (17 index tests, treat all, treat none) which was judged to be a barrier to interpretation. Comparison of naïve curves rather than harm-adjusted curves was done because (a) the pen-and-paper nature of the tests meant that other than time there was little or no harm from any test (b) harm adjusted curves have lower net benefit than naïve curves, making comparisons between curves difficult because tests can have negative net benefit and therefore not be displayed on the curve, especially at higher threshold probabilities.

3.7.5.2 Three test combination: GP 360 tests

As described, for a credible clinical diagnosis it would be preferable for a combined test to include both informant- and patient- completed measures and therefore the accuracy of three combined tests (GP judgement and a patient-completed test and an informant-completed test) for diagnosing both target conditions was analysed. For each target condition, the four GP combined tests (t^pA and t^pB; t^iA and t^iB) with the highest LRP were also evaluated as a combination of three tests (termed GP 360 test, as in a 360°evaluation).

The results of the logistic regression equation for three tests in combination resulted in eight categories: (+|+|+; +|+|-; +|-|+; -|+|+; -|-|+; +|-|-; -|+|-; -|-|-). For each GP 360 test discrimination and calibration were assessed as previously. For the target condition dementia the GP 360 test was dichotomised at a 80% predicted probability of dementia, and for normal as the target condition the four GP 360 tests were dichotomised at a 60% predicted probability of normal (Section 3.7.4.1). Discrimination of both the dichotomised and eight-category GP 360 tests are presented on the same axis in the Results Chapter. Calibration plots are presented in Appendix I. For both the dichotomised tests, and the continuous probabilities from the logistic regression equation measures of diagnostic

[3] two **p**atient-completed (t^pA, t^pB) and two **i**nformant-completed (t^iA, $t^{\mathbf{i}}$B) GP combined tests, selected on the basis of the highest LRP for the GP combined test

accuracy were calculated using `diagt`, which as described in paragraph *logistic regression equation* dichotomised the continuous probabilities so that the group with the highest predicted probability was taken to indicate test positive. Sensitivity, specificity, likelihood ratios, and predictive values were calculated, together with 95% confidence intervals. For each target condition four naïve decision curves were plotted on the same axis for each of the four possible GP 360 tests (t^pA|t^iA, t^pA|t^iB, t^pB|t^iA, t^pB|t^iB), both dichotomised and continuous. The predicted classification of people according to each of the four GP 360 tests was tabulated using natural frequencies. For each target condition the optimal GP 360 test was identified.

3.7.6 Sensitivity analysis

Sensitivity analyses are detailed in Sections below. Sensitivity analyses were not done to investigate whether diagnostic accuracy for the single tests varied with age and sex because it would be typical in clinical practice to combine the index tests with GP clinical judgement. Sensitivity analyses were not done on GP 360 tests because it was judged this analysis would be under-powered. Sensitivity analysis on the order effect of tests within the battery, and the order of index test assessment and specialist evaluation, were not done because these factors were randomised. Sensitivity analyses exploring the accuracy of individual tests at different thresholds were not done because of the risk of chance findings due to multiple analyses.

3.7.6.1 Continuous test scores

In clinical practice clinicians typically categorise continuous test scores as either test positive or negative (Section 1.1.2.4). However, this dichotomisation results in a loss of data and when combining tests may attenuate the diagnostic accuracy. Diagnostic accuracy was calculated for the GP combined tests when using the component patient-completed and informant-completed index tests as continuous scores, for each target condition. The predicted probabilities of the target condition from the logistic regression were used to calculate diagnostic accuracy, as described in paragraph *logistic regression equation* above. The predicted probabilities were dichotomised at a threshold of 80% for dementia and 60% for normal (see Section 3.7.4.1 for rationale for these thresholds), to facilitate comparisons between the main analysis and the sensitivity analysis, with predicted probabilities above this indicating test positive and probabilities below this indicating test negative for the target condition. This was done because the `diagt` command dichotomises the test at the highest predicted probability of the target condition, which would vary between the categorical and continuous index tests, and thus artificially impact diagnostic accuracy.

3.7.6.2 Age and sex

Sensitivity analyses were also done to investigate whether diagnostic accuracy for the GP combined tests varied with age and sex. Age was dichotomised as age < 80 years or age ≥ 80 years because there

is evidence that predictors vary in performance in these groups [90]. Logistic regression was done restricted to each group (Men | Women; Age < 80 years | Age ≥ 80 years) and used to generate predicted probabilities of the target condition. The predicted probabilities were dichotomised at a threshold of 80% for dementia and 60% for normal. The `diagt` command could not be used to calculate diagnostic accuracy because of small numbers in the analysis, and so the `estat classification` command was used to generate the sensitivity and specificity directly from the results of the logistic regression equation, and confidence intervals were then calculated by extracting the values of the regression coefficients and classification table, calculating the test positive/negatives and the total tested, and then using `cii proportions` to calculate the confidence interval for a proportion, for each of the 17 GP combined tests, for each target condition.

For heterogeneity by age and sex, the proportion in each group (Men | Women; Age < 80 years | Age ≥ 80 years) with dementia, MCI and normal cognition was calculated, along with the mean age, ACE3 score and school leaving age. Linear regression was used to investigate an association between the independent variable (age | sex) and the dependent variables ACE3 score and school leaving age.

3.7.6.3 Combining tests

Further sensitivity analyses were done to explore the diagnostic accuracy when each of the 12 GP combined patient-completed index tests were combined with each of the five informant completed tests, for each target condition. The `estat classification` procedure described in the above paragraph was used to calculate the diagnostic accuracy for each of the 60 combinations (12 × 5) for each target condition.

3.7.6.4 Geography

A final sensitivity analysis investigated heterogeneity in GP judgement by setting, with two groups Bristol (which had the diagnosis scheme, see Section 3.2.2), and South Gloucestershire/North Somerset which did not. The proportion of people with dementia according to the reference standard was also calculated for Bristol and South Gloucestershire/North Somerset.

3.7.7 Sample size

Standard tables [354] were used to calculate that a minimum sample size of 200 was needed for a 95% confidence interval lower bound of 80%, based on a specificity of 95%, based on prior studies [255, 273] and a 75% proportion of people with dementia in a population of people with cognitive complaints, based on data from Bristol memory clinic data prior to July 2013; this allowed evaluation of between 30-45 diagnostic indicators, based on the five events per variable rule [355].

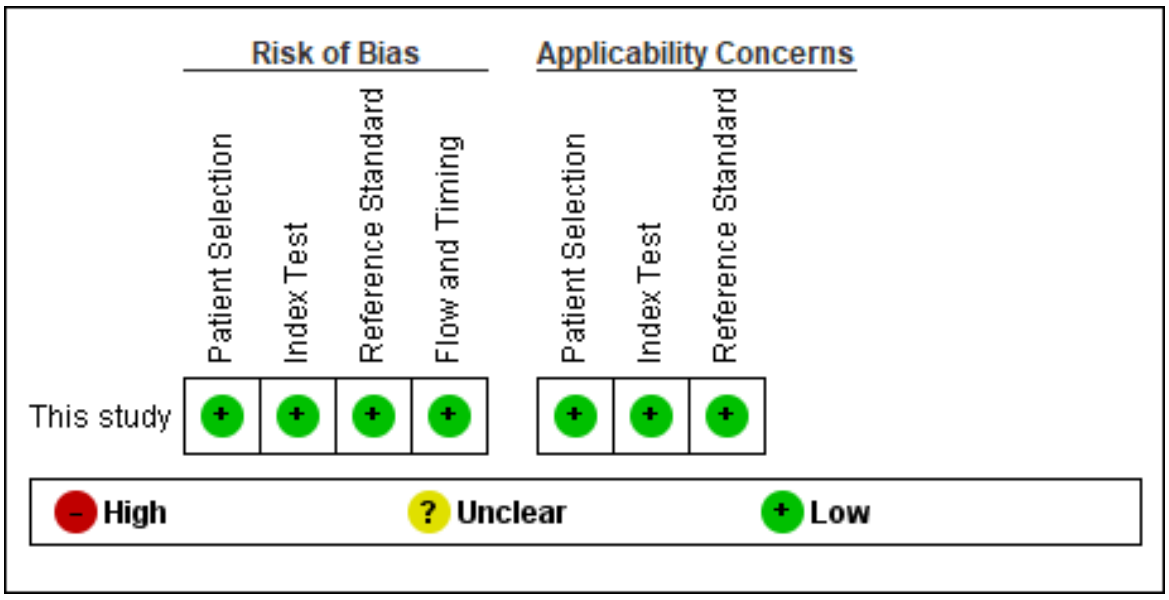
3.8 Risk of bias

Figure 3.3 shows the appraisal of the risk of bias in the described work using the QUADAS-2 tool [199]. QUADAS-2 is a tool to judge the risk of bias in diagnostic test accuracy studies and comprises four domains, each with signalling questions to assess the risk of bias. QUADAS-2 is summarised as:

1. **Patient selection**: Was a consecutive or random series of patients enrolled? Was a case control design avoided? Did the study avoid inappropriate exclusions?
2. **Index test**: Were the index test results interpreted without knowledge of the results of the reference standard? If a threshold was used, was it pre-specified?
3. **Reference standard**: Is the reference standard likely to correctly classify the target condition? Were the reference standard results interpreted without knowledge of the results of the index test?
4. **Flow and timing**: Was there an appropriate interval between index test(s) and reference standard? Did all patients receive a reference standard? Did all patients receive the same reference standard? Were all patients included in the analysis?

The first three domains also consider concerns about applicability of the study to the review question, which is less relevant for this assessment of risk of bias but is relevant to the applicability of the findings to clinical practice.

Figure 3.3: Risk of bias in this study using the QUADAS-2 [199] tool



Overall in this study all four domains were judged to be at low risk of bias. Patient selection was judged to be at low risk of bias because a consecutive sample of patients was enrolled (the impact of

GP referral bias is discussed in Chapter 6), and a case control design and inappropriate exclusions were avoided. The index test domain was judged to be at low risk of bias because the results were interpreted independently of the reference standard, and pre-specified thresholds were used. The reference standard was judged at low risk of bias because it was likely to correctly classify the target condition and it was interpreted without knowledge of the index test. The flow and timing was judged at low risk of bias because there was an appropriate interval between index test and reference standard (done same day), ICD-10 and Peterson MCI definitions were used consistently and applied by a specialist, and all patients were included in the analysis.

3.8.1 Differences to protocol

A study protocol was published [356]. The methods in this Chapter differ to the methods in the published protocol in two aspects. Firstly, it was not possible to ascertain the diagnosis according to different reference standards within the time limits of the thesis due to difficulties with the consensus panel. Secondly, multiple imputation was not used in the analysis because the missing data were judged to violate the missing at random assumption.

Chapter Summary

This Chapter has presented the methods for the empirical quantitative investigation of diagnostic test accuracy. The target conditions for evaluating diagnostic accuracy were dementia of any aetiology, and normal. Participants were people aged at least 70 years with symptoms of dementia, who had been referred by their GP to a research study conducted in general practices. Participants were recruited from 21 practices with a total population of 249,556, of whom 34,956 were aged at least 70 years.

Referring GPs were asked to provide their clinical judgement about the extent of cognitive problems on a referral form which was sent to a secure nhs.net email account. Electronic prompts were used at the point of care to facilitate referral of a consecutive series of patients. Standard procedures were used by research staff to process referrals, contact potential participants, confirm eligibility, and arrange appointments and transport.

At the research clinic two examiners undertook a separate blinded examination on each participant. All index test evaluations were conducted by one physician, and all reference test evaluations were conducted by a second physician. The reference standard for dementia was expert specialist assessment according to ICD-10 criteria and Peterson criteria [93] for MCI. Medical records were reviewed for the period six months before and after the research clinic.

Characteristics of participants and tests were summarised using plots and regression analyses. Accuracy of tests was investigated as single tests, in combination with GP judgement (GP combined tests) and as a combination of GP judgement, informant-completed and patient-completed tests (GP 360 tests). Performance of tests was evaluated using measures of discrimination, calibration, diagnostic accuracy, and net benefit (decision curves).

CHAPTER 4

RESULTS

This Chapter covers the results of the quantitative diagnostic accuracy study. Firstly details of ascertainment and missing data are provided. Secondly, the characteristics of participants are described, with respect to age, sex, and cognitive status. Thirdly the characteristics of the index tests are described regarding how the scores and duration were associated with cognition. Fourthly the target conditions are described, the prevalence of the target conditions reported, and GP judgement cross tabulated against the reference standard. Fifthly the informant measures are described, weighted kappa is reported for the informant completed and interview completed tests, and ROC curves of the paired informant tests are plotted. In the sixth Section diagnostic accuracy is reported for both target conditions, firstly classification accuracy is reported, then diagnostic accuracy of single tests is provided, and the meta-analysis of GP judgement from Chapter 2 is updated. The accuracy of index tests in combination with GP clinical judgement (GP combined tests) is then reported for both target conditions. Next, the accuracy of GP clinical judgement, a patient completed test, and an informant completed test is reported (GP 360 test). The optimal combination of tests for each target condition is identified. In the seventh Section the sensitivity analyses are reported.

4.1 Ascertainment and missing data

Table 4.1 shows the recruitment of participants by each practice, with details of the number of people on each GP practice list in each of five age groups: 70-74 years; 75-79 years; 80-84 years; 85-89 years; 90+ years. Clevedon (L81102) had the highest overall proportion of people aged over 70 years out of the total list size (19%), other practices with a high proportion of people aged over 70 years were Portishead which had 18% eligible, and Backwell, Clevedon (L81040), Frampton Cotterell, and Yate which all had 17% eligible. In contrast Redfield had only 6% eligible, Horfield 7%, Clifton 8%

and Stoke Bishop 9%. The number of participants recruited per practice varied from two (Redfield) to 31 (Westbury on Trym). Because sites had different set up dates (see Methods Table 3.2) naïve comparisons of the recruitment figures for practices, such as the calculated proportion who were recruited out of the total who were potentially eligible, are likely to be misleading. Table 4.1 also shows the number of people who were predicted to be eligible based on the person years at risk. The person years at risk estimate of eligible population accounts for duration of recruitment from each practice but does not account for factors such as changes in practice list size *during* recruitment due to practice mergers; and the impact of mergers and other practice business, such as change in staffing or other contractual commitments, that may have led a recruiting practice to be fallow (effectively non-recruiting) for a period.

Table 4.1: Practice recruitment

Location (code) ∗	**List size**†	**by age**	**group**	**in years**		**Predicted**	**Predicted**	**Recruited**
	70-74	*75-79*	*80-84*	*85-89*	*90+*	**eligible** ‡	**eligible PYAR**	
Backwell (L81060)	925	568	473	286	126	65	88	20
Bedminster (L81053)	321	288	285	168	98	39	85	4
Clevedon (L81102)	420	338	266	180	179	50	29	7
Clevedon (L81040)	893	682	502	385	239	87	174	29
Clifton (L81081)	382	270	196	109	85	31	35	3
Close Farm (L81050)	314	204	164	82	39	21	30	11
Fishponds (L81087)	356	253	203	130	82	32	62	3
Frampton Cotterell (L81014)	823	710	534	321	165	76	89	9
Hanham (L81079)	1098	825	713	454	253	104	225	30
Henleaze (L81131)	390	341	307	267	196	59	90	4
Horfield (L81022)	437	262	199	137	98	35	73	5
Kingswood (L81063)	521	406	339	234	132	52	98	12
Long Ashton (L81056)	302	253	206	119	59	28	25	2
Mendip Vale (L81086)	1390	902	623	367	204	94	188	20
Portishead (L81004)	1193	821	585	394	242	96	112	9
Redfield (L81061)	191	138	87	63	45	16	19	2
Shirehampton (L81008)	419	318	237	159	95	38	80	6
Southmead (L81098)	296	230	205	132	88	32	41	5
Stoke Bishop (L81622)	140	97	75	34	20	10	18	4
Westbury on Trym (L81017)	434	309	254	215	163	50	103	31
Yate (L81047)	815	659	404	216	105	57	71	25

∗ Location is given rather than practice name because practice names change with mergers and change of doctors. Practice code is a unique identifier held by NHS England

† based on NHS England data for 2018 [188]

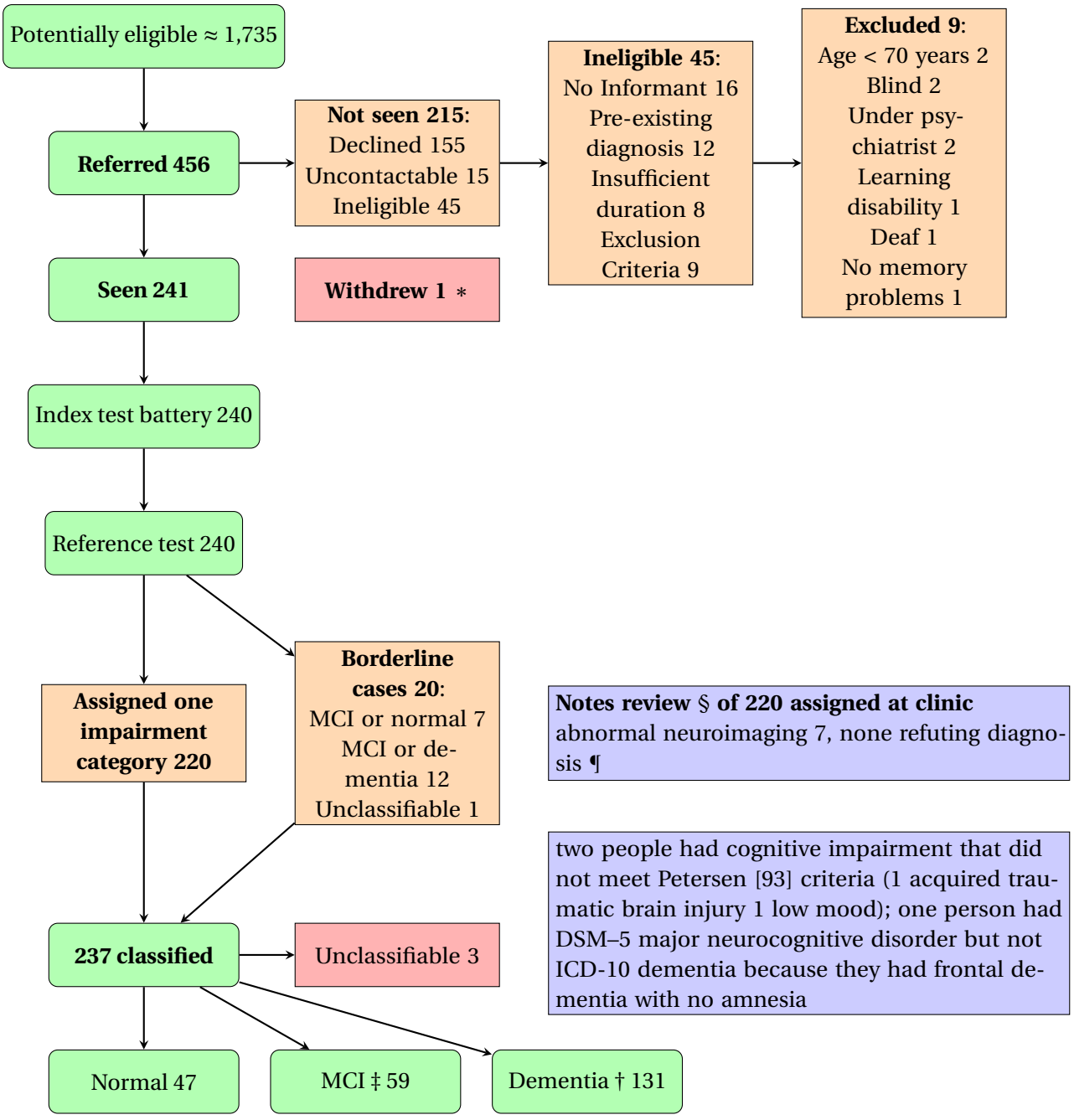
‡ based on age specific incidence per 1000 person years at risk [14], multiplied by list size.
The meta-analytic incidence of dementia per 1000 person years at risk by age group is:
9.3 (70-74 years), 17.3 (75-79 years), 32.0 (80-84 years), 57.0 (85-89 years), and 122.4 (90+years) [14]

PYAR Person years at risk

As outlined in Methods Table 3.2 there were a total of 249,556 people on the combined GP practice lists in April 2018 and of these 34,956 were aged at least 70 years and so were potentially eligible for the study if they developed symptoms of dementia. The meta-analytic incidence of dementia per 1000 person years at risk by age group is: 9 (70-74 years), 17 (75-79 years), 32 (80-84 years), 57 (85-89 years), and 122 (90+years) [14] (Section 1.1.3.3). Applying these age specific incidence rates to the study population described in Table 4.1 it follows that the potentially eligible population who might be expected to develop symptoms based on person years at risk would be 1,735 people based on person years at risk.

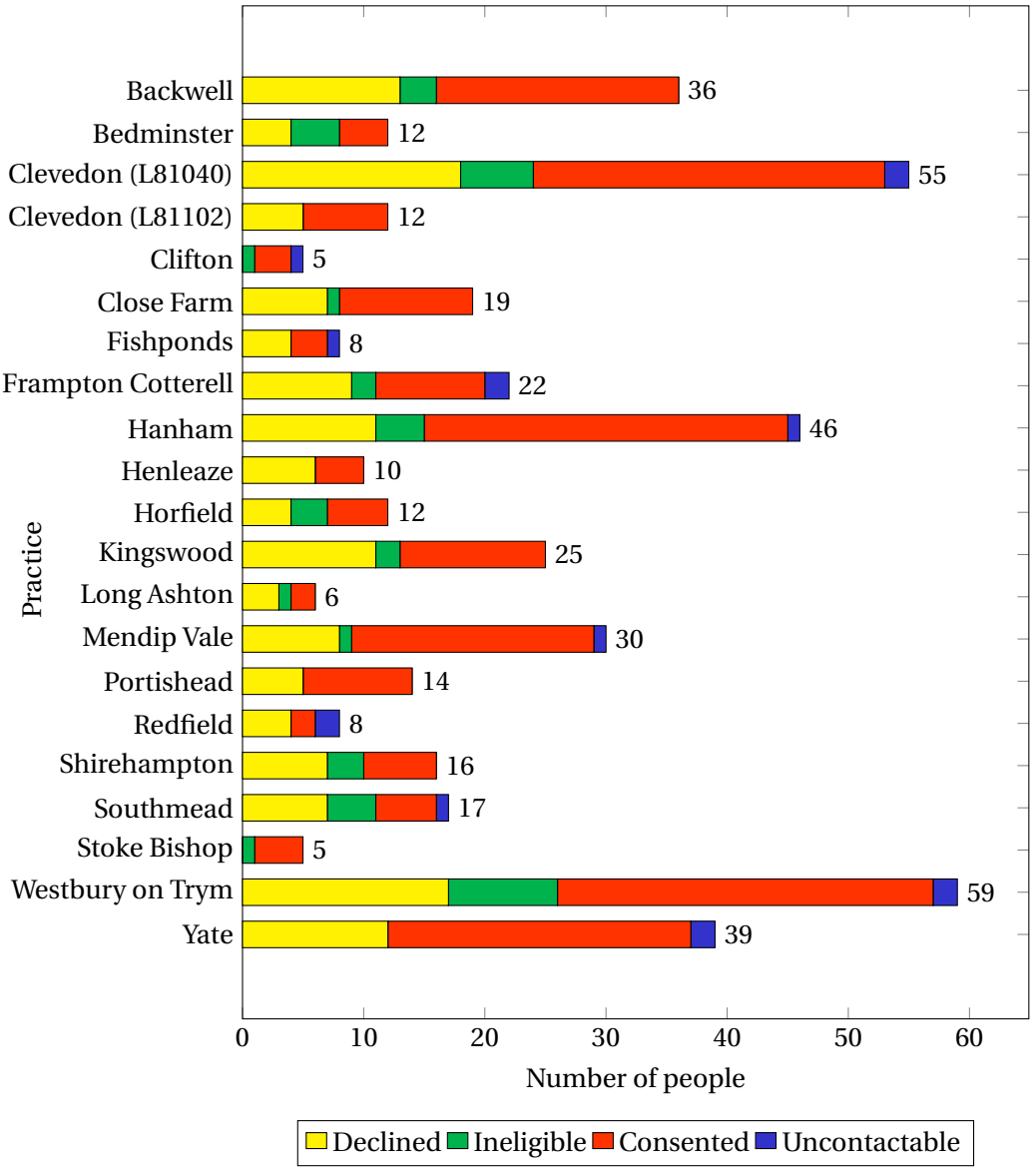
Figure 4.1 presents the STARDdem [357] participant flowchart. Of the 34,956 people aged over 70 years in the recruiting practices, 456 were referred, which is 26% (95% CI 24% to 28%) of the estimated potentially eligible population. Of the 456 people who were referred 241 (53%; 95% CI 48% to 57%) participated, 170 (37%; 95% CI 33% to 42%) were uncontactable or declined, and 45 (10%; 95% CI 7% to 13%) were ineligible. Figure 4.2 shows the distribution by practice of people consenting to participate, who were referred but deemed ineligible, who were eligible but declined to participate, or who were uncontactable. Of the 456 people who were referred, 53% (95% CI 48% to 58%) were contactable, eligible, and consented to participate. Overall 34% (95% CI 30% to 39%) of referred people declined to participate, ranging from 0% in Clifton and Stoke Bishop (95% CI 0% to 52%) to 60% (95% CI 26% to 88%) in Henleaze. In contrast 10% (95% CI 7% to 13%) of referred people were ineligible; with a number of practices referring 0% ineligible people: Henleaze (95% CI 0% to 31%), Clevedon (L81102) (95% CI 0% to 26%), Fishponds and Redfield (95% CI 0% to 37%), Portishead (95% CI 0% to 23%), and Yate (95% CI 0% to 10%). In contrast, of the 12 referrals from Bedminster 33% (95% CI 10% to 65%) were of ineligible people. The sustained efforts to make contact with people were generally successful and overall only 3% (95% CI 2% to 5%) were uncontactable.

Figure 4.1: STARDdem flowchart



∗ one person, who was randomly assigned to the expert assessment first (Section 3.4) had to withdraw part way through the reference test as they were acutely ill
† Dementia according to ICD-10 definition
‡ MCI according to [93] § see Section 3.5
¶ 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.

Figure 4.2: Referrals from practices



4.2 Characteristics of participants

Table 4.2 summarises the characteristics of participants and shows that there were a total of 240 participants included in the analysis; as described in Figure 4.1 one person (a man) withdrew after consenting due to acute illness. Also as described in Figure 4.1 three people could not be classified on the ICD-10 reference standard, of these two people had cognitive impairment that did not meet Petersen [93] criteria (1 acquired traumatic brain injury 1 low mood) and one person had DSM–5 major neurocognitive disorder but not ICD-10 dementia because they had frontal dementia with no amnesia.

Table 4.3 shows the known characteristics for people who declined compared to people who participated. There was weak evidence of an association between declining and a GP 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). There was some evidence for an association between declining and age (odds ratio per year 1.08; 95% CI 1.04 to 1.12), or female sex (odds ratio 1.88; 95% CI 1.21 to 2.92). For participants, the median time between referral (the date of GP judgement) and the clinic appointment (reference standard) was 47 days (IQR 30 to 72 days), the longest interval was 177 days, approximately six months, which was due to difficulties with 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; p=0.254). All other index tests other than GP judgement were done on the same day as the reference standard.

Of the 240 people in the analysis, there were 114 women and 126 men. Overall the median age of participants at the research clinic was 80 years (interquartile range 75 years to 85 years). Compared to people with normal cognition (mean age 76.5 years; s, standard deviation, 5.2 years) people with MCI were a mean 2.9 years older (95% CI 0.8 years to 5.0 years) and people with dementia a mean 5.5 years older (95% CI 3.7 years to 7.3 years). The three people who could not be classified according to the ICD-10 reference standard (*other* in Table 4.2) were a mean 3.7 years older than people with normal cognition (95% CI -2.7 years to 10 years). School leaving age was available for 234 people who reported a median leaving age of 15 years (interquartile range 15 years to 16 years). Age at retirement was available for 207 people, who reported a median retirement age of 60 years (interquartile range 58 years to 65 years).

Duration of symptoms was available for 171 people, who reported a median 24 months symptoms (interquartile range 12 to 36 months). The reported duration of symptoms was similar in people with normal cognition, MCI, and dementia, and those classified other. Most (87%) participants reported a gradual onset of symptoms, a similar proportion (88%) reported no fluctuation in their symptoms, and the findings were similar regardless of cognition. In contrast, most (76%) participants reported that their symptoms were progressive but this was reported more often by people with dementia (84%, 95% CI 77% to 90%) than people who had MCI (69%, 95% CI 56% to 80%), or normal (62%, 95% CI 46% to 75%) or other cognition (67%, 95% CI 9% to 99%).

Two people could not complete the ACE3 because English was not their first language, they had both been offered an interpreter and declined. In both cases sufficient information was available

Table 4.2: Characteristics of participants

Characteristic	Cognitive diagnosis ∗ Dementia *n=131*	MCI *n=59*	Other *n=3*	Normal *n=47*
Sex *n (row %)*				
Male	67 (53)	34 (27)	2 (2)	23 (18)
Female	64 (56)	25 (22)	1 (1)	24 (21)
Age (years) *Median (IQR)*				
At clinic	82 (77-87)	80 (75-83)	80 (79-82)	75 (72-80)
Left School	15 (15-16)	15 (15-16)	17 (15-18)	16 (15-16)
Retired	60 (58-65)	61 (59-67)	60 (56-60)	61 (58-65)
Symptom onset				
Median (IQR) (months)				
Time ago	24 (12-36)	18 (12-24)	24 (24-24)	21 (12-36)
Type, n (column %)				
Gradual	110 (84)	54 (92)	2 (67)	43 (91)
Sudden	13 (10)	4 (7)	1 (33)	0 (-)
Uncertain	8 (6)	1 (2)	0 (-)	4 (9)
Symptom pattern *n (column %)*				
Course				
Progressive	110 (84)	41 (69)	2 (67)	29 (62)
Stepwise	2 (2)	0 (-)	0 (-)	0 (-)
Regressive	1 (1)	0 (-)	1 (33)	1 (2)
Static	5 (4)	7 (12)	0 (-)	9 (19)
Uncertain	13 (10)	11 (19)	0 (-)	8 (17)
Fluctuation				
None	111 (85)	51 (86)	3 (100)	45 (96)
Within one day	12 (9)	5 (8)	0 (-)	1 (2)
Over several days	8 (6)	3 (5)	0 (-)	1 (2)
ACE3 † score *Median (IQR)*				
Total (max 100)	69 (61-74)	82 (76-87)	90 (70-94)	93 (90-95)
Attention (max 18)	14 (11-16)	16 (15-17)	17 (15-18)	17 (16-18)
Memory (max 26)	13 (10-17)	19 (14-22)	22 (22-25)	23 (22-25)
Fluency (max 14)	7 (5-9)	9 (7-11)	10 (7-12)	11 (11-13)
Language (max 26)	22 (19-24)	24 (23-26)	25 (22-25)	26 (25-26)
Visuospatial (max 16)	12 (10-14)	14 (12-15)	16 (12-16)	16 (15-16)
GP opinion *n (row %)*				
Normal	6 (18)	8 (24)	1 (3)	19 (56)
Cognitive impairment	51 (43)	40 (33)	2 (2)	27 (23)
Dementia	74 (86)	11 (13)	0 (-)	1 (1)

∗ *Dementia* according to ICD-10 definition *MCI* according to Peterson definition
Other impairment: not normal but impairment not meeting formal criteria
Due to low mood in one case and traumatic brain injury in one case
† *ACE3* Addenbrooke's Cognitive Examination III

Table 4.3: Comparison of people who participated and declined

Facet	Participated	Declined
Demographics		
Male n (%)	126 (53)	47 (37)
Female n (%)	114 (47)	80 (63)
Median age (years)	80	83
IQR age (years)	75 to 84	79 to 88
GP judgement		
Dementia % (95% CI)	36 (30 to 42)	48 (39 to 58)
CIND % (95% CI)	50 (44 to 56)	42 (33 to 51)
Normal % (95% CI)	14 (10 to 19)	10 (5 to 17)

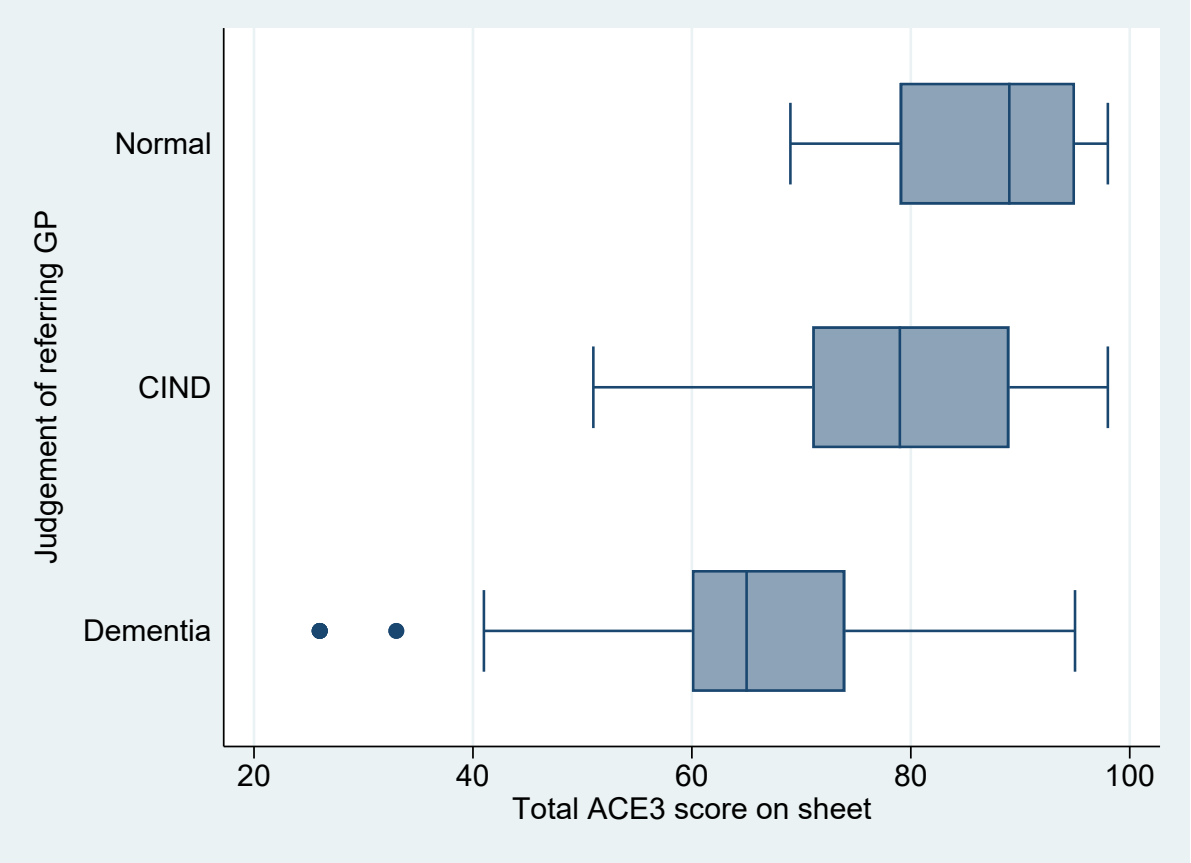
IQR interquartile range
CIND cognitive impairment, but not dementia

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 ACE3 score, the median was 75 (interquartile range 65 to 87). Compared to people with normal cognition (who scored a mean (s) 92 (5) points) people with MCI scored a mean 12 points less (95% CI 16 points less to 8 points less) on the ACE3, and people with dementia scored a mean 26 points less (95% CI 29 points less to 22 points less).

Referring GPs judged that 34 people were normal, 86 had dementia, and 120 had cognitive impairment, but not dementia (CIND); the one person who withdrew from the study due to acute illness was judged by the referring GP to have CIND. People who GPs judged as having dementia had a mean age 81.6 years (σ 5.7 years; 95% CI 80.4 years to 82.8 years) compared to people who they judged as not having dementia (including people judged as normal and CIND) who had a mean age 79.5 years (σ 5.7 years; 95% CI 78.6 years to 80.5 years). People that GPs judged as having dementia had a total ACE3 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 (cognitive impairment, but not dementia) had an ACE3 score IQR 71 to 89. By way of comparison, in contrast to GP judgement, people who were test positive on MOCA had an ACE3 score IQR of 65 to 85, people who were test positive on Minicog had an ACE3 score IQR of 61 to 75, people who were IQCODE test positive had an ACE3 score IQR of 65 to 82, and people who were test positive on AD8 had an ACE3 score IQR of 65 to 83.

As shown in Table 4.4, in the attention sub domain of ACE3 (max score 18) people with normal cognition scored a mean of 17 points, and people with dementia scored 4 points less (95% CI 3 points less to 5 points less). In the memory sub domain (max score 26) people with normal cognition scored a mean of 23, people with MCI scored 5 points less (95% CI 4 points less to 7 points less) and people with dementia scored 10 points less (95% CI 9 points less to 11 points less). In the fluency sub domain (max score 14) people with normal cognition scored a mean of 12, people with MCI scored 3 points less (95% CI 2 points less to 4 points less) and people with dementia scored 5 points less (95% CI 4

Figure 4.3: Scores on ACE3 by GP judgement



The box plots the median (darker middle line) and the quartiles (box edges), the whiskers enclose the lower (Q1 - 1.5 × interquartile range) and upper (Q3 + 1.5 × interquartile range) adjacent values, and the dots mark the outlying values.

points less to 6 points less). In the language sub domain (max score 26) people with normal cognition scored a mean of 25 and people with dementia scored a mean 4 points less (95% CI 3 points less to 6 points less). In the visuospatial skills sub domain (max score 16) people with normal cognition scored a mean of 15 points, people with MCI 2 points less (95% CI 1 point less to 3 points less), and people with dementia 3 points less (95% CI 2 points less to 4 points less). Comparisons with the other cognition group are available in Table 4.4 but are not discussed in the text because the small numbers (n=3) result in a high degree of uncertainty in the estimates. Figure 4.4 summarises the standardised scores on the ACE3 sub domains, to facilitate comparisons because of the different maximum scores in different domains. In general people with dementia had lower scores in all ACE3 sub domains than people with MCI who in turn had generally lower scores than people with normal cognition. People with dementia had median standardised scores in memory that were lower than (sequentially) those in visuospatial, fluency, language and attention. In contrast people with MCI had median standardised scores in fluency that were lower than (sequentially) those in visuospatial, attention, memory and language domains. In turn, people with normal cognition had

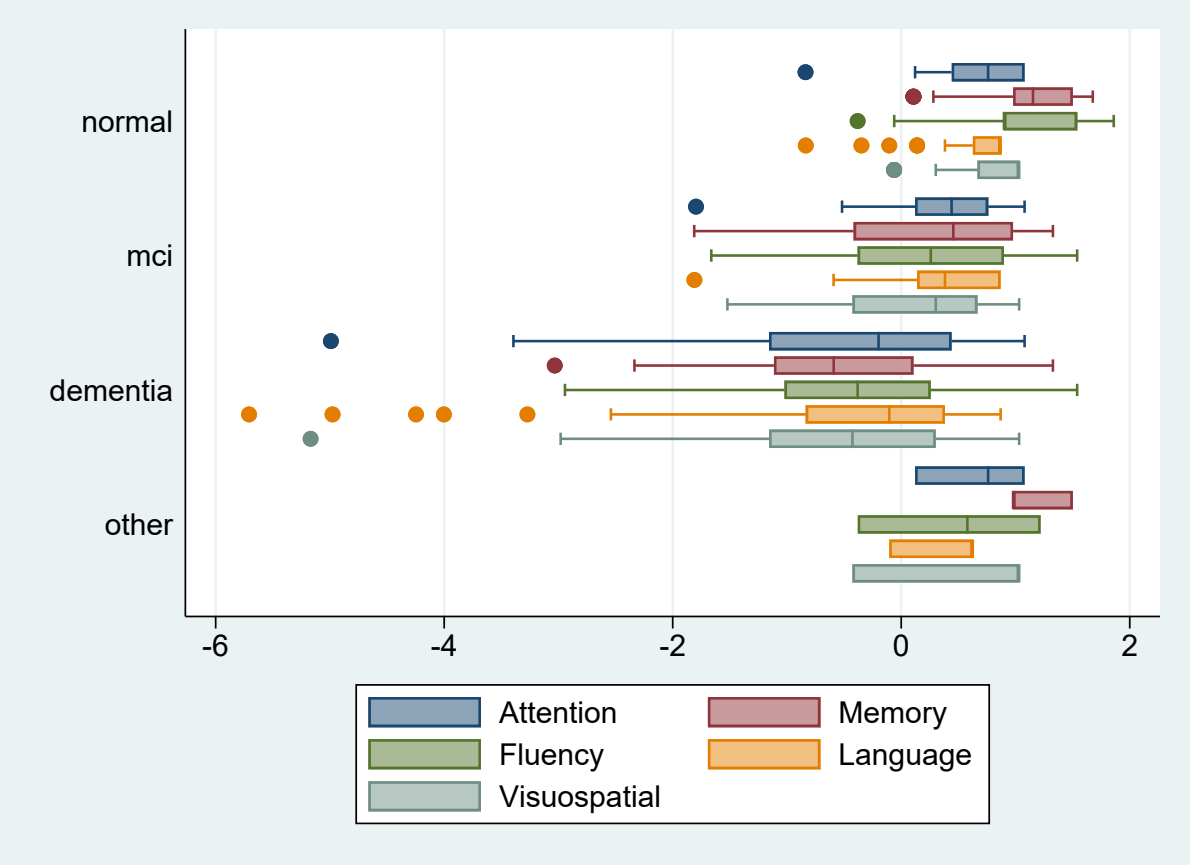
Table 4.4: Mean difference in ACE3 sub domain scores, compared to normal cognition

	Cognitive diagnosis ∗			
	Dementia	**MCI**	**Other**	**Normal**
ACE3 domain	*n=131*	*n=59*	*n=3*	*n=47*
	Compared to normal (95% CI) †			$\bar{x}$ (s) ‡
Total	-26 (-29 to -22)	-12 (-16 to -8.0)	-7.4 (-19 to 4.7)	92.0 (4.8)
Attention	-3.9 (-4.7 to -3.0)	-0.9 (-2.0 to 0.1)	-0.3 (-3.4 to 2.8)	17.0 (1.2)
Memory	-9.9 (-11 to -8.5)	-5.3 (-7.0 to -3.7)	-0.1 (-5.0 to 5.0)	23.1 (2.4)
Fluency	-4.8 (-5.6 to -3.9)	-2.7 (-3.7 to -1.7)	-1.8 (-4.8 to 1.1)	11.5 (1.7)
Language	-4.2 (-5.5 to -3.0)	-1.3 (-2.7 to 0.1)	-1.1 (-5.4 to 3.3)	25.0 (1.5)
Visuospatial	-3.2 (-4.1 to -2.4)	-1.8 (-2.7 to -0.8)	-0.7 (-3.6 to 2.1)	15.4 (1.0)

∗ *Dementia* according to ICD-10 definition *MCI* according to Peterson definition
Other impairment: not normal but impairment not meeting formal criteria
Due to low mood in one case, traumatic brain injury in one case,
and DSM–5 dementia not meeting criteria for ICD-10 dementia in one case
† *ACE3* Addenbrooke's Cognitive Examination III
‡ $\bar{x}$ mean *s* standard deviation

median standardised scores in the attention domain that were lower than their scores in language, fluency; and had highest standardised median scores in visuospatial and especially memory. People classified as other had similar scores in ACE3 sub-domains as those with normal cognition; this group had lowest standardised scores in the language domain, with serially higher standardised scores in fluency, visuospatial skills, attention, and highest standardised scores in memory. Scores in the language domain were most tightly clustered with the smallest IQR but with a number of outliers, with a similar pattern (though not distribution) observed in people who were normal and those with MCI or dementia. Conversely the fluency domain generally had a wide distribution of scores with a wide IQR.

Figure 4.4: Standardised scores on ACE3 sub domains



The box plots the median (darker middle line) and the quartiles (box edges), the whiskers enclose the lower (Q1 - 1.5 × interquartile range) and upper (Q3 + 1.5 × interquartile range) adjacent values, and the dots mark the outlying values.

4.3 Characteristics of tests

Table 4.5 summarises the characteristics of the 12 patient completed index tests. Five tests were completed by all participants, and Minicog was derived from GPCOG, but MAT (n=34), Sniffin sticks (n=188), MOCA (Montreal Cognitive Assessment, n=206), 6CIT (Six Item Cognitive Impairment Test, n=236), TUG (Timed up-and-go, n=236), and Phototest (n=238) were not completed by all participants. MOCA and MAT had incomplete completion by participants because after data collection commenced Sniffin sticks were introduced and MOCA replaced MAT (see Section 3.4.1). Other missing data in the index tests was fully attributable to participants declining to complete the tests. Accounting for the changes to the case report form, 230 participants completed all of the index tests, nine people declined one test, and one person declined two tests. There was no evidence of association between cognition and the number of declined tests, with beta coefficients (β) for MCI and dementia, compared to normal, of -0.03 (95% CI -0.11 to 0.06) and 0.02 (95% CI -0.06 to 0.10) respectively, or between the number of declined tests and TUG as a indicator of physical ill health β -0.01 (95% CI -0.04 to 0.03). Similarly in logistic regression analyses the odds ratio for MCI was 0.39 (95% CI 0.03 to 4.4) and for dementia the odds ratio was 1.3 (95% CI 0.25 to 6.3); whereas the odds ratio for the TUG was 0.97 (95% CI 0.82 to 1.14).

As shown in Table 4.5 scores on index tests were lower in people with dementia than those with MCI or normal cognition. For all index tests except one (MAT) there was evidence against the null hypothesis of no difference in mean index test score compared to people with normal cognition, for people with both dementia and MCI. People with dementia had lower mean scores on MAT compared to normal, but there was no evidence against the null hypothesis of no difference in mean MAT score for people with MCI compared to people with normal cognition.

Table 4.5: Characteristics of tests

Index test †	**Cognitive diagnosis** ∗ **Dementia** *n=131*	**MCI** *n=59*	**Other** *n=3*	**Normal** *n=47*
	Compared to normal			$\bar{x}$ (s) ‡
MAT *n=34* [0-50] ⇑ •• <28 §				
Test score (95% CI)	-14 (-20 to -6.6)	-7.0 (-15 to 1.2)	2.5 (-12 to 17)	39 (3.7)
Test duration (95% CI)	3.3 (1.6 to 5.1)	1.9 (-0.20 to 4.0)	0.25 (-3.4 to 3.9)	4.8 (0.50)
MOCA *n=206* [0-30] ⇑ •• <26				
Test score (95% CI)	-8.7 (-10 to -7.2)	-3.8 (-5.6 to -2.1)	-4.5 (-11 to 1.6)	24 (2.7)
Test duration (95% CI)	2.6 (1.8 to 3.3)	1.3 (0.46 to 2.2)	2.1 (-0.92 to 5.2)	9.4 (1.4)
Eurotest *n=240* [0-35] ⇑ •• <21				
Test score (95% CI)	-11 (-13 to -9.1)	-4.4 (-6.6 to -2.2)	-5.4 (-12 to 1.2)	27 (3.5)
Test duration (95% CI)	1.9 (1.5 to 2.4)	1.1 (0.52 to 1.6)	1.7 (0.0053 to 3.4)	5.6 (1.2)
Time & Change *n=240* ⇑ •• <2				
Pass (n, col %)	95 (73)	55 (93)	3 (100)	47 (100)
Test duration (95% CI)	0.56 (0.32 to 0.81)	0.27 (-0.02 to 0.56)	0.20 (-0.67 to 1.1)	1.5 (0.55)
Phototest *n=238* [0-∞] ⇑ •• <27				
Test score (95% CI)	-12 (-15 to -10)	-6.2 (-8.9 to -3.5)	-3.6 (-12 to 4.7)	37 (6.2)
Test duration (95% CI)	0.22 (0.0079 to 0.43)	0.25 (0.0048 to 0.48)	0.20 (-0.54 to 0.93)	2.8 (0.54)
SPMT *n=240* [0-23] ⇑ •• <10				
Test score (95% CI)	-7.9 (-9.2 to -6.6)	-3.9 (-5.4 to -2.4)	-1.5 (-6.1 to 3.0)	15 (3.5)
Test duration (95% CI)	0.015 (-0.46 to 0.49)	0.0037 (-0.54 to 0.55)	-0.23 (-1.9 to 1.4)	6.6 (1.2)
6CIT *n=238* [0-28] ⇓ •• >7				
Test score (95% CI)	9 .9 (8.1 to 12)	5.1 (2.9 to 7.2)	3.7 (-2.7 to 10)	2.6 (2.9)
Test duration (95% CI)	0.52 (0.30 to 0.75)	0.25 (-0.013 to 0.53)	-0.063 (-0.89 to 0.76)	1.1 (0.64)
Minicog *n=240* [0-5] ⇑ •• <3				
Test score (95% CI)	-2.6 (-3.0 to -2.1)	-1.4 (-2.0 to -0.88)	-0.80 (-2.4 to 0.82)	4.5 (0.8)
Duration not available as calculated				
GPCOG *n=240* [0-9] ⇑ •• ∗∗				
Test score (95% CI)	-4.4 (-5.1 to -3.7)	-2.4 (-3.2 to -1.5)	-1.1 (-3.6 to 1.4)	8.1 (1.2)
Test duration (95% CI)	1.1 (0.80 to 1.5)	0.64 (0.26 to 1.0)	0.57 (-0.60 to 1.7)	1.8 (0.60)
TUG *n=236* [0-∞] ⇓ •• > 15				
Test time, seconds (95% CI)	4.7 (2.6 to 6.8)	2.7 (0.3 to 5.1)	1.5 (-5.8 to 8.8)	8.8 (2.1)
Test duration (95% CI)	0.35 (-0.011 to 0.71)	0.16 (-0.25 to 0.58)	-0.70 (-2.0 to 0.57)	2.4 (1.1)
EPSS *n=240* [0-28] ⇓ •• > 1				
Test score (95% CI)	3.3 (2.2 to 4.4)	2.1 (0.87 to 3.4)	-0.34 (-4.2 to 3.5)	2.3 (1.8)
Test duration (95% CI)	0.18 (-0.048 to 0.41)	-0.063 (-0.33 to 0.20)	0.021 (-0.78 to 0.83)	1.0 (0.53)

Index test †	Cognitive diagnosis ∗ Dementia *n=131*	MCI *n=59*	Other *n=3*	Normal *n=47*
	Compared to normal			$\bar{x}$ (s) ‡
Sniffin sticks *n=188* [0-16] ⇑ •• < 11				
Test score (95% CI)	-2.2 (-3.1 to 1.3)	-1.5 (-2.6 to -0.59)	-1.4 (-4.8 to 2.0)	7.9 (2.6)
Test duration (95% CI)	0.42 (-0.053 to 0.89)	0.65 (0.11 to 1.2)	0.52 (-1.3 to 2.4)	4.0 (1.0)

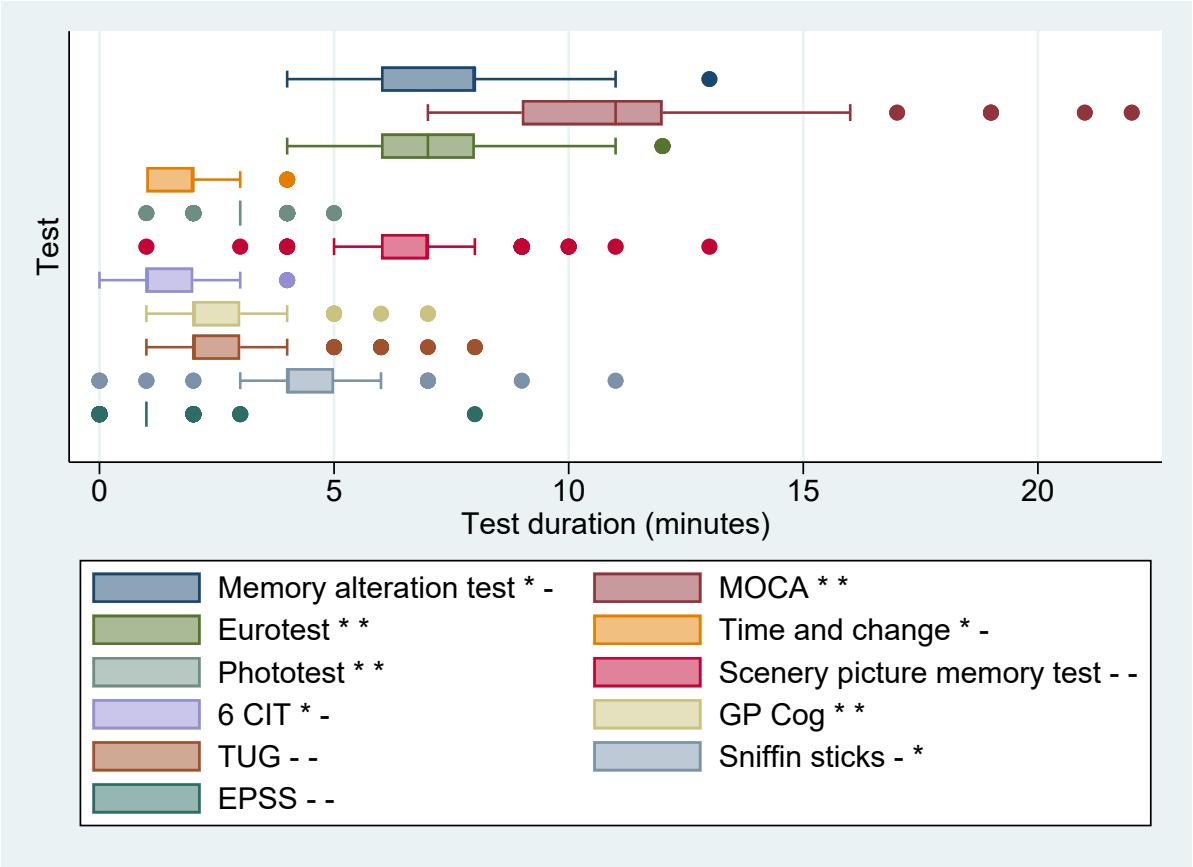
∗ *Dementia* according to ICD-10 definition *MCI* according to Peterson definition
Other impairment: not normal but impairment not meeting formal criteria, see Figure 4.1
† *MAT* Memory alteration test
MOCA Montreal Cognitive Assessment *SPMT* Scenery Picture Memory Test
6CIT Six Item Cognitive Impairment Test
GPCOG General Practitioner Assessment of Cognition (± *i* informant)
TUG Timed up-and-go *EPSS* Extra pyramidal signs scale ‡ $\bar{x}$ mean *s* standard deviation
§ *[n-n]* range of possible scores ⇑ higher scores indicate better cognition
⇓ lower scores indicate better cognition •• *threshold for dementia*
∗∗ for GPCOG a two stage approach to scoring is used whereby scores >8 indicate normal and <5 indicate abnormal and scores 5-8 indicate GPCOGi needed, where scores <4 indicate abnormal
Duration for minicog not given because the score was calculated using items from the GPCOG

Figure 4.5 displays the duration of the index tests that were done by participants who were people with cognitive symptoms (patients), times of the tests are also provided in Table 4.5. Mini-cog is not included in the figure because the score was calculated using items from the GPCOG (General Practitioner Assessment of Cognition). The MOCA had the longest median duration (11 minutes) of the tests and the 6CIT (1 minute) had the shortest duration. Six tests took longer in people with dementia than people with normal cognition, these were: MAT, MOCA, Eurotest, Phototest, 6CIT, and GPCOG. In contrast four tests took longer in people with MCI than people with normal cognition: MOCA, Eurotest, GPCOG, and Sniffin sticks.

As shown by the outliers in Figure 4.5, some participants took much longer to complete tests. The tests were classified by 95th centile of duration (see Section 3.7), the short tests had duration <5 minutes, medium tests ≥ 5 minutes but ≤ 10 minutes, and longer tests (> 10 minutes). The short duration tests were: EPSS (C_{95} 2 minutes), TAC (C_{95} 3 minutes), Phototest (C_{95} 4 minutes), 6CIT (C_{95} 3 minutes), GPCOG (C_{95} 4 minutes). The medium duration tests were: Eurotest (C_{95} 10 minutes), SPMT (C_{95} 9 minutes), TUG (C_{95} 5 minutes), Sniffin sticks (C_{95} 6 minutes). The long duration tests were: MAT (C_{95} 11 minutes), MOCA (C_{95} 15 minutes).

Characteristics of GP clinical judgement are not presented in either Table 4.5 or Figure 4.5 because there was no test score and test duration was not known. A summary of the GP opinions is provided in Table 4.2.

Figure 4.5: Duration of index tests, all cognition categories



The box plots the median (darker middle line) and the quartiles (box edges), the whiskers enclose the lower (Q1 - 1.5 × interquartile range) and upper (Q3 + 1.5 × interquartile range) adjacent values, and the dots mark the outlying values. - indicates no evidence and ∗ indicates evidence against the null hypothesis of no difference in mean test duration, compared to people with normal cognition, for (respectively) people with dementia and people with MCI

4.4 Target conditions

As described in Section 3.1.1 the two target conditions for evaluating diagnostic accuracy were dementia and normal. When assessing the diagnostic accuracy of the index tests, whether as single tests or in combination, the participants were classified as follows:

- Dementia as target condition
  - Disease positive: 131 people with reference standard diagnosis of ICD-10 dementia and one person with reference standard diagnosis of other on ICD-10 who had DSM–5 frontotemporal dementia but no objective memory impairment. *Total 132 people*.
  - Disease negative: 47 people with reference standard diagnosis of ICD-10 normal, 59 people with reference standard diagnosis of Peterson MCI, and two people with reference standard diagnosis of other on ICD-10 of whom one had an affective disorder and one had acquired brain injury. *Total 108 people*.
- Normal as target condition
  - Disease positive: 47 people with reference standard diagnosis of ICD-10 normal
  - Disease negative: 131 people with reference standard diagnosis of ICD-10 dementia, 59 people with reference standard diagnosis of ICD-10 MCI, and three people with reference standard diagnosis of other on ICD-10. *Total 193 people*.

4.4.1 Prevalence of target condition

With dementia as the target condition there were 132 disease positives, for a prevalence of 55% (95% CI 48% to 61%). With normal as the target condition there were 47 target-condition positives, meaning that 193 people had cognitive impairment, for a prevalence of 80% (95% CI 75% to 85%). Table 4.6 cross tabulates the judgement of GPs against the cognitive category according to the reference standard. Of the 34 people who GPs judged as having normal cognition, 19 (56%) were normal according to the reference standard, 8 (24%) had MCI and 6 (25%) had dementia. In contrast, of the 86 people who GPs judged as having dementia, 1 (1%) had normal cognition according to the reference standard, 11 (13%) had MCI and 74 (86%) had dementia.

4.5 Informant tests

Five informant measures were completed as standalone index tests: Functional Activities Questionnaire (FAQ [327]), Lawton instrumental activities of daily living scale (IADL [328]), Katz index of activities of daily living (ADL [329]), Galvin AD8 Dementia Screening Interview (AD8 [299]), and Short Form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE [304]). The

Table 4.6: GP judgement against target conditions

	Reference standard			
GP judgement	**Normal**	**MCI**	**Dementia**	**Other**
Normal	19	8	6	1
CIND	27	40	51	2
Dementia	1	11	74	0

MCI Mild Cognitive Impairment
CIND cognitive impairment, but not dementia
ICD-10 definition for Dementia and Peterson for MCI

five measures which were completed as standalone index tests were completed both by independent completion of a paper questionnaire (first), and shortly afterwards (within 30 minutes) by GP researcher interview.

Kappa Table 4.7 shows the results when weighted kappa was calculated for the whole instrument for each of the five tests, both as continuous measures and as a categorical test at the published threshold. For all tests, weighted kappa was≥ 89% when comparing informant completed tests to those done by interview.

Diagnostic accuracy As indicated by the DOR provided in Table 4.7 and the AUROC shown in Figure 4.6, the informant measures had similar accuracy for the diagnosis of dementia regardless of whether the test was completed by the interviewer or the informant.

Analysis of diagnostic accuracy of informant measures Based on the high weighted kappa between informant and interviewer completed measures for all of the informant tests, and the comparable diagnostic accuracy based on the odds ratio, the informant completed measures were used in preference to calculate diagnostic accuracy for single tests and in combinations, because these would be less burdensome for completion in clinical practice.

Table 4.7: Characteristics of informant tests

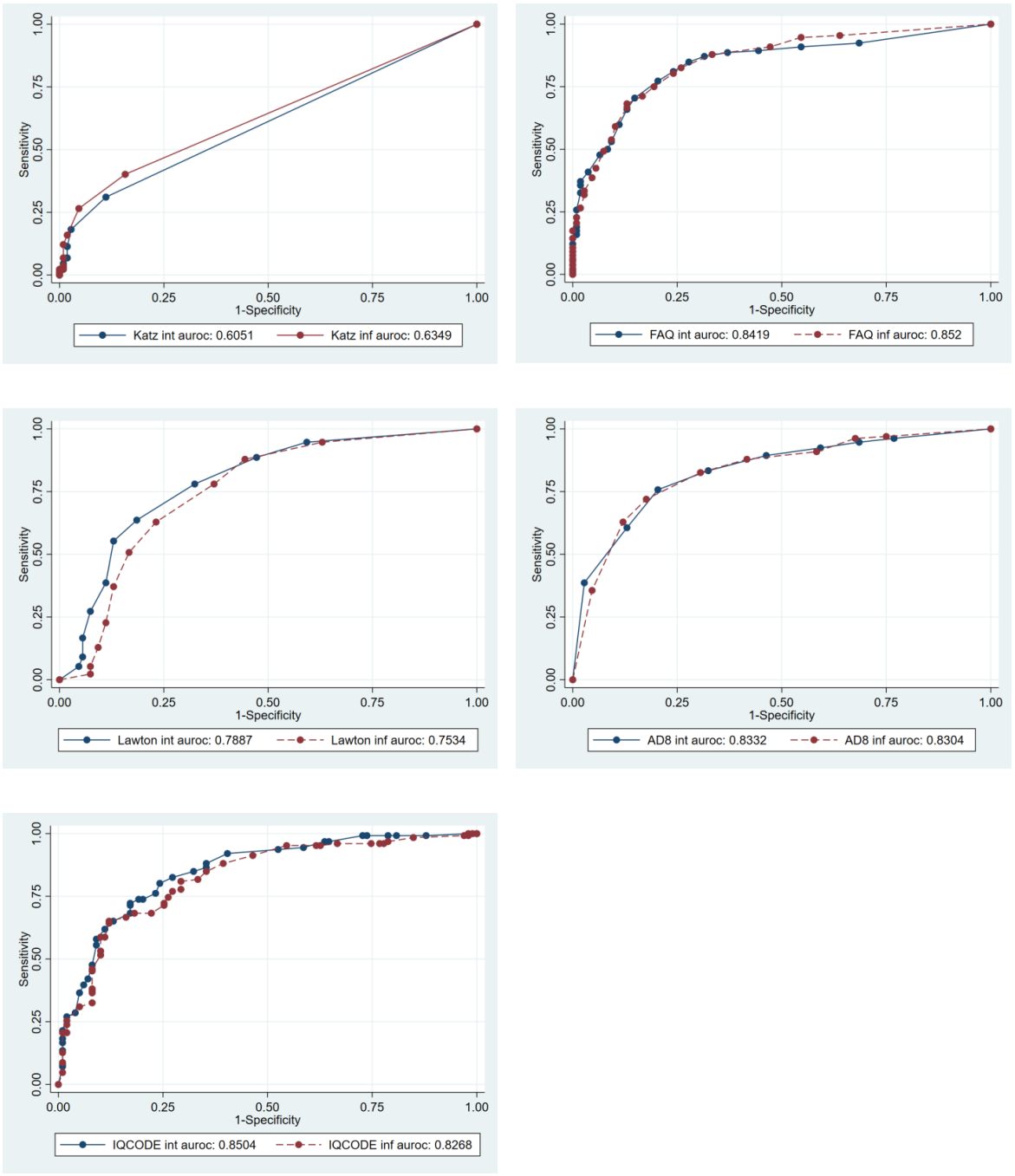
Test	Agreement	Expected Agreement	Kappa	Standard Error	z	p>z
Continuous measures						
ADL	96%	89%	0.6592	0.0433	15.2	<0.001
FAQ	91%	70%	0.7071	0.0407	17.4	<0.001
IADL	92%	67%	0.7673	0.0408	18.8	<0.001
AD8	89%	61%	0.7077	0.0435	16.3	<0.001
IQCODE	93%	68%	0.7727	0.0422	18.3	<0.001
Categorical measures at published threshold $*$						
ADL	93%	76%	0.7202	0.0632	11.4	<0.001
FAQ	90%	72%	0.6402	0.0644	9.9	<0.001
IADL	91%	61%	0.7632	0.0645	11.8	<0.001
AD8	90%	72%	0.6287	0.0645	9.7	<0.001
IQCODE	91%	70%	0.7074	0.0666	10.6	<0.001

Diagnostic accuracy for dementia at published threshold $*$

Informant completed	Odds ratio	95% CI	Interviewer completed	Odds ratio	95% CI
ADL	7.8	2.4 to 25	ADL	7.2	2.8 to 19
FAQ	15	4.7 to 46	FAQ	13	4.0 to 40
IADL	5.0	2.5 to 10	IADL	4.0	2.1 to 7.6
AD8	8.2	3.5 to 19	AD8	12	4.7 to 31
IQCODE	18	6.6 to 52	IQCODE	13	5.2 to 31

$*$ based on following threshold indicating abnormal: *ADL* >1 [259]
FAQ dependent in >2 activities [327] *IADL* <5 [259] *AD8* >1 [349] *IQCODE* > 3.3 [350]

Figure 4.6: ROC curves for informant tests



The plots show the comparative AUROC for informant and interviewer completed tests: ADL, FAQ, IADL, AD8, and IQCODE. Completed by int *Interviewer (full lines) and* inf *informant (dashed lines).* AUROC area under ROC curve

4.6 Diagnostic accuracy of tests

The evaluated index tests are listed in Chapter 3 Section 3.7. Test accuracy is numerically associated with prevalence of disease, and in particular positive predictive values are positively correlated with higher prevalence of disease in the study sample (Section 1.1.2.3). The impact of these considerations is discussed in Chapter 6, for example Section 6.4.3 considers how test accuracy could vary with prevalence. Chapter 6 also discusses important matters relating to comparisons between tests; for example in particular Section 6.2.6 emphasises that the overlapping confidence intervals for single and combined tests means that apparent differences in test accuracy could be chance findings, especially given the multiple comparisons. Confidence intervals for accuracy are important because based on point estimates the accuracy of test A may be lower than test B, but if test B has a narrower confidence interval (indicating greater certainty) that is generally higher than test A, then this may support the use of test B in preference to test A.

4.6.1 Classification accuracy

Table 4.8 shows the error rate (calculated as $\frac{FP+FN}{TP+TN+FP+FN}$) for each index test, with the classification accuracy being calculated as 1-error rate. Accounting for variation in the denominator, the test that correctly classified most people for dementia as the target condition was SPMT which correctly classified 78% (186 people) (56 TN (true negative); 123 TP (true positive)) and incorrectly classified 54 people (30 FN (false negative); 24 FP (false positive)). In addition to SPMT, with dementia as the target condition, six of the 17 index tests correctly classified at least 70% of participants: Eurotest correctly classified 75% (87 TN, 93 TP), GPCOG correctly classified 75% (56 TN, 123 TP), 6CIT correctly classified 73% (75 TN, 99 TP), clinical judgement correctly classified 71% (96 TN, 74 TP) IQCODE correctly classified 71% (38 TN, 125 TP), and Minicog correctly classified 71% (79 TN, 92 TP). In contrast, with dementia as the target condition, the tests that incorrectly classified the highest proportion of participants were Sniffin sticks 46% (0 FN, 86 FP) and TUG 46% (99 FN, 10 FP).

The test that correctly classified most people with normal as the target condition was GPCOG which correctly classified 85% (204 people; 166 TN, 38 TP) and incorrectly classified 36 people (9 FN, 27 FP). In addition to GPCOG, with normal cognition as the target condition, four of the 17 index tests correctly classified ≥ 80% of people: GP clinical judgement, MOCA, and IQCODE correctly classified 82% (respectively 178, 158 and 166 TN; 19, 11 and 24 TP), and AD8 correctly classified 80% (172 TN, 19 TP). In contrast, with normal as the target condition, the tests that incorrectly classified the highest proportion of participants were FAQ 65% (0 FN, 155 FP), TUG and TAC 64% (respectively 1 and 0 FN, 151 and 153 FP).

4.6.2 Dementia as target condition: single tests

Table 4.10 shows the accuracy of single tests for diagnosing dementia and Figure 4.9 shows the diagnostic accuracy of the different tests in natural frequencies; the true results (TP and TN) are

Table 4.8: Error rate of single tests for diagnosing dementia and normal cognition

	Error rate % (95% CI) for Target condition	
Test	**Dementia**	**Normal**
EPSS	40 (34 to 46)	25 (19 to 31)
TAC	42 (35 to 48)	64 (57 to 70)
Phototest	32 (26 to 38)	45 (38 to 51)
6CIT	27 (21 to 33)	27 (22 to 33)
GPCOG	25 (20 to 31)	15 (11 to 20)
Minicog	29 (23 to 35)	32 (26 to 38)
Eurotest	25 (20 to 31)	35 (29 to 41)
SPMT	23 (17 to 28)	30 (24 to 36)
TUG	46 (40 to 53)	64 (58 to 71)
Sniffin sticks	46 (38 to 53)	21 (15 to 27)
MAT	32 (17 to 51)	38 (22 to 56)
MOCA	40 (33 to 47)	18 (13 to 24)
GP judgement	29 (23 to 35)	18 (13 to 23)
FAQ	42 (35 to 48)	65 (58 to 71)
ADL	43 (36 to 49)	64 (58 to 70)
IADL	40 (34 to 47)	55 (48 to 61)
AD8	33 (27 to 39)	20 (16 to 26)
IQCODE	29 (24 to 36)	18 (13 to 23)

∗ *TAC* Time and change *6CIT* Six Item Cognitive Impairment Test
GPCOG General Practitioner Assessment of Cognition
SPMT Scenery Picture Memory Test *TUG* Timed up-and-go *MAT* Memory alteration test
MOCA Montreal Cognitive Assessment *FAQ* Functional Activities Questionnaire
ADL Katz index of activities of daily living
IADL Lawton instrumental activities of daily living scale

plotted above the line of zero and the false results (FP and FN) are plotted below the line.

For dementia as the target condition, positive likelihood ratios for single tests ranged from 1 (95% CI 1 to 1) for EPSS, MOCA, Sniffin sticks, and AD8 (which had an upper 95% CI of 2 rather than 1) to 10 (95% CI 3 to 30) for FAQ. Other single tests with a LRP of five or more [1] were GP judgement LRP 5 (95% CI 3 to 9), ADL LRP 6 (95% CI 2 to 14) and TAC LRP 7 (95% CI 3 to 20). Of these no test had an LRP 95% CI lower bound of more than 3, and confidence intervals overlapped, indicating no real evidence of superiority for any one test.

In contrast, negative likelihood ratios for single tests ranged from 0 (95% CI incalculable) for MOCA and Sniffin sticks to 0.85 for TUG (95% CI 0.76 to 0.95). Other single tests with an LRN of less than 0.2 [2] were GPCOG LRN 0.13 (95% CI 0.07 to 0.25), AD8 LRN 0.12 (95% CI 0.05 to 0.29), and IQCODE LRN 0.12 (95% CI 0.05 to 0.27). Of these only GPCOG, AD8 and IQCODE had an LRN 95% CI upper bound of less than 0.30, providing some support for using these three tests in preference to

[1] an LRP of 5 corresponds approximately to a 30% increase in probability of disease if test positive [358]
[2] an LRN of 0.20 corresponds approximately to a 30% reduction in the probability of disease if test negative [358]

others for ruling out dementia; however the confidence intervals overlapped between these tests and others indicating no real evidence for superiority of any one single test.

Chapter 2 described a systematic review of the accuracy of GP clinical judgement for the diagnosis of dementia. Table 4.10, Figure 4.7 and Figure 4.8 show the updated meta-analysis of the accuracy of GP clinical judgement for the diagnosis of dementia and cognitive impairment. The data from Chapter 2 are included and updated with the new data. Overall the updated accuracy including the new data is very similar to the findings of the systematic review.

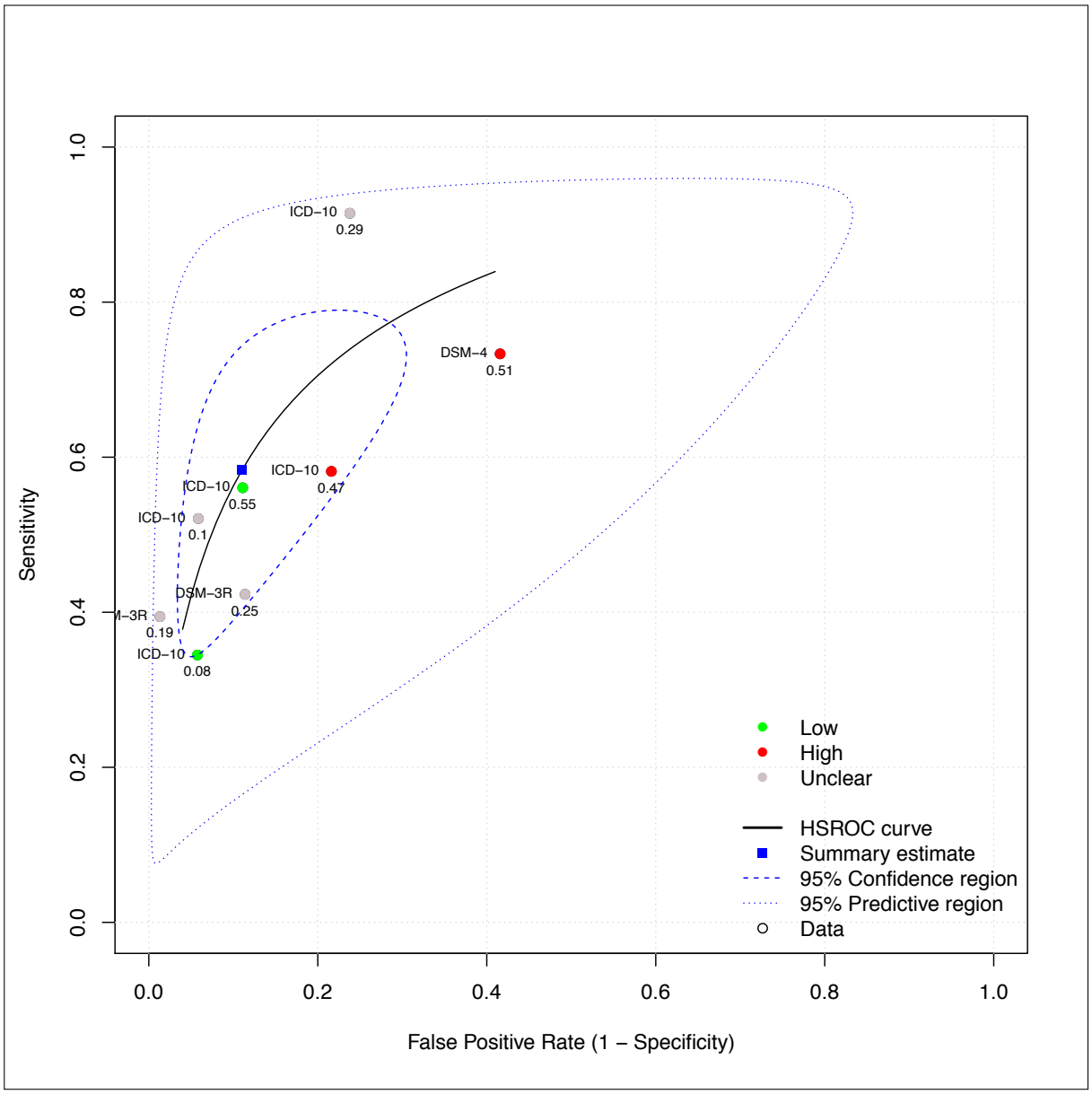
The updated meta-analysis of the accuracy for GP judgement for the diagnosis of dementia was: sensitivity 58% (95% CI 43% to 72%) specificity 89% (95% CI 79% to 95%) which is very similar to the results in Chapter 2: sensitivity 59% (95% CI 41% to 74%) specificity 89% (95% CI 77% to 95%). In contrast the updated meta-analysis of the accuracy for GP judgement for the diagnosis of cognitive impairment was sensitivity 84% (95% CI 60% to 95%) specificity 73% (95% CI 50% to 88%), compared to Chapter 2 sensitivity 80% (95% CI 45% to 95%) specificity 79% (95% CI 57% to 92%). Figure 4.7 and 4.8 show the summary ROC curves including the new data and indicate that for dementia as the target condition the new findings lie very close to the summary point and near to the summary ROC curve. In contrast, for the cognitive impairment target condition the new findings lie close to the summary ROC curve but are much further from the summary point, with a lower specificity and higher sensitivity. Chapter 6 discusses possible reasons for this, discusses the new findings in the context of the systematic review in Chapter 2 in more depth, and considers the extent to which the results can be attributed to differences in study methodology or chance. As described in Chapter 2 there is substantial heterogeneity in the data and this prevents the ability to draw any firm conclusions.

As described in Section 3.7.4.1, Decision curves for all of the single tests are provided in Appendix H.

Table 4.9: Updated meta-analysis of GP judgement for diagnosis of dementia and cognitive impairment

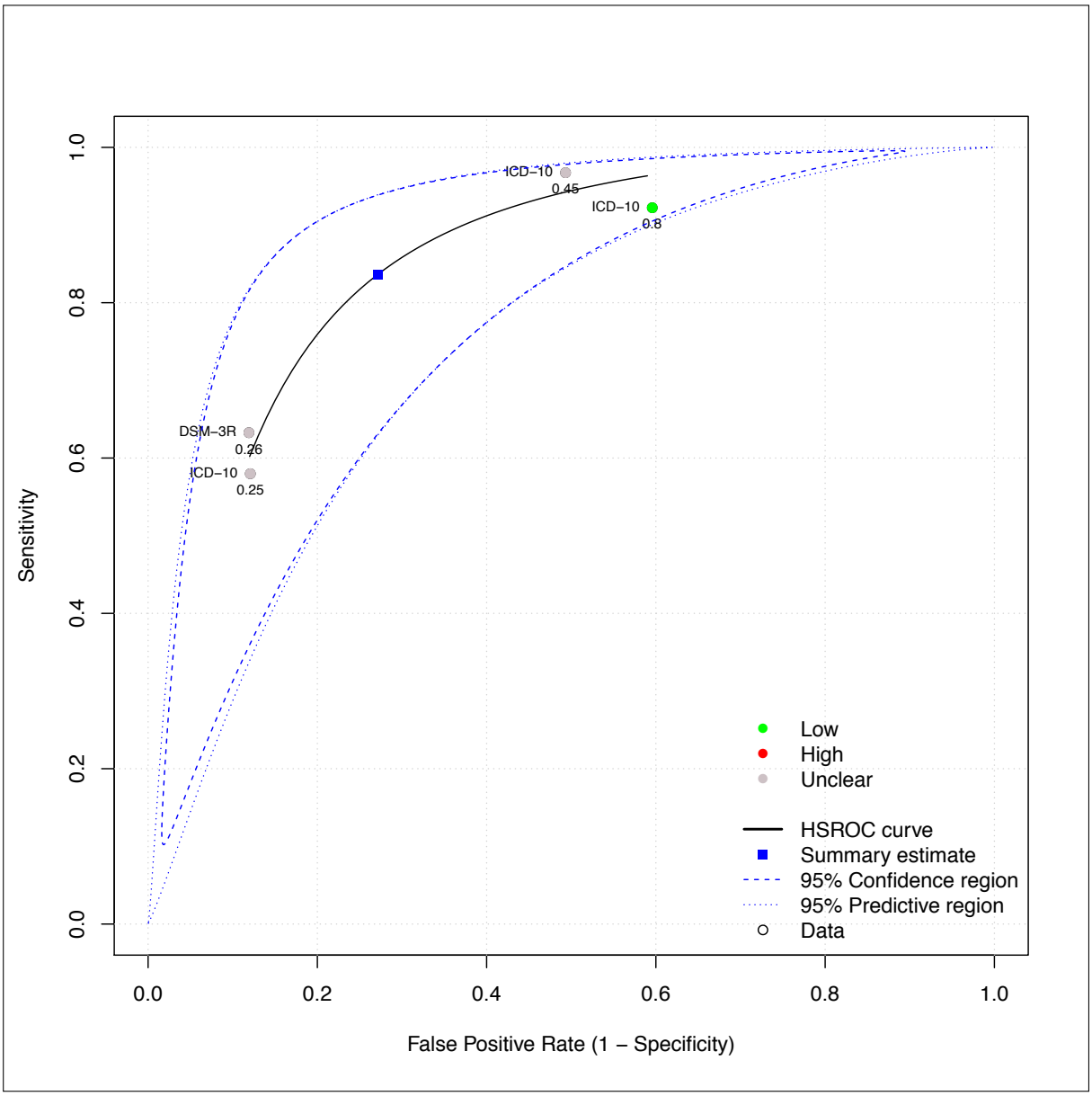
Analysis	**Studies**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**DOR (95% CI)**
Dementia as target condition				
Chapter 2	7	59 (41 to 74)	89 (77 to 95)	11 (6 to 22)
New findings	1	56 (47 to 65)	89 (81 to 94)	10 (5 to 22)
Updated	8	58 (43 to 72)	89 (79 to 95)	11 (5 to 18)
Cognitive impairment as target condition				
Chapter 2	3	80 (45 to 95)	79 (57 to 92)	15 (8 to 29)
New findings	1	92 (88 to 96)	40 (26 to 56)	8 (4 to 18)
Updated	4	84 (60 to 95)	73 (50 to 88)	14 (8 to 19)

Figure 4.7: Updated summary plot of clinical judgement of GPs for diagnosis of dementia



Colours of dots: indicate risk of bias in flow and timing domain; decimals indicate prevalence of target condition
new findings: the study at low risk of bias and prevalence 0.55 (55%)

Figure 4.8: Updated summary plot of clinical judgement of GPs for diagnosis of cognitive impairment



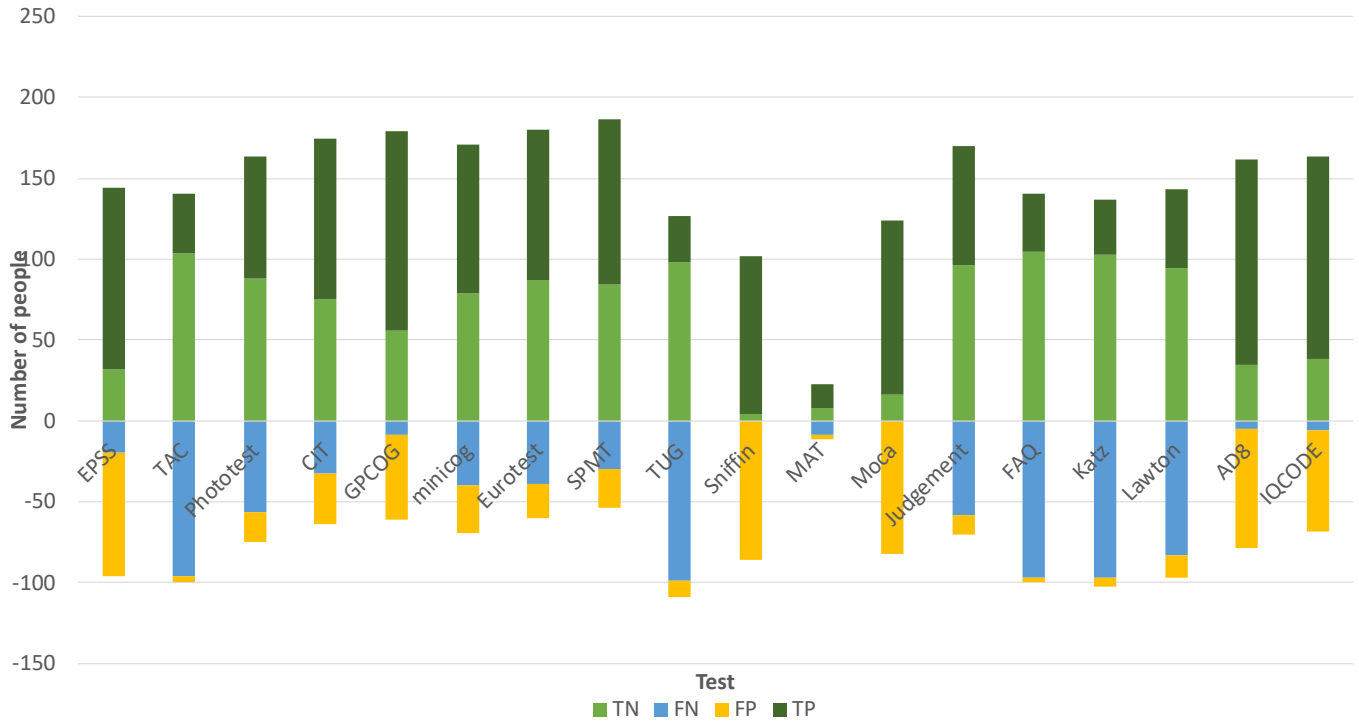
Colours of dots: indicate risk of bias in flow and timing domain; decimals indicate prevalence of target condition
new findings: the study at low risk of bias and prevalence 0.80 (80%)

Table 4.10: Diagnostic accuracy of single tests for dementia

Test	Sensitivity (95% CI)	Specificity (95% CI)	PPV ∗ (95% CI)	NPV (95% CI)	LRP (95% CI)	LRN (95% CI)	DOR (95% CI)
Brief tests †							
EPSS	85 (78 to 90)	30 (21 to 39)	60 (52 to 67)	62 (47 to 75)	1.2 (1.0 to 1.4)	0.51 (0.31 to 0.84)	2 (1 to 5)
TAC	27 (20 to 36)	96 (91 to 99)	90 (76 to 97)	52 (45 to 59)	7.4 (2.7 to 20)	0.76 (0.68 to 0.84)	10 (3 to 39)
Phototest	57 (48 to 66)	82 (74 to 89)	80 (70 to 87)	61 (53 to 69)	3.2 (2.1 to 5.0)	0.52 (0.42 to 0.65)	6 (3 to 12)
6CIT	76 (67 to 83)	70 (60 to 79)	76 (67 to 83)	70 (60 to 79)	2.5 (1.9 to 3.4)	0.35 (0.25 to 0.48)	7 (4 to 13)
GPCOG	93 (87 to 97)	52 (42 to 62)	70 (63 to 77)	86 (75 to 93)	1.9 (1.6 to 2.4)	0.13 (0.07 to 0.25)	15 (7 to 36)
Minicog	70 (61 to 77)	73 (64 to 81)	76 (67 to 83)	66 (57 to 75)	0.4 (0.3 to 0.5)	0.41 (0.31 to 0.55)	6 (3 to 11)
Medium duration tests							
Eurotest	70 (62 to 78)	81 (72 to 88)	82 (73 to 88)	69 (60 to 77)	3.6 (2.4 to 5.4)	0.37 (0.28 to 0.48)	10 (5 to 19)
SPMT	77 (69 to 84)	78 (69 to 85)	81 (73 to 87)	74 (65 to 81)	3.5 (2.4 to 5.0)	0.29 (0.21 to 0.41)	12 (6 to 23)
TUG	23 (16 to 31)	91 (84 to 95)	74 (58 to 87)	50 (43 to 57)	2.4 (1.3 to 4.8)	0.85 (0.76 to 0.95)	3 (1 to 7)
Sniffin sticks	100 (96 to 100)	4 (1 to 11)	53 (46 to 61)	100 (40 to 100)	1.0 (1.0 to 1.1)	0 (. to .)	. (1 to .)
Longer tests							
MAT	63 (41 to 81)	80 (44 to 97)	88 (64 to 99)	47 (23 to 72)	3.1 (0.9 to 11)	0.47 (0.26 to 0.86)	7 (1 to 74)
MOCA	100 (97 to 100)	16 (10 to 25)	57 (49 to 64)	100 (79 to 100)	1.2 (1.1 to 1.3)	0 (. to .)	. (5 to .)
Clinician and informant tests							
GP judgement	56 (47 to 65)	89 (81 to 94)	86 (77 to 93)	62 (54 to 70)	5 (2.9 to 8.8)	0.49 (0.4 to 0.61)	10 (5 to 22)
FAQ	27 (19 to 35)	97 (92 to 99)	92 (79 to 98)	52 (45 to 59)	9.5 (3.0 to 30)	0.76 (0.68 to 0.84)	13 (4 to 66)
ADL	26 (19 to 34)	95 (90 to 98)	87 (73 to 96)	52 (44 to 59)	2.5 (1.6 to 4.1)	0.78 (0.7 to 0.87)	7 (3 to 24)
IADL	37 (29 to 46)	87 (79 to 93)	78 (66 to 87)	53 (45 to 61)	2.9 (1.7 to 4.9)	0.72 (0.62 to 0.84)	4 (2 to 8)
AD8	96 (91 to 99)	32 (24 to 42)	64 (56 to 70)	88 (73 to 96)	1.4 (1.2 to 1.6)	0.12 (0.05 to 0.29)	12 (4 to 41)
IQCODE	95 (90 to 98)	38 (28 to 48)	67 (60 to 74)	86 (73 to 95)	1.5 (1.3 to 1.8)	0.12 (0.05 to 0.27)	13 (5 to 39)

∗ *TAC* Time and change *6CIT* Six Item Cognitive Impairment Test *GPCOG* General Practitioner Assessment of Cognition
SPMT Scenery Picture Memory Test *TUG* Timed up-and-go *MAT* Memory alteration test
MOCA Montreal Cognitive Assessment *FAQ* Functional Activities Questionnaire *ADL* Katz index of activities of daily living
IADL Lawton instrumental activities of daily living scale
† see Section 3.7.2 for test thresholds
. indicates incalculable results

Figure 4.9: Natural frequencies on tests for dementia



The figure plots the diagnostic accuracy of the index tests using natural frequencies. Plots above the line are true results and those below the line are false results.

4.6.3 Normal cognition as target condition: single tests

Table 4.11 shows the accuracy of single tests for diagnosing normal cognition and Figure 4.10 shows the diagnostic accuracy of the different tests in natural frequencies; the true results (TP and TN) are plotted above the line of zero and the false results (FP and FN) are plotted below the line.

For normal as the target condition, positive likelihood ratios for single tests ranged from 1 (95% CI 1 to 1) for TAC, TUG, FAQ, ADL, and IADL (which had an upper 95% CI of 2 rather than 1) to 11 (95% CI 1 to 101) for Sniffin sticks. Other single tests with a LRP of five or more were GP judgement LRP 5 (95% CI 3 to 9), IQCODE LRP 5 (95% CI 3 to 8) GPCOG LRP 6 (95% CI 4 to 8), and MOCA LRP 8 (95% CI 3 to 23). Of these only GPCOG had an LRP 95% CI lower bound of more than 3, providing some support for using this test in preference to others for ruling in normal; however the confidence intervals overlapped between this test and others indicating no real evidence for superiority of any one single test.

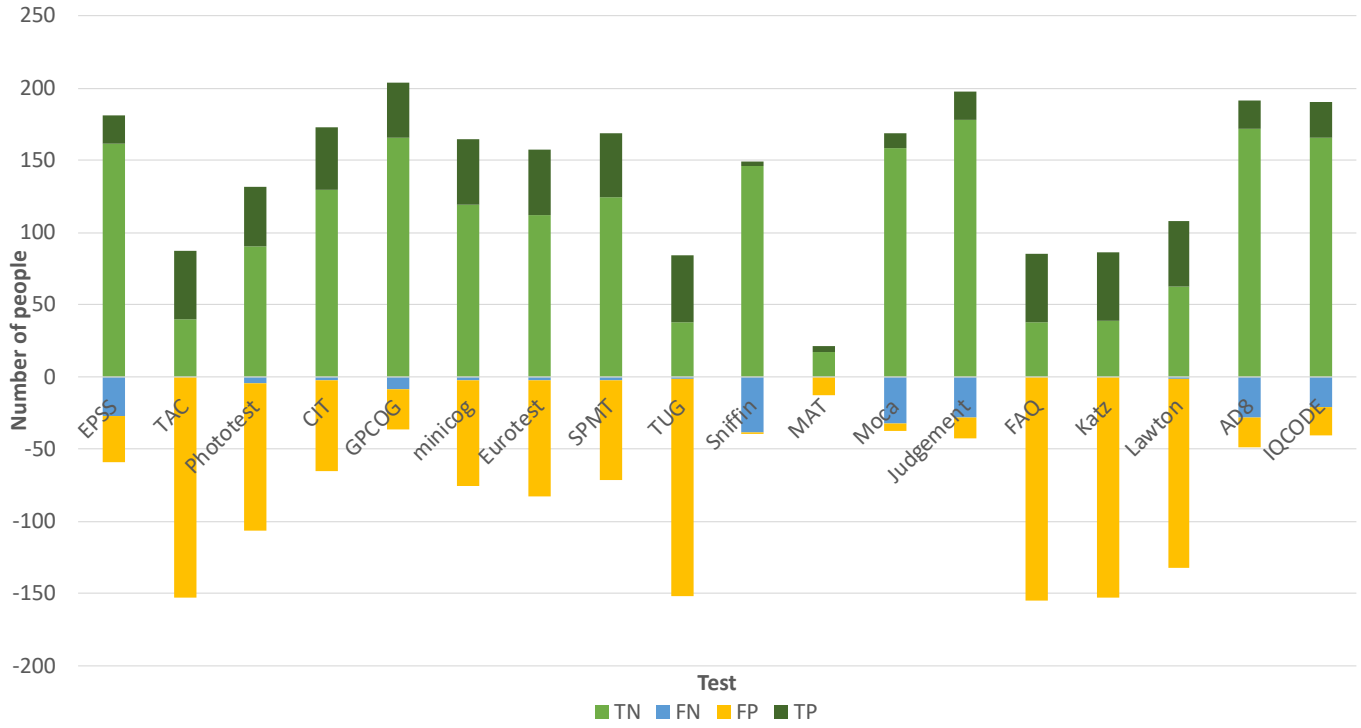
In contrast, negative likelihood ratios for single tests ranged from 0 (95% CI incalculable) for TAC, MAT, FAQ and ADL to LRN 0.93 for Sniffin sticks (95% CI 0.86 to 1.0). Other single tests with an LRN of less than 0.20 were Phototest LRN 0.19 (95% CI 0.07 to 0.48), TUG LRN 0.11 (95% CI 0.01 to 0.75), Minicog LRN 0.07 (95% CI 0.02 to 0.27), Eurotest LRN 0.07 (95% CI 0.02 to 0.29), SPMT LRN 0.07 (95% CI 0.02 to 0.26), IADL LRN 0.07 (95% CI 0.01 to 0.47) and 6CIT LRN 0.06 (95% CI 0.02 to 0.25). Of these tests only Eurotest, SPMT and 6CIT had an LRN 95% CI upper bound of less than 0.30, providing some support for using these three tests in preference to others for ruling out normal; however the confidence intervals overlapped between these tests and others indicating no real evidence for superiority of any one single test.

Table 4.11: Diagnostic accuracy of single tests for normal cognition

Test	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LRP (95% CI)	LRN (95% CI)	DOR (95% CI)
Brief tests †							
EPSS	43 (28 to 58)	83 (77 to 88)	38 (25 to 53)	86 (80 to 90)	2.6 (1.6 to 4.1)	0.69 (0.53 to 0.89)	4 (2 to 8)
TAC	100 (92 to 100)	21 (15 to 27)	24 (18 to 30)	100 (91 to 100)	1.3 (1.2 to 1.4)	0 (. to .)	. (3 to .)
Phototest	91 (79 to 98)	47 (40 to 54)	29 (22 to 37)	96 (89 to 99)	1.7 (1.5 to 2.0)	0.19 (0.07 to 0.48)	9 (3 to 37)
6CIT	96 (85 to 99)	67 (60 to 74)	41 (32 to 51)	98 (95 to 100)	2.9 (2.4 to 3.6)	0.06 (0.02 to 0.25)	45 (11 to 390)
GPCOG	81 (67 to 91)	86 (80 to 91)	58 (46 to 71)	95 (90 to 98)	5.8 (4.0 to 8.4)	0.22 (0.12 to 0.4)	26 (11 to 67)
Minicog	96 (85 to 99)	62 (54 to 69)	38 (29 to 47)	98 (94 to 100)	0.1 (0.0 to 0.3)	0.07 (0.02 to 0.27)	36 (9 to 313)
Medium duration tests							
Eurotest	96 (85 to 99)	58 (51 to 65)	36 (27 to 45)	98 (94 to 100)	2.3 (1.9 to 2.7)	0.07 (0.02 to 0.29)	31 (8 to 269)
SPMT	96 (85 to 99)	64 (57 to 71)	39 (30 to 49)	98 (94 to 100)	2.7 (2.2 to 3.3)	0.07 (0.02 to 0.26)	40 (10 to 350)
TUG	98 (89 to 100)	20 (15 to 27)	23 (18 to 30)	97 (87 to 100)	1.2 (1.1 to 1.3)	0.11 (0.01 to 0.75)	12 (2 to 479)
Sniffin sticks	7 (2 to 20)	99 (96 to 100)	75 (19 to 99)	79 (73 to 85)	11 (1.1 to 101)	0.93 (0.86 to 1.02)	12 (1 to 609)
Longer tests							
MAT	100 (40 to 100)	57 (37 to 75)	24 (7 to 50)	100 (80 to 100)	2.3 (1.5 to 3.5)	0 (. to .)	. (1 to .)
MOCA	26 (14 to 41)	97 (93 to 99)	69 (41 to 89)	83 (77 to 88)	8.3 (3.1 to 23)	0.77 (0.64 to 0.92)	11 (3 to 42)
Clinician and informant tests							
GP judgement	40 (26 to 56)	92 (88 to 96)	56 (38 to 73)	86 (81 to 91)	5.2 (2.9 to 9.5)	0.65 (0.51 to 0.82)	8 (3 to 19)
FAQ	100 (92 to 100)	20 (14 to 26)	23 (18 to 30)	100 (91 to 100)	1.2 (1.2 to 1.3)	0 (. to .)	. (3 to .)
ADL	100 (92 to 100)	20 (15 to 27)	24 (18 to 30)	100 (91 to 100)	1.3 (1.1 to 1.5)	0 (. to .)	. (3 to .)
IADL	98 (89 to 100)	32 (26 to 39)	26 (20 to 33)	98 (91 to 100)	1.4 (1.3 to 1.6)	0.07 (0.01 to 0.47)	22 (4 to 892)
AD8	40 (26 to 56)	89 (84 to 93)	48 (32 to 64)	86 (80 to 90)	3.7 (2.2 to 6.3)	0.67 (0.53 to 0.85)	6 (2 to 12)
IQCODE	53 (38 to 68)	89 (84 to 93)	55 (39 to 70)	89 (83 to 93)	5.0 (3.0 to 8.1)	0.52 (0.38 to 0.72)	9 (4 to 21)

∗ *TAC* Time and change *6CIT* Six Item Cognitive Impairment Test *GPCOG* General Practitioner Assessment of Cognition
SPMT Scenery Picture Memory Test *TUG* Timed up-and-go *MAT* Memory alteration test
MOCA Montreal Cognitive Assessment *FAQ* Functional Activities Questionnaire *ADL* Katz index of activities of daily living
IADL Lawton instrumental activities of daily living scale
† see Section 3.7.2 for test thresholds
. indicates incalculable results

Figure 4.10: Natural frequencies on tests for normal as target condition



The figure plots the diagnostic accuracy of the index tests using natural frequencies. Plots above the line are true results and those below the line are false results.

4.6.4 Index tests in combination

4.6.4.1 Two test combination: GP combined tests

Tables 4.10 and 4.11 give full details of the diagnostic accuracy of GP judgement as a single test and indicate that a GP judgement of dementia has LRP 5 (95% CI 3 to 9) LRN 0.49 (95% CI 0.40 to 0.61) and GP judgement of normal has LRP 5 (95% CI 3 to 9) LRN 0.65 (95% CI 0.51 to 0.82). Table 4.12 and Table 4.13 give the accuracy of GP clinical judgement, in combination with each individual index test, for the target condition of dementia and normal respectively.

With dementia as the target condition, the combined tests with the highest LRP were GP + Eurotest which had an LRP of 16 (95% CI 5 to 49) and GP + ADL which had LRP 16 (95% CI 2 to 115). Other GP combined tests with an LRP of ten or more were GP + SPMT LRP 13 (95% CI 5 to 35) and GP + FAQ LRP 10 (95% CI 2 to 42). Of these only GP + Eurotest and GP + SPMT had an LRP 95% CI lower bound of more than 3, providing some support for using these combined tests in preference to others for ruling in dementia; however the confidence intervals overlapped between these tests and others, indicating no real evidence for superiority of any one GP combined test. In contrast the GP combined test with the lowest LRN was GP + MAT LRN 0.46 (95% CI 0.30 to 0.71) and no test had an upper 95% CI of LRN that was less than 0.60, indicating relatively poor performance of GP combined tests for ruling out dementia.

With normal as the target condition, the combined tests with the highest LRP were GP + AD8 which had an LRP of 45 (95% CI 6 to 341) and GP + IQCODE which had LRP 29 (95% CI 7 to 123). Other GP combined tests with an LRP of ten or more were GP + 6CIT LRP 13 (95% CI 5 to 30), GP + GPCOG LRP 11 (95% CI 4 to 30), and GP + Minicog LRP 12 (95% CI 5 to 29). Of these five tests, four had an LRP 95% CI lower bound of more than 3 (GP + GPCOG did not), providing some support for using these four GP combined tests in preference to others for ruling in normal; however the confidence intervals overlapped between these tests and others indicating no real evidence for superiority of any one GP combined test. In contrast the GP combined test with the lowest LRN was GP + SPMT LRN 0.62 (95% CI 0.49 to 0.79) and no test had an upper 95% CI of LRN that was less than 0.79, indicating relatively poor performance of GP combined tests for ruling out normal.

For normal as the target condition it was not possible to analyse the combination of GP judgement and MAT because of perfect prediction, MAT as a single test had 100% sensitivity for the diagnosis of normal. MAT was completed by 34 people of whom 32 were judged by the referring GP to have some form of impairment. In contrast, four of the 34 people who completed MAT were normal according to the reference standard and 30 had some impairment (of whom 1 had other impairment). All four of the people with normal cognition were judged by the referring GP to have CIND and were normal on MAT.

Figure 4.11 shows naïve decision curves for all of the index tests, on a separate axis for each target condition. The pair of curves plot the net benefit across all thresholds, along with the treat-all and treat-none lines. This plot provides the reader with an overview of the shape of the curves across threshold probabilities. The second larger pair of curves (Figure 4.12 and Figure 4.13) plot the net

benefit at the thresholds that were judged as being clinically plausible, as described in Section 3.7.4.1.

For the target condition dementia Figure 4.12 indicates that at threshold probabilities of 80% to 90% the combinations GP judgement + SPMT, and GP judgement + Eurotest have the largest net benefit. At threshold probabilities of ≥ 93% to 98% the combination test GP judgement + TAC has largest net benefit and at threshold probabilities of more than 98% the treat none approach has largest net benefit.

For the target condition normal Figure 4.13 indicates that at threshold probabilities of up to 81% the combination test GP judgement + IQCODE has largest net benefit, with GP judgement + AD8 having similar but slightly lower net benefit. At threshold probabilities of ≥ 82% to 93% the combination test GP judgement + MOCA has largest net benefit and at threshold probabilities of more than 93% the treat none approach has largest net benefit.

Table 4.12: Diagnostic accuracy of tests for target condition dementia, in combination with GP judgement

Test	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LRP (95% CI)	LRN (95% CI)	DOR (95% CI)	Probability ‡
Brief tests †								
EPSS	44 (35 to 53)	91 (84 to 95)	85 (75 to 93)	57 (49 to 64)	5 (3 to 9)	0.62 (0.53 to 0.73)	8 (4 to 18)	0.89
TAC	17 (11 to 24)	100 (97 to 100)	100 (85 to 100)	50 (43 to 56)	. (. to .)	0.83 (0.77 to 0.90)	. (6 to .)	0.98
Phototest	37 (29 to 46)	94 (88 to 98)	89 (78 to 96)	55 (48 to 63)	7 (3 to 15)	0.66 (0.58 to 0.76)	10 (4 to 30)	0.92
6CIT	46 (37 to 55)	93 (86 to 97)	88 (78 to 95)	58 (50 to 66)	6 (3 to 12)	0.59 (0.50 to 0.69)	10 (5 to 27)	0.91
GPCOG	55 (46 to 64)	91 (84 to 95)	88 (79 to 94)	62 (54 to 70)	6 (3 to 11)	0.49 (0.40 to 0.60)	12 (6 to 28)	0.88
Minicog	42 (34 to 51)	94 (87 to 97)	89 (78 to 95)	57 (49 to 64)	7 (3 to 14)	0.62 (0.53 to 0.72)	11 (4 to 29)	0.92
Medium duration tests								
Eurotest	44 (35 to 53)	97 (92 to 99)	95 (86 to 99)	59 (51 to 66)	16 (5 to 49)	0.58 (0.49 to 0.67)	27 (8 to 140)	0.94
SPMT	49 (40 to 58)	96 (91 to 99)	94 (86 to 98)	61 (53 to 68)	13 (5 to 35)	0.53 (0.44 to 0.63)	25 (9 to 99)	0.92
TUG	14 (9 to 21)	98 (93 to 100)	90 (68 to 99)	49 (42 to 56)	8 (2 to 32)	0.88 (0.81 to 0.94)	9 (2 to 78)	0.93
Sniffin sticks	51 (41 to 61)	88 (80 to 94)	83 (71 to 92)	61 (52 to 70)	4 (2 to 8)	0.55 (0.45 to 0.69)	8 (4 to 19)	0.83
Longer tests								
MAT	54 (33 to 74)	100 (69 to 100)	100 (75 to 100)	48 (26 to 70)	. (. to .)	0.46 (0.30 to 0.71)	. (3 to .)	0.95
MOCA	55 (45 to 64)	87 (77 to 93)	84 (74 to 92)	59 (50 to 68)	4 (2 to 7)	0.52 (0.42 to 0.66)	8 (4 to 18)	0.84
Informant tests								
FAQ	19 (13 to 27)	98 (93 to 100)	93 (76 to 99)	50 (43 to 57)	10 (2 to 42)	0.83 (0.76 to 0.90)	12 (3 to 110)	0.97
ADL	15 (9 to 22)	99 (95 to 100)	95 (75 to 100)	49 (42 to 56)	16 (2 to 115)	0.86 (0.80 to 0.93)	18 (3 to 761)	0.97
IADL	26 (19 to 34)	96 (91 to 99)	89 (75 to 97)	51 (44 to 59)	7 (3 to 19)	0.77 (0.69 to 0.86)	9 (3 to 36)	0.92
AD8	55 (46 to 63)	92 (85 to 96)	89 (80 to 95)	62 (54 to 70)	7 (3 to 12)	0.50 (0.41 to 0.60)	13 (6 to 32)	0.89
IQCODE	56 (47 to 64)	91 (84 to 96)	89 (80 to 95)	61 (53 to 69)	6 (3 to 12)	0.49 (0.40 to 0.60)	13 (6 to 31)	0.88

∗ *TAC* Time and change *6CIT* Six Item Cognitive Impairment Test *GPCOG* General Practitioner Assessment of Cognition
SPMT Scenery Picture Memory Test *TUG* Timed up-and-go *MAT* Memory alteration test
MOCA Montreal Cognitive Assessment *FAQ* Functional Activities Questionnaire *ADL* Katz index of activities of daily living
IADL Lawton instrumental activities of daily living scale
† see Section 3.7.2 for test thresholds
. indicates incalculable results
‡ predicted probability of dementia in test positive group, from logistic regression equation, see Methods Section 3.7.5.1

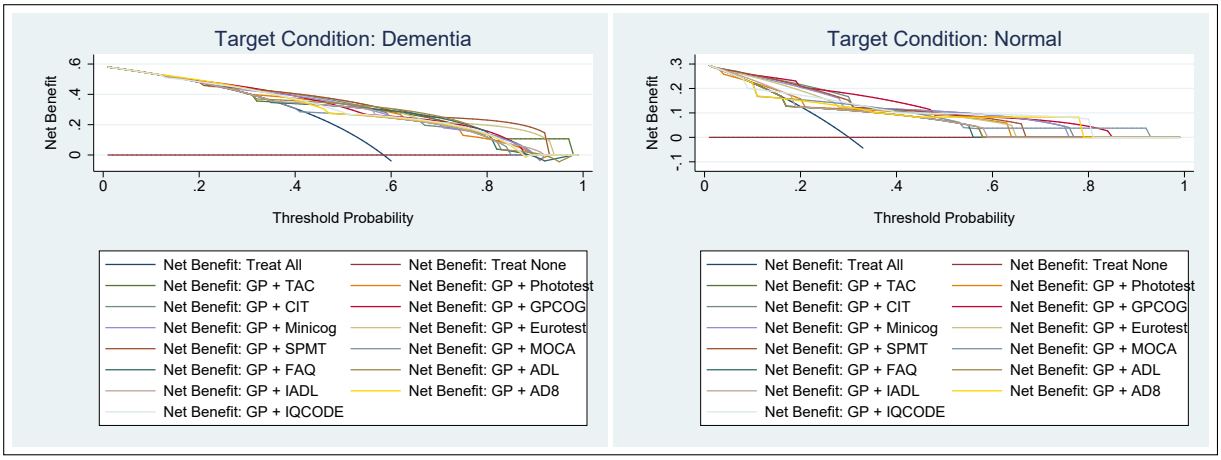
Table 4.13: Diagnostic accuracy of tests for target condition normal cognition, in combination with GP judgement

Test	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LRP (95% CI)	LRN (95% CI)	DOR (95% CI)	Probability ‡
Brief tests †								
EPSS	13 (5 to 26)	98 (96 to 100)	67 (30 to 93)	82 (77 to 87)	8 (2 to 32)	0.89 (0.79 to 0.99)	9 (2 to 59)	0.79
TAC	40 (26 to 56)	91 (85 to 95)	58 (39 to 75)	83 (77 to 89)	4 (2 to 8)	0.66 (0.52 to 0.83)	7 (3 to 16)	0.58
Phototest	39 (25 to 55)	95 (91 to 97)	64 (44 to 81)	87 (81 to 91)	8 (4 to 15)	0.64 (0.51 to 0.81)	12 (5 to 31)	0.64
6CIT	39 (25 to 55)	97 (93 to 99)	75 (53 to 90)	87 (82 to 91)	13 (5 to 30)	0.63 (0.50 to 0.79)	20 (7 to 65)	0.76
GPCOG	30 (17 to 45)	97 (94 to 99)	74 (49 to 91)	85 (80 to 89)	11 (4 to 30)	0.72 (0.60 to 0.87)	16 (5 to 59)	0.85
Minicog	38 (25 to 54)	97 (93 to 99)	75 (53 to 90)	87 (81 to 91)	12 (5 to 29)	0.64 (0.51 to 0.80)	19 (7 to 63)	0.76
Medium duration tests								
Eurotest	40 (26 to 56)	95 (91 to 97)	66 (46 to 82)	87 (81 to 91)	8 (4 to 16)	0.63 (0.50 to 0.80)	12 (5 to 33)	0.64
SPMT	40 (26 to 56)	95 (91 to 98)	68 (48 to 84)	87 (81 to 91)	9 (4 to 18)	0.62 (0.49 to 0.79)	14 (5 to 38)	0.67
TUG	40 (26 to 56)	94 (89 to 97)	61 (42 to 78)	86 (81 to 91)	6 (3 to 12)	0.64 (0.50 to 0.81)	10 (4 to 25)	0.6
Sniffin sticks	2 (0 to 13)	100 (98 to 100)	100 (3 to 100)	79 (72 to 84)	. (. to .)	0.98 (0.93 to 1.02)	. (0 to .)	0.94
Longer tests								
MAT	. (. to .)	. (. to .)	. (. to .)	. (. to .)	. (. to .)	. (. to .)	. (. to .)	.
MOCA	12 (4 to 25)	100 (98 to 100)	100 (48 to 100)	81 (75 to 86)	. (. to .)	0.88 (0.79 to 0.98)	. (5 to .)	0.93
Informant tests								
FAQ	40 (26 to 56)	90 (85 to 94)	56 (38 to 73)	83 (77 to 89)	4 (2 to 8)	0.66 (0.52 to 0.84)	6 (3 to 15)	0.56
ADL	40 (26 to 56)	91 (85 to 95)	58 (39 to 75)	83 (77 to 89)	4 (2 to 8)	0.66 (0.52 to 0.83)	7 (3 to 16)	0.58
IADL	40 (26 to 56)	93 (89 to 96)	59 (41 to 76)	87 (81 to 91)	6 (3 to 11)	0.64 (0.50 to 0.81)	9 (4 to 23)	0.59
AD8	23 (12 to 38)	99 (97 to 100)	92 (62 to 100)	84 (79 to 89)	45 (6 to 341)	0.77 (0.66 to 0.90)	59 (8 to 2537)	0.78
IQCODE	31 (18 to 47)	99 (96 to 100)	88 (62 to 98)	86 (80 to 90)	29 (7 to 123)	0.70 (0.57 to 0.85)	42 (9 to 384)	0.8

∗ *TAC* Time and change *6CIT* Six Item Cognitive Impairment Test *GPCOG* General Practitioner Assessment of Cognition
SPMT Scenery Picture Memory Test *TUG* Timed up-and-go *MAT* Memory alteration test
MOCA Montreal Cognitive Assessment *FAQ* Functional Activities Questionnaire *ADL* Katz index of activities of daily living
IADL Lawton instrumental activities of daily living scale
† see Section 3.7.2 for test thresholds
. indicates incalculable results
‡ predicted probability of dementia in test positive group, from logistic regression equation, see Methods Section 3.7.5.1

Figure 4.11: Decision curves for all tests, for both target conditions, at all threshold probabilities

The plot provides the reader with an overview of the shape of the curves across threshold probabilities.
Figure 4.12 and Figure 4.13 provide greater clarity for interpretation at the thresholds judged to be clinically relevant

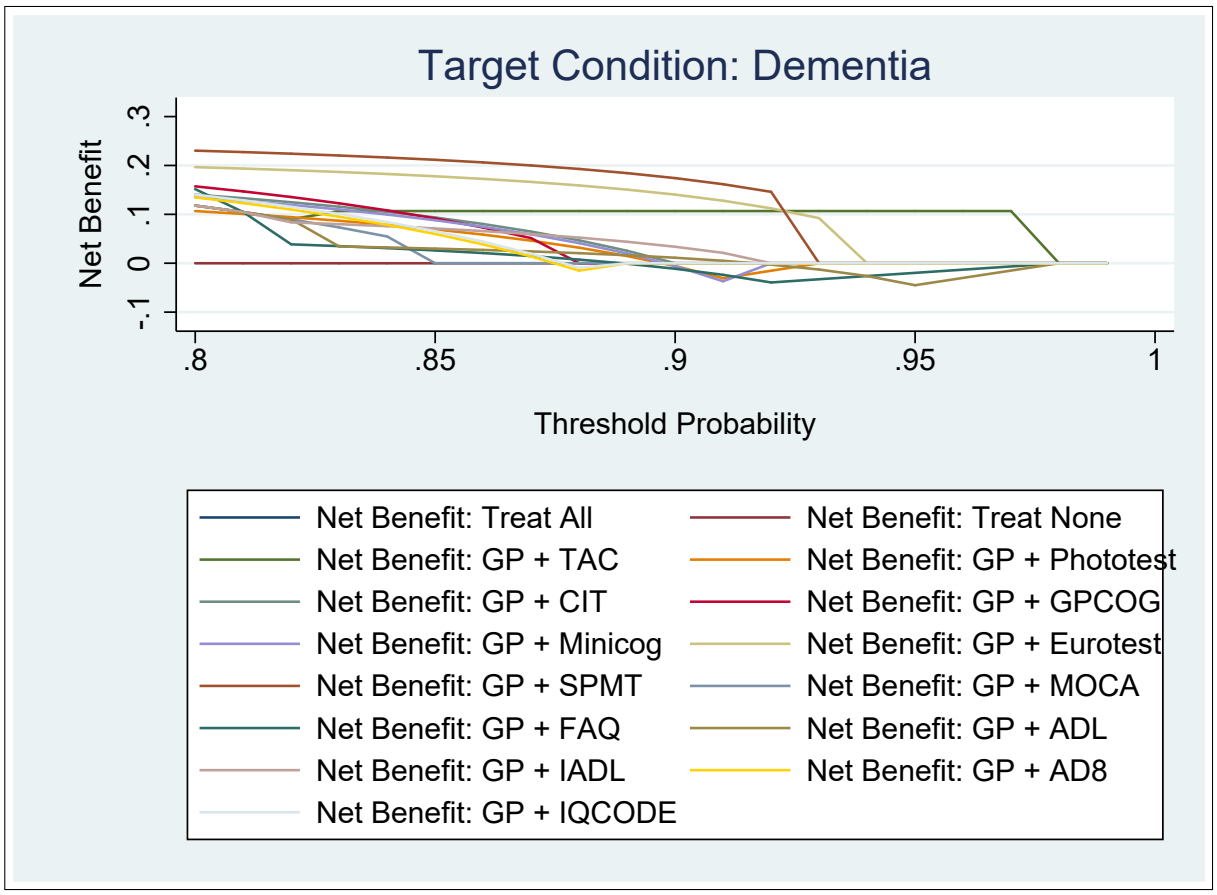


The curves plot the net benefit across a range of thresholds, along with the treat-all and treat-none lines.
Curves are smoothed, as recommended to minimise unstable results [168]

Calibration plots for each GP combined test, for each target condition, are provided in Appendix I. The GP combined tests were all well calibrated. Discrimination using AUROC is shown in Table 4.14. For dementia as the target condition the GP combined test with the largest AUROC was GP + SPMT which had AUROC 0.83 (95% CI 0.79 to 0.88). GP + Eurotest, GP + GPCOG GP + 6CIT and GP + Minicog also had AUROC of >0.8 for dementia, as GP combined tests. For normal as the target condition the same five GP combined tests had AUROC >0.8, but the combined test with the largest AUROC was GP + GPCOG which had AUROC 0.89 (95% CI 0.83 to 0.94). Overlapping confidence limit the ability to draw inference about the the superiority of one test over another regarding discrimination.

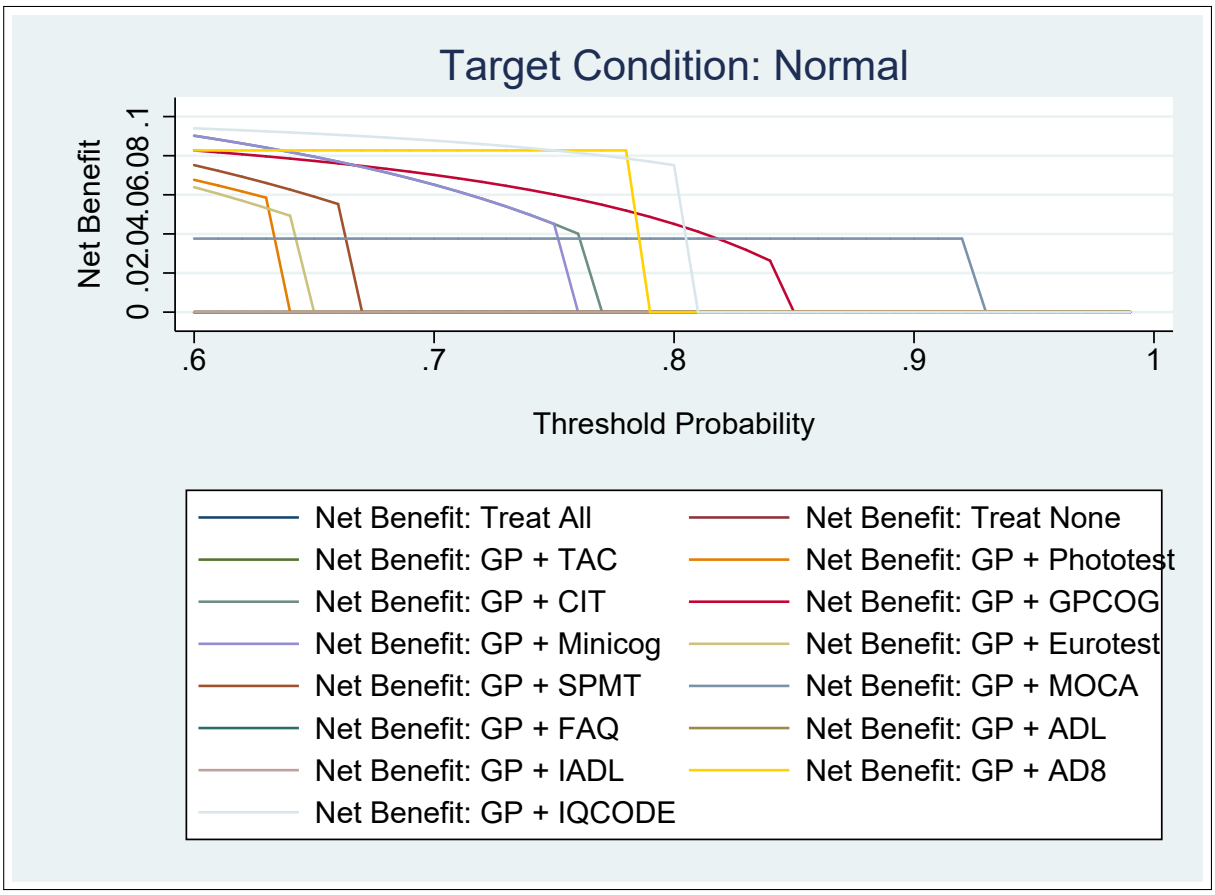
As described in Section 3.7.5.1 GP combined tests with AUROC <0.75 were not for considered combination as a GP 360 combination of three tests. For dementia as the target condition two GP combined tests had an AUROC of <0.75: Sniffin sticks AUROC 0.70 and MOCA AUROC. For normal as the target condition five tests had a AUROC of <0.75: TAC 0.66, Sniffin sticks 0.69, FAQ 0.65, ADL 0.66, and AD8 0.72. Overlapping confidence limit the ability to draw inference about the the superiority of one test over another regarding discrimination.

Figure 4.12: Decision curves for all tests, with dementia as the target condition, at threshold probabilities of more than 80%



The curves plot the net benefit across a range of thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]

Figure 4.13: Decision curves for all tests, with normal as the target condition, at threshold probabilities of more than 60%



The curves plot the net benefit across a range of thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]

Table 4.14: Discrimination, using AUROC, of the GP combined tests

	AUROC (95% CI) for Target condition	
Test	**Dementia**	**Normal**
EPSS	0.77 (0.72 to 0.82)	0.76 (0.68 to 0.83)
TAC	0.77 (0.72 to 0.82)	0.66 (0.58 to 0.73)
Phototest	0.79 (0.74 to 0.85)	0.78 (0.71 to 0.84)
6CIT	0.81 (0.76 to 0.86)	0.87 (0.83 to 0.91)
GPCOG	0.82 (0.77 to 0.87)	0.89 (0.83 to 0.94)
Minicog	0.81 (0.75 to 0.86)	0.85 (0.80 to 0.90)
Eurotest	0.82 (0.78 to 0.87)	0.83 (0.78 to 0.88)
SPMT	0.83 (0.79 to 0.88)	0.85 (0.80 to 0.90)
TUG	0.75 (0.7 to 0.81)	0.72 (0.65 to 0.79)
Sniffin sticks	0.70 (0.64 to 0.76)	0.69 (0.60 to 0.77)
MAT	0.79 (0.65 to 0.92)	. (. to .)
MOCA	0.71 (0.65 to 0.77)	0.74 (0.66 to 0.82)
FAQ	0.76 (0.71 to 0.81)	0.65 (0.58 to 0.73)
ADL	0.78 (0.72 to 0.83)	0.66 (0.58 to 0.73)
IADL	0.76 (0.70 to 0.81)	0.75 (0.69 to 0.82)
AD8	0.79 (0.74 to 0.84)	0.72 (0.64 to 0.8)
IQCODE	0.79 (0.74 to 0.84)	0.76 (0.68 to 0.84)

∗ *TAC* Time and change *6CIT* Six Item Cognitive Impairment Test
GPCOG General Practitioner Assessment of Cognition
SPMT Scenery Picture Memory Test *TUG* Timed up-and-go *MAT* Memory alteration test
MOCA Montreal Cognitive Assessment *FAQ* Functional Activities Questionnaire
ADL Katz index of activities of daily living
IADL Lawton instrumental activities of daily living scale
† see Section 4.3. GP+ MAT was not possible for normal as target condition
. indicates incalculable results

4.6.4.2 Three test combination: GP 360 tests

Accounting for discrimination and calibration, the tests with the highest LRP were selected for each target condition. For dementia as the target condition the two patient completed tests were t^pA = TAC and t^pB = Eurotest, and the two informant completed tests were t^iA = FAQ and t^iB=ADL (Section 3.7.5.1)[3]). For normal as the target condition the two patient completed tests were t^pA = 6CIT and t^pB = Minicog, and the two informant completed tests were t^iA = IQCODE and IADL. Section 3.7.5.2 describes how the tests were combined and dichotomised.

Table 4.15 shows the tabulated classification of the GP 360 tests using natural frequencies in a theoretical population. Based on the study sample, of 1000 people who have presented to a GP with symptoms of dementia; based on the study sample 546 of the 1000 have dementia. Table 4.15 shows the natural frequencies for both the continuous tests and the dichotomised tests (at a predicted probability of 80% for dementia and 60% for normal, see Section 3.7.5.2). Table 4.16 shows the diagnostic accuracy of the GP 360 tests for both target conditions. Based on the new findings GPs would be expected to identify 358 people as having dementia, and of those 308 would have dementia, 46 would have MCI and 4 would have normal cognition. In contrast, GPs would be expected to identify 141 people as being normal, and of these 79 would be normal, 33 would have MCI, 25 would have dementia and 4 would have other cognition.

Table 4.15 shows that for the target condition dementia in the 358 people classified by the GP as having dementia, the dichotomised test GP + Eurotest + FAQ would have 292 test positives of whom 275 would have dementia and 17 would have MCI. In contrast, using GP + TAC + FAQ, which from Table 4.16 has the highest LRP, would identify 183 people as having dementia, of whom 175 would have dementia and eight would have MCI. Using Table 4.16 and Table 4.15 together it can be seen that while GP + TAC + FAQ has the highest LRP 17 (95% CI 4 to 69) for dementia, using GP + Eurotest + FAQ which has LRP 14 (95% CI 5 to 36) identifies an additional 100 TP cases of dementia at a cost of 9 additional FP, all of whom have MCI.

Figure 4.14 shows that for the target condition dementia at threshold probabilities above 80% to around 93%, the dichotomised GP 360 test with GP judgement, Eurotest, and FAQ has the largest net benefit, whereas GP + TAC + FAQ has largest net benefit at thresholds between 93% and 97%.

Table 4.15 shows that for the target condition normal in the 141 people classified by the GP as being normal the dichotomised test GP + 6CIT + IQCODE would have 65 test positives of whom 61 would be normal and four would have MCI. In contrast, using continuous GP + Minicog + IQCODE, which from Table 4.16 has the highest LRP (∞), would identify 61 people as being normal, of whom all would be normal.

Figure 4.15 shows that for the target condition normal at threshold probabilities above 60% to around 93%, the dichotomised GP 360 test with GP judgement, 6CIT, and IQCODE has the largest net benefit. The continuous test GP + Minicog + IQCODE has largest net benefit of the continuous

[3]two **p**atient-completed (t^pA, t^pB) and two **i**nformant-completed (t^iA, t^iB) GP combined tests, selected on the basis of the highest LRP for the GP combined test; see Section 3.7.5.2

tests, at thresholds up to 90%, but the curve is unstable (indicated by the upward slant at thresholds of 63%).

Table 4.15: Tabulated classification of combined index tests

Test ∗	**Normal** †	**MCI**†	**Dementia**†	**Other**	**Predicted probability of dementia if test positive** ‡
in 1000 people, of whom 546 have dementia, of 358 people classified by GP as having dementia					
GP Dementia	4	46	308	0	
Target condition: dementia	*Number of people who test positive* §				
Binary tests					
GP + TAC + FAQ	0	8	175	0	1
GP + TAC + ADL	0	8	151	0	1
GP + Eurotest + FAQ	0	17	275	0	1
GP + Eurotest + ADL	0	25	293	0	1
Continuous probability tests					
GP + TAC + FAQ	0	0	33	0	0.99
GP + TAC + ADL	0	0	33	0	1
GP + Eurotest + FAQ	0	4	100	0	0.98
GP + Eurotest + ADL	0	0	75	0	0.98
in 1000 people, of whom 454 are normal, of 141 people classified by GP as normal					
GP Normal	79	33	25	4	
Target condition: normal	*Number of people who test positive* §				
Binary tests					
GP + 6CIT + IQCODE	61	4	0	0	1
GP + Minicog + IQCODE	78	9	9	4	1
GP + 6CIT + IADL	76	13	8	4	1
GP + Minicog + IADL	75	8	8	4	1
Continuous probability tests					
GP + 6CIT + IQCODE	61	4	0	0	0.87
GP + Minicog + IQCODE	61	0	0	0	0.89
GP + 6CIT + IADL	76	13	8	4	0.76
GP + Minicog + IADL	75	8	8	4	0.78

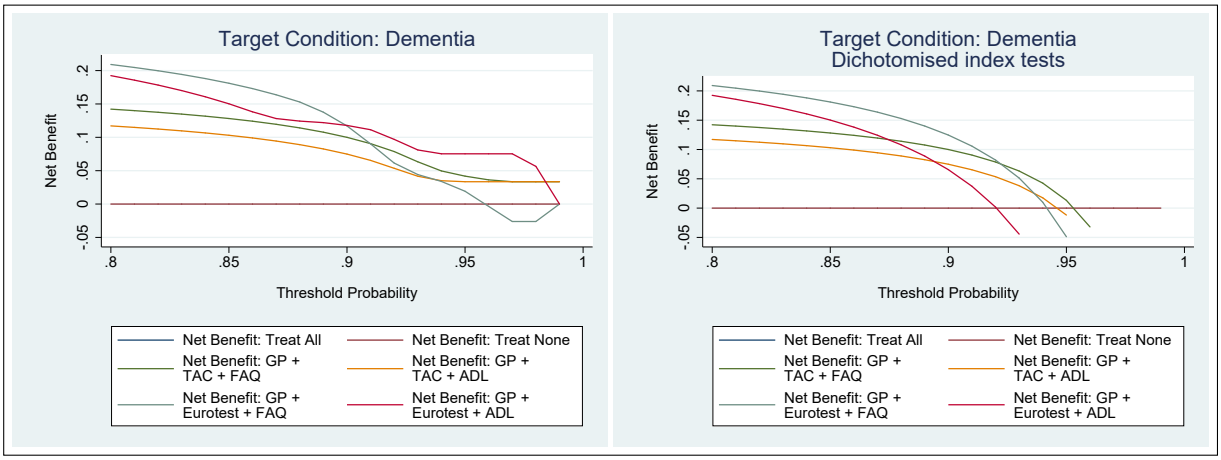
∗ *TAC* Time and change *FAQ* Functional Activities Questionnaire
ADL Katz index of activities of daily living
6CIT Six Item Cognitive Impairment Test
IQCODE Short Form of the Informant Questionnaire on Cognitive Decline in the Elderly
IADL Lawton instrumental activities of daily living scale
† Reference standard diagnosis; Dementia ICD-10 Peterson MCI; see Figure 4.1
‡ predicted probability of dementia in test positive group, from logistic regression equation, Methods Section 3.7.5.1 probability is 1 where tests are dichotomised
§indicating for example 183 (175+8) people test positive on GP + TAC + FAQ of the 358 identified by GP as having dementia. The remainder (358-183 = 175) test negative on the GP 360 test and may be normal, MCI or have dementia

Table 4.16: Diagnostic accuracy of GP 360 tests for both target conditions

Test	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LRP (95% CI)	LRN (95% CI)	DOR (95% CI)
Dementia as the target condition							
Binary tests							
GP + TAC + FAQ	32 (24 to 40)	98 (93 to 100)	95 (85 to 99)	54 (47 to 61)	17 (4 to 69)	0.69 (0.62 to 0.78)	25 (6 to 215)
GP + TAC + ADL	27 (20 to 36)	98 (93 to 100)	95 (82 to 99)	53 (46 to 60)	15 (4 to 60)	0.74 (0.66 to 0.82)	20 (5 to 175)
GP + Eurotest + FAQ	50 (41 to 59)	96 (91 to 99)	94 (86 to 98)	61 (53 to 69)	14 (5 to 36)	0.52 (0.44 to 0.62)	26 (9 to 102)
GP + Eurotest + ADL	53 (45 to 62)	94 (88 to 98)	92 (84 to 97)	63 (55 to 70)	10 (4 to 21)	0.49 (0.41 to 0.6)	20 (8 to 58)
Continuous probability tests							
GP + TAC + FAQ	6 (3 to 12)	100 (97 to 100)	100 (63 to 100)	47 (40 to 53)	. (. to .)	0.94 (0.9 to 0.98)	. (2 to .)
GP + TAC + ADL	6 (3 to 12)	100 (97 to 100)	100 (63 to 100)	47 (40 to 53)	. (. to .)	0.94 (0.9 to 0.98)	. (2 to .)
GP + Eurotest + FAQ	18 (12 to 26)	99 (95 to 100)	96 (80 to 100)	50 (43 to 57)	20 (3 to 143)	0.83 (0.76 to 0.9)	24 (4 to 987)
GP + Eurotest + ADL	14 (8 to 21)	100 (97 to 100)	100 (81 to 100)	49 (42 to 56)	. (. to .)	0.86 (0.81 to 0.92)	. (4 to .)
Normal as the target condition							
Binary tests							
GP + 6CIT + IQCODE	32 (19 to 48)	99 (97 to 100)	93 (68 to 100)	86 (81 to 90)	59 (8 to 436)	0.69 (0.56 to 0.84)	86 (12 to 3648)
GP + Minicog + IQCODE	40 (26 to 56)	97 (94 to 99)	78 (56 to 93)	87 (82 to 91)	15 (6 to 38)	0.62 (0.49 to 0.78)	24 (8 to 88)
GP + 6CIT + IADL	39 (25 to 55)	97 (93 to 99)	75 (53 to 90)	87 (82 to 91)	13 (5 to 30)	0.63 (0.5 to 0.79)	20 (7 to 65)
GP + Minicog + IADL	38 (25 to 54)	97 (94 to 99)	78 (56 to 93)	87 (81 to 91)	15 (6 to 38)	0.63 (0.51 to 0.79)	23 (7 to 85)
Continuous probability tests							
GP + 6CIT + IQCODE	32 (19 to 48)	99 (97 to 100)	93 (68 to 100)	86 (81 to 90)	59 (8 to 436)	0.69 (0.56 to 0.84)	86 (12 to 3648)
GP + Minicog + IQCODE	31 (18 to 47)	100 (98 to 100)	100 (77 to 100)	86 (80 to 90)	. (. to .)	0.69 (0.57 to 0.84)	. (21 to .)
GP + 6CIT + IADL	39 (25 to 55)	97 (93 to 99)	75 (53 to 90)	87 (82 to 91)	13 (5 to 30)	0.63 (0.5 to 0.79)	20 (7 to 65)
GP + Minicog + IADL	38 (25 to 54)	97 (94 to 99)	78 (56 to 93)	87 (81 to 91)	15 (6 to 38)	0.63 (0.51 to 0.79)	23 (7 to 85)

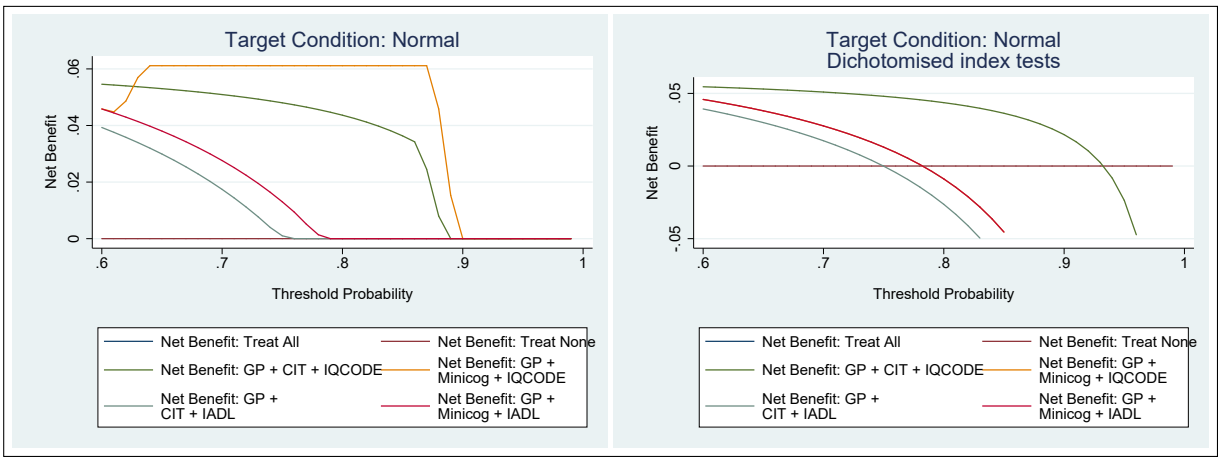
∗ *TAC* Time and change *6CIT* Six Item Cognitive Impairment Test *GPCOG* General Practitioner Assessment of Cognition
SPMT Scenery Picture Memory Test *TUG* Timed up-and-go *MAT* Memory alteration test
MOCA Montreal Cognitive Assessment *FAQ* Functional Activities Questionnaire *ADL* Katz index of activities of daily living
IADL Lawton instrumental activities of daily living scale
† see Section 3.7.2 for test thresholds
. indicates incalculable results

Figure 4.14: Decision curves for the GP 360 tests, with dementia as the target tests condition, at threshold probabilities of more than 80%



The curves plot the net benefit across a range of thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]

Figure 4.15: Decision curves for the GP 360 tests, with normal as the target condition, at threshold probabilities of more than 60%



The curves plot the net benefit across a range of thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]

For the GP 360 test with Eurotest and FAQ, the regression coefficients in the logistic regression equation were GP judgement 1.88, Eurotest 1.87, FAQ 1.65 and constant -1.33. From these it can be deduced that if all tests indicate dementia then the score from the regression coefficients would be 4.07, the exponent of which is 58.5, for a probability of dementia of 98% (derived from $\frac{odds}{1+odds}$). Conversely if GP judgement indicates dementia but the other tests indicate no dementia the probability of dementia is 63%.

For the GP 360 test with 6CIT and IQCODE, the regression coefficients in the logistic regression

equation were GP judgement 1.64, 6CIT 3.40, IQCODE 1.55 , and a constant of -4.66. From these, it can be deduced that if all tests indicate normal cognition then the score from the regression coefficients would be 1.93, the exponent of which is 6.88, for a probability of normal of 87%. Conversely if GP judgement indicates normal cognition but the other tests indicate impairment the probability of normal cognition is 5%.

4.7 Sensitivity analyses

The method of the sensitivity analyses is described in Section 3.7.6. The full results are provided in Appendix J. For technical reasons, detailed in Section 3.7.6 the results are not the same as reported in Table 4.12 and Table 4.13. A sensitivity analysis of TAC using continuous scores could not be done for normal as the target condition because of perfect prediction, similarly a sensitivity analysis of AD8 for dementia as the target condition could not be done by sex because of perfect prediction. Findings related to MAT are reported in the full results in the Appendix but are not discussed in the text because of small numbers leading to significant uncertainty in the estimates.

4.7.1 Continuous test scores

For the target condition dementia, analysis of GP combined tests as continuous scores had similar sensitivity compared to when they were analysed at the published threshold. Some tests had lower sensitivity, and this was most marked for Sniffin Sticks, which had a sensitivity of 37% (95% CI 27% to 47%) when analysed as a continuous test, compared to a sensitivity of 51% (95% CI 41% to 61%) when analysed at the published threshold. In contrast Eurotest had a higher sensitivity of 61% (95% CI 52% to 69%) when analysed as a continuous test, compared to 44% (95% CI 35% to 53%) when analysed at the published threshold. Phototest also had a higher sensitivity of 47% (95% CI 38% to 55%) when analysed as a continuous test, compared to 37% (95% CI 29% to 46%) at the published threshold. Other differences in sensitivity when tests were analysed as continuous tests were less than 10 percentage points. Specificity of tests when analysed as continuous measures was similar to when they were analysed at the published threshold, with all differences being ≤ 6 percentage points.

For the target condition normal there were no differences when GP combined tests were analysed as continuous tests or at the published threshold. This finding was checked and confirmed.

4.7.2 Age

For the target condition dementia some GP combined tests had lower sensitivity in people aged < 80 years than people aged ≥ 80 years. ADL had a sensitivity of 4% (95% CI 1% to 15%) in people aged < 80 years compared to sensitivity 69% (95% CI 58% to 79%) in people aged ≥ 80 years. Six other tests had a sensitivity in younger people that was more than 10 percentage points less than the sensitivity in older people: FAQ, MOCA, TAC, SPMT, Phototest and Eurotest. In contrast, specificity on GP combined tests was often higher in people aged < 80 years. ADL had a specificity of 99% (95%

CI 92% to 100%) in people aged < 80 years, compared to 76% (95% CI 60% to 89%) in people aged ≥ 80 years. Five other GP combined tests had a specificity in younger people that was more than 10 percentage points more than the specificity in older people: TAC, SPMT, FAQ, MOCA, and Phototest.

For normal as the target condition, some GP combined tests had higher sensitivity in people aged < 80 years than people aged ≥ 80 years. Sniffin sticks had a sensitivity of 44% (95% CI 26% to 62%) in people aged < 80 years compared to a sensitivity of 0% (0% to 37%) in people aged ≥ 80 years. Nine other GP combined tests had a sensitivity in younger people that was more than 10 percentage points more than the sensitivity in older people: TAC, Phototest, Eurotest, FAQ, ADL, IADL, MOCA, AD8, and GPCOG. In contrast, specificity on GP combined tests was often slightly lower in people aged < 80 years. Eurotest had a specificity of 86% (95% CI 73% to 95%) in people aged < 80 years compared to a specificity of 100% (95% CI 97% to 100%) in people aged ≥ 80 years. Five other GP combined tests had a specificity in younger people that was more than 10 percentage points less than the specificity in older people: IADL, ADL, FAQ, Sniffin sticks, and TAC.

Fewer people aged < 80 years than those aged ≥ 80 years had dementia: < 80 years 40% (95% CI 31% to 50%) ≥ 80 years 68% (95% CI 59% to 76%); but the proportion with MCI was similar: < 80 years 29% (95% CI 21% to 38%) ≥ 80 years 20% (95% CI 14% to 29%); normal cognition was more common in younger people: < 80 years 29% (95% CI 21% to 38%) ≥ 80 years 10% (95% CI 6% to 17%). Mean ACE3 scores were 79 (σ 14) in people aged < 80 years and 71 (σ 14) in people ≥ 80 years, and mean school leaving age was the same: 16 (σ 1) in people < 80 years and ≥ 80 years. There was weak evidence of an association between age and school leaving age with a coefficient for being ≥ 80 years of -0.20 (95% CI -0.55 to 0.15). In contrast there was some evidence of an association between total ACE3 score and age with a coefficient for being ≥ 80 years of -7.4 (95% CI -10.9 to -3.8).

4.7.3 Sex

In general, for the target condition dementia GP combined tests were more sensitive in women than men. For example, ADL had a sensitivity of 73% (60% to 80%) in women compared to 9% (95% CI 3% to 18%) in men. Similarly, IQCODE had a sensitivity of 62% (95% CI 49% to 74%) in women, compared to a sensitivity of 0% (95% CI 0% to 5%) in men. The GP combined tests with the smallest difference in sensitivity between women and men were GPCOG which had a sensitivity of 63% (95% CI 50% to 74%) in women compared to 49% (95% CI 36% to 61%) in men, and Eurotest which had a sensitivity of 52% (95% CI 39% to 64%) in women compared to 37% (95% CI 25% to 49%) in men. In contrast, specificity of GP combined tests was similar between men and women and differed by less than 10 percentage points.

For the target condition normal, some GP combined tests had lower sensitivity in women compared to men. For example, ADL had a sensitivity of 0% (95% CI 0% to 14%) in women compared to 43% (95% CI 23% to 66%) in men. Six other GP combined tests had lower sensitivity in women compared to men: TAC, SPMT, TUG, FAQ, IADL, Phototest, and 6CIT. Two GP combined tests were more sensitive in women than men, MOCA had a sensitivity of 30% (95% CI 13% to 53%) in women

compared to 10% (95% CI 1% to 32%) in men, and GPCOG had a sensitivity of 83% (95% CI 63% to 95%) in women compared to 30% (95% CI 13% to 53%) in men. In contrast, specificity of GP combined tests was similar between men and women and differed by less than 10 percentage points. The proportion of men and women with each category of cognition was similar: dementia men 53% (95% CI 44% to 62%) women 56% (95% CI 47% to 65%); MCI men 27% (95% CI 19% to 36%) women 22% (95% CI 15% to 31%); normal cognition men 18% (95% CI 12% to 26%) women 21% (95% CI 14% to 30%). The average age in women (81 years σ 6 years) was similar to that in men (80 years σ 6 years). Mean ACE3 scores were 73 (σ 15) in women and 77 (σ 14) in men, and mean school leaving age was 15 (σ 1) in women and 16 (σ 1) in men. There was weak evidence of an association between sex and school leaving age or ACE3 score, the coefficients for being male were respectively 0.39 (95% CI 0.03 to 0.73) and 3.2 (95% CI -0.46 to 6.9).

4.7.4 All combinations of GP combined tests

As described in Section 3.7.5.2 and reported in Section 4.6.4.2 for the target condition dementia, four tests were selected for analysis as GP 360 tests on the basis of discrimination, calibration and decision curve analysis:

1. **GP + TAC + FAQ** sensitivity 32% (95% CI 23% to 40%) specificity 98% (95% CI 93% to 100%)
2. **GP + TAC + ADL** sensitivity 27% (95% CI 20% to 36%) specificity 98% (95% CI 93% to 100%)
3. **GP + Eurotest + FAQ** sensitivity 50% (95% CI 41% to 59%) specificity 96% (95% CI 91% to 99%)
4. **GP + Eurotest + ADL** sensitivity 53% (95% CI 45% to 62%) specificity 94% (95% CI 88% to 98%)

Appendix J shows the full results for the analysis of all GP 360 tests. For the target condition dementia, in comparison to the four tests which were selected for analysis, two tests had higher sensitivity at the same specificity: GP + SPMT + FAQ had sensitivity 54% (95% CI 45% to 62%) specificity 96% (95% CI 91% to 99%) and GP + SPMT + ADL had sensitivity 59% (95% CI 50% to 67%) specificity 96% (95% CI 91% to 99%).

For the target condition normal, four tests were selected for analysis as GP 360 tests on the basis of discrimination, calibration and decision curve analysis:

1. **GP + 6CIT + IQCODE** sensitivity 32% (95% CI 19% to 48%) specificity 99% (95% CI 97% to 100%)
2. **GP + Minicog + IQCODE** sensitivity 40% (95% CI 26% to 56%) specificity 97% (95% CI 94% to 99%)
3. **GP + 6CIT + IADL** sensitivity 39% (95% CI 25% to 55%) specificity 97% (95% CI 93% to 99%)
4. **GP + Minicog + IADL** sensitivity 38% (95% CI 25% to 54%) specificity 97% (95% CI 94% to 99%)

For the target condition normal, in comparison to the four tests which were selected for analysis, two tests had higher sensitivity at the same specificity: GP + MOCA + AD8 had sensitivity 40% (95% CI 25% to 56%) specificity 99% (95% CI 96% to 100%) and GP + MOCA + IQCODE had sensitivity 44% (95% CI 28% to 60%) specificity 99% (95% CI 95% to 100%).

4.7.5 Geography

Of participants from Bristol practices, 57% (95% CI 45% to 69%) had dementia according to the reference standard, compared to 53% (95% CI 46% to 61%) of patients from South Gloucestershire and North Somerset practices.

Diagnostic accuracy of GP judgement for dementia in South Gloucestershire and North Somerset practices was LRP 4 (95% CI 2 to 7) LRN 0.49 (95% CI 0.38 to 0.63), compared to diagnostic accuracy of LRP 15 (95% CI 2 to 103) LRN 0.49 (95% CI 0.35 to 0.69) in Bristol practices. In contrast the diagnostic accuracy for normal in South Gloucestershire and North Somerset practices was LRP 6 (95% CI 3 to 12) LRN 0.60 (95% CI 0.45 to 0.80), compared to diagnostic accuracy of LRP 3 (95% CI 0.6 to 12) LRN 0.85 (95% CI 0.59 to 1.22) in Bristol practices.

Chapter Summary

This Chapter has presented the results for the empirical quantitative investigation of diagnostic test accuracy. From 21 participating GP surgeries with a total population of 34,956 aged over 70 years, 465 people were referred to a diagnostic test accuracy study, 241 were seen (comprising 126 men and 114 women) and 240 were evaluated. Of the 240 participants, 132 had dementia, and 47 had normal cognition. The informant completed measures were similar whether they were completed by the interviewer or informant.

For dementia as the target condition, GP judgement as a single test had a sensitivity of 56% and a specificity of 89%; five tests had an LRP of more than 3 but none had a lower bound of the 95% CI of more than 3 (Eurotest LRP 4 (95% CI 2 to 5); GP judgement LRP 5 (95% CI 3 to 9); ADL LRP 6 (95% CI 2 to 14); TAC LRP 7 (95% CI 3 to 20); FAQ LRP 10 (95% CI 3 to 30)). In contrast three tests had an LRN of less than 0.20 but none had a upper bound of the 95% CI which was less than 0.20 (GPCOG LRN 0.13 (95% CI 0.07 to 0.25); AD8 LRN 0.12 (95% CI 0.05 to 0.29); IQCODE LRN 0.12 (95% CI 0.05 to 0.27)). Confidence intervals overlapped and prevented inference about the superiority of any one test.

For the target condition dementia, at threshold probabilities of 80% to 93%, the tests with the largest net benefit in combination with GP judgement were SPMT and Eurotest. For the target condition normal at threshold probabilities of 60% to 81%, the tests with the largest net benefit in combination with GP judgement were IQCODE and AD8.

For the target condition dementia at threshold probabilities of 80% to 93% GP + Eurotest + FAQ largest net benefit, whereas GP + TAC + FAQ had largest net benefit at threshold probabilities of 93% to 97%, and at threshold probabilities above 97% the treat-none approach had largest net benefit. For the target condition normal at threshold probabilities of 60% to 93% GP + 6CIT + IQCODE had largest net benefit, whereas at threshold probabilities above 93% the treat-none approach had largest net benefit.

In sensitivity analyses for the target condition dementia GP combined tests generally had similar diagnostic accuracy to when they were analysed at published thresholds, but Phototest and Eurotest had higher sensitivity when analysed as a continuous test. There was some evidence that for the target condition dementia GP combined tests were more sensitive in women, compared to men, and people aged 80 years or over, than those aged between 70 and 80 years. When all GP 360 tests were analysed for both target conditions, two tests had higher sensitivity at the same specificity than the tests which were selected in advance for analysis on the basis of discrimination calibration and decision curve analysis of the GP combined tests; for the target condition dementia these two GP 360 tests were SPMT and either FAQ or ADL; for the target condition normal the two GP 360 tests were MOCA and either AD8 or IQCODE.

CHAPTER 5

ACCEPTABILITY OF DIAGNOSIS OF DEMENTIA BY A GENERAL PRACTITIONER

This Chapter describes the qualitative aspects of the empirical research. The rationale and research questions are outlined. The methods, including the participants, sampling, data collection and management, and data analysis are described. A series of qualitative interviews were done with people who had symptoms of dementia, who were asked if they wished to be accompanied by a person of their choice, typically a family member. The results of the qualitative research are also presented, together with an interpretation of the findings through the lens of a theoretical model. In this Chapter, the term *kin* is used to refer to the person who was chosen to accompany the person with symptoms of dementia because this person was not necessarily a *carer*, indeed this term was intensely disliked by some participants, or a family member, or even their "nearest and dearest". The word kin was judged to best capture the nature of the accompanying parties in a succinct way, defined as:

> *"The group of persons who are related to one; one's kindred, kinsfolk, or relatives, collectively."* *[359]*

5.1 Rationale and research questions

Just because a test is accurate does not necessarily mean that it is appropriate or acceptable to be used in clinical practice; it is important to understand how the evidence about the accuracy of diagnostic tests can be applied in practice. Therefore, this qualitative component of the empirical research investigated how acceptable a general practice based diagnosis of dementia would be to patients and their kin.

5.1.1 Qualitative research questions

The aim of this study was to understand the acceptability, to people with cognitive symptoms and their kind, of a GP diagnosing dementia without specialist input. The objective was to obtain a range of views on how acceptable people with cognitive problems and their kind perceived the suggestion of GPs making a diagnosis of dementia, rather than specialists such as geriatricians or old age psychiatrists. The qualitative research questions was:

- How acceptable is it to patients and their kin for general practitioners to make a diagnosis of dementia independent of specialist input?

It is proposed[1] that components of a diagnostic pathway that are "acceptable" may be categorised under three headings. Firstly some details require consideration regardless of whether the diagnostician is a GP or specialist, such as the number, frequency and geographic setting of contacts in the evaluation process. Secondly other aspects, such as preferences for particular interpersonal or consulting characteristics, are specific to individual clinicians and patients. Thirdly, some features are hallmarks of an acceptable diagnostic process that define the circumstances under which one process resembles another sufficiently well for it to be acceptable, and these were of particular interest to explore.

5.2 Participants and Sampling

Participants were people with cognitive problems (patients) and their kin who were taking part in the quantitative diagnostic test accuracy study.

Participants in the main quantitative study who the GP researcher judged as having sufficient cognitive and physical reserves were purposively sampled for maximum variation in characteristics of referring practice, age and sex, and invited to partake in the additional optional qualitative interview. To ensure maximum variation in these sample characteristics, potential participants were invited at each of the three research sites (see Section 3.3), with particular efforts to invite older people who had sufficient reserves to take part. In addition to the three sampling characteristics in which maximum variation was sought, particular efforts were made to recruit people who might offer a novel or original opinion in the interviews, specifically those with mild sensory impairment, or of non-Caucasian ethnic origin. Additionally, partly because of changes to GP surgeries during the practices, including mergers of practices, and partly because of difficulties in recruiting people from certain practices to the qualitative study, as recruitment progressed sampling for maximum variation in practice evolved into maximum variation in the *size* of practice (smaller than, or bigger than, average list size - approximately 11,900; see Table 3.2 for details).

The GP researcher selected and invited people to participate in the qualitative study when they had completed the research clinic. To avoid bias in the quantitative study, the GP researcher knew

[1] by the author

little about the participants, other than their sex, age and referring GP surgery. Specifically, the GP researcher did not know the opinion that the specialist had reached about the degree of cognitive impairment, but had completed a number of index tests with the participants and therefore had some insight into the degree of cognitive difficulties that the person experienced. People who the GP researcher judged as being in an advanced stage of frailty, either physically or cognitively, were not invited to avoid the research process being overly burdensome. Limitations of human resource regarding researchers meant that only people who attended a research clinic in the final six months (9 November 2016 to 9 May 2017) of the quantitative study (which commenced on 19 April 2015) were invited to take part in a qualitative interview.

Having identified a person as potentially suitable, the GP researcher invited people to the qualitative study by explaining that it was an additional, optional study that would take place on a separate day, at a time and place convenient to the participants. An information leaflet about the qualitative study was discussed and willing participants signed a consent form indicating their agreement to be contacted to arrange an interview at a later date.

5.2.1 Data collection and management

Semi-structured interviews with participants were done in their homes, at a convenient time. The participant with cognitive problems was encouraged to have a second person present at the interview, either the same person who had attended the research clinic with them, or another person. The attendance of a second person at the interview was encouraged but was not a mandatory inclusion criteria. Interviews were usually arranged a week or so in advance of the meeting, and participants received a confirmation phone call on the day. The intention was to recruit between 15-25 participants. However, recruitment continued until there were no new emerging themes in the analysis.

At the start of the interviews the consent form for interviews was discussed together with the patient information leaflet. The patient information leaflet for the qualitative interviews had been provided previously at the main research clinic but further copies were available at the interview if required. An opportunity for questions was provided and willing participants indicated their informed consent to participate in the interview on a signed consent form. Written consent to participate was taken from both the patient and (if present) their kin; each party consented for themselves only and not as a proxy for one another. As outlined in Section 3.4 all participants in the main quantitative study were assessed as having capacity to consent to their own involvement, and this was also the case in the qualitative interviews.

The interviews explored the range of views for both the patient and their kin. The interview guide was drafted before any interviews were done, but was then refined in response to the process of interviewing people. The original interview guide elicited experiences of receiving care at the participants' GP surgeries and memory difficulties; their ideas about who should take the lead for diagnosing memory problems; their thoughts about a GP diagnosing dementia without getting a specialist involved; their thoughts on a hypothetical process where a specially trained GP in a local

area triaged referrals and held intermediate care clinics; their thoughts on the hypothetical role of a specialist nurse compared to a specialist GP; their priorities when trying to establish the cause of memory problems; their suggestions for improvements to services in future. Participants were then asked about how their particular memory problems had first been identified and to talk through their personal experience of being diagnosed.

The original intention was that participants would be invited to complete five of the index tests that had been done at the main research clinic, so that they could provide feedback on the burden that they experienced by completing the test. These tests were the Montreal cognitive assessment [333], timed up and go [323], Eurotest [319], GPCOG [301], and IQCODE [304]. These tests were prioritised for investigation because they cover a range of domains and, with the exception of Eurotest, are commonly used in clinical practice. A description of these tests is provided in Section 3.4.1. Patient participants were to be invited to complete these tests, which were also included in the main quantitative study.

The interview guide was iterated and refined in response to experience of doing the interviews, in discussion with research supervisors. The interviewer was vigilant and responsive to the burden that the participant appeared to be experiencing and modified the depth of exploratory questions as appropriate to the circumstances. The section regarding the personal experience of being diagnosed was dropped from the interview guide, as it emerged that this section contributed little information that was relevant to the research question. It became apparent within the first few interviews that the original intention to ask the person to complete five index tests from the research clinic in order to assess the burden of the test would be overly burdensome for participants, due to the time taken to repeat the tests, and also because participants had already the completed many cognitive tests with the research team and NHS service; therefore this section was dropped completely after nine interviews with a diverse group of people. The ethical approval for the qualitative interviews had been obtained as an amendment (Amendments 4 May 2015 and 10 May 2016) to the main study (IRAS id 143065, Favourable Opinion 26 Nov 2014, London - Bromley Research Ethics Committee, REC reference 14/LO/2025) and specified that the interview would last up to 30 minutes. It became clear that repeating the index tests at the qualitative interview would itself take more than the 30 minutes and this would reduce the opportunity to gain valuable insights into the the first qualitative research question, which was prioritised for investigation. It was judged as being overly burdensome to extend the interview so re-applying for a further amendment to allow a longer interview was decided to be inappropriate.

In contrast, some additional areas were added to the interview guide in response to interviews. A question was added to explore whether the need to typically drive to an appointment mitigated the desire for the appointment to be especially close to home. Some people were interested in the idea of trying to improve, or preserve, their cognition and so a question was added about the perceived value of being given activities or advice on interventions that might improve cognition. Finally, the discordant view that it was not a problem to wait for an appointment because cognitive problems

typically progress slowly prompted the addition of a question to explore the extent to which this view was shared. The original and final interview guides are provided in Appendix F. Notes were made after each interview and are provided in Appendix G.

5.3 Characteristics of qualitative subsample

During the period of recruitment to the qualitative study research clinics were held at one of three locations: Clevedon (L81040), Hanham (L81079) or Shirehampton (L81008). Table 5.1 shows the number of people from each practice who were seen in the quantitative study during the period of recruitment to the qualitative study. Of the 15 practices who referred people to the quantitative study during recruitment to the qualitative study, participants from 11 different practices had a qualitative interview, this included two practices with an average list size (9 participants), six practices with a smaller than average list size (9 participants) and three practices with a bigger than average list size (8 participants).

Table 5.2 presents an overview of the participants in the qualitative interviews, giving details of the diagnosis according to the reference standard, ACE3 score, sex, age, participants in the interview, whether specific tests were done and the perception of them explored, together with the study identifier for use with quotations. Of a total 36 people who were invited to take part, 26 participated in interviews that took place between three and seven months after the quantitative research clinic appointment (median 5 months). Participants came from 11 different GP surgeries (including urban, semi-rural and rural locations), were aged from 70 to 89 years, included 10 women and 16 men. Participants had a range of cognitive diagnoses (6 normal, 11 mild cognitive impairment, 9 dementia) and performance on Addenbrooke's Cognitive Examination III (ACE3), a standardised cognitive test, with participant scores ranging from 62 to 98 out of 100; higher scores indicate better cognitive function, scores below 82 are often consistent with dementia. Participants were interviewed alone (6), with a spouse (18), a friend (1), or a daughter (1). Of the 26 participants, 25 agreed to video recording for least part of the interview and one interview was audio-recorded only.

Three interviews had factors which are potentially relevant when considering their quotations: E30 and her husband were both nurses while they were working; S13 was a GP when working; and R13 stated she had been diagnosed with a brain tumour.

5.4 Data analysis

5.4.1 Analytical approach

The data were managed and analysed using a framework approach [360, 361] which is an appropriate methodology to use when research aims to generate recommendations within a limited time frame about a specific policy issue [360]. Framework analysis is broadly speaking a thematic analysis or qualitative content analysis, and has no strict requirement for either an inductive or deductive approach,

Table 5.1: Referrals to qualitative study from each practice

Location∗ (code †)	**Number seen in ‡ quantitative clinic during qualitative recruitment**	**Number recruited to qualitative study §**	**Practice list size ¶**
Backwell (L81060)	6	4	A
Bedminster (L81053)	0	-	-
Clevedon (L81102)	7	2	S
Clevedon (L81040)	8	5	B
Clifton (L81081)	2	0	A
Close Farm (L81050)	4	3	S
Fishponds (L81087)	0	-	-
Frampton Cotterell (L81014)	4	0	B
Hanham (L81079)	3	0	B
Henleaze (L81131)	3	1	S
Horfield (L81022)	1	0	B
Kingswood (L81063)	0	-	-
Long Ashton (L81056)	1	0	S
Mendip Vale (L81086)	3	1	B
Portishead (L81004)	2	2	B
Redfield (L81061)	0	-	-
Shirehampton (L81008)	0	-	-
Southmead (L81098)	2	1	S
Stoke Bishop (L81622)	0	-	-
Westbury on Trym (L81017)	8	2	S
Yate (L81047)	6	5	A
Total	60	26	

∗ Location is given rather than practice name because practice names change with mergers and change of doctors
† Practice code is a unique identifier held by NHS England
‡ includes total number seen, not necessarily eligible
§ - indicates zero referrals; recruitment impossible
¶ *A* Average [list size 10,000-14,000] *B* Bigger than average [list size >14,000] *S* Smaller than average [list size <10,000]

or rigid adherence to any particular epistemological, philosophical, or theoretical underpinning [362]. Additional features of framework analysis that made it a suitable approach are that it is heavily based in the accounts of people which it relates to, is systematic and repeatable, responsive to new data, comprehensive, and allows for comparisons between data [360]. While framework analysis cannot be used with heterogeneous data that cover a plethora of issues [362], this was not the case in the empirical semi-structured interviews. Phenomenological approaches were considered to be less appropriate than a comparative approach, and ethnographic methods, narrative analysis and discourse analysis were rejected because of time constraints [363].

Table 5.2: Characteristics of qualitative participants

Study identifier ∗	**Diagnosis** †	**ACE3** ‡	**Sex**	**Age**	**Participants** §	**Specific tests**
C20	MCI	93	male	72	h+w	done
C25	MCI	91	female	88	solo	not done
E19	dementia	76	male	89	h+w	planned, not done
E23	dementia	77	female	76	friend	done
E25	dementia	72	male	73	h+w	not done
E27	MCI	83	male	80	solo	not done
E30	dementia	76	female	82	h+w	not done
G18	normal	92	male	88	h+w	not done
L3	MCI	90	male	76	h+w	done
L7	dementia	73	male	80	h+w	not done
M18	dementia	62	male	77	h+w	done
M22	MCI	84	female	89	solo	not done
M23	MCI	74	male	73	solo	not done
M24	normal	92	male	71	h+w	not done
M25	dementia	69	female	87	solo	not done
N2	dementia	81	male	86	h+w	not done
P5	MCI	79	male	74	h+w	not done
R13	normal	95	female	72	h+w	done
R14	normal	97	female	73	solo	done
R15	MCI	89	male	84	h+w	done
R22	MCI	87	female	77	h+w	not done
S13	normal	90	male	72	h+w	not done
S14	MCI	87	female	85	daughter	not done
S15	dementia	81	male	85	h+w	not done
Z9	normal	98	female	73	h+w	done
Z11	MCI	79	male	79	h+w	not done

Specific tests indicates whether views on five specific index tests were elicited (see Section 5.2.1)
∗ Study identifier is an anonymous identifier but is used consistently with quotations in this thesis
† Diagnosis as assigned by the reference standard evaluation
‡ *ACE3* Addenbrooke's Cognitive Examination III
§ *h+w* husband and wife

5.4.2 Procedure for analysis

A combined inductive / deductive approach was taken, as is common with a Framework analysis [362]. The following seven-step approach was taken to data analysis, as is recommended [362]: Transcription, Familiarisation, Coding, Developing a framework, Applying the Framework, Charting the data, Interpreting the data.

5.4.2.1 Transcription

Transcription of the audio files was done verbatim by a professional approved service. Transcriptions were of the whole interview and so included attributed contributions from both the patient and (where relevant) their kin. Data were transferred to the professional transcribers using best practice in approved data management. Transcriptions were imported into NVivo 11 qualitative data analysis software [364] and verified against the original data files by listening to the original recordings; any corrections to the professional transcription were clearly marked on the transcript. Contributions to the interview from both the patient and their kin were transcribed and clearly attributed.

Data management Interviews were electronically recorded using a secure encrypted laptop. All interviews were audio-recorded and when participants consented the interviews were also simultaneously electronically video recording using the same equipment. Data were moved from the secure encrypted laptop that had been used to record the interview onto a secure encrypted University network drive for storage of research data (and deleted from the laptop) immediately on return from the interview to the university. Data were stored in compliance with relevant regulation and the University of Bristol data management policy. All audio recordings were anonymised. All personally identifiable data was stored separately to study data on secure University of Bristol servers.

5.4.2.2 Familiarisation with the interview

The full interviews were listened to and checked against the professional transcript. All interviews had been conducted by the same GP researcher and so there was already a good familiarity with the content. Ethnographic notes were also reviewed at the time of reviewing the interviews. If a video was available this was also viewed whilst reading the transcript to help provide immersion in the context of the transcribed material.

5.4.2.3 Coding

Transcripts were read line-by-line, and codes were applied to relevant parts of the transcript by the GP researcher using NVivo. Codes were derived from the words that were used in the transcript and so emerged from the data. Values and emotions about the acceptability were especially identified for coding. Particular effort was made to identify and code data that expressed an opinion or stance that was discordant with the majority. Impressionistic data on the reliability of the testimony were coded because some participants had a greater degree of cognitive impairment than others. Only data that were relevant to the issue of the research question on the acceptability of a GP diagnosis of dementia were coded. A 20% selection of transcripts were double coded by a second coder, a supervisor with significant experience in qualitative research, for quality assurance.

5.4.2.4 Developing a working analytical framework

A set of codes was discussed and agreed between coders after five transcripts had been coded. The codes emerged from the data and had not been specified rigidly in advance, although the GP researcher acknowledged some general preconceived ideas about the topic. The initial codes were:

- GP diagnosis
- Trained non-specialist diagnosis
- Specialist diagnosis
- Accessibility
- Credibility and competency
- Waiting / nurturing / uncertainty / prediagnosis
- Continuity
- Self care
- Organisational competence

These codes were developed and iterated into the the working analytical framework outlined in Table 5.3. Two sets of codes were used. Chart 1 (aspects of acceptability) contained coded data relating to GP diagnosis and role, diagnosis by a dedicated, trained but non-specialist clinician, specialist diagnosis, self care, and the extent to which communication was effective. Chart 2 (organisational competence) contained coded data relating to accessibility, nurturing and handling uncertainty, and continuity.

Table 5.3: Working analytical framework

Code	Operationalised as data relating to	Example data
Chart 1: Aspects of acceptability		
GP diagnosis and role	GP role and capacity for making diagnosis	*"it starts with your GP and then he can refer you on "[P5 informant]*
Diagnosis by dedicated, trained non specialist clinician	Dedicated non-consultant specialist making a diagnosis	*"Well I suppose as long as somebody diagnoses the problem and gets it right it doesn't really matter who does it "[R22 patient]*
Specialist diagnosis	Diagnosis by a specialist	*"Well, what's a specialist? They've learnt that by um, over the years by um, observing and being involved with them anyway. I mean some people are better at it than others that's life anyway "[M22 patient]*
Self-care	Self management and empowerment	*"I can live with it now. I've learned - it seems that you learn little tricks that you can play with your brain "[M23 patient]*
Effective communication	Extent to which communication transmitted by participant reflected their intended meaning	*"at the moment I'm not upset I'm a bit tangled "[M25 patient]*
Chart 2: Organisational competence		
Accessibility	Waiting times, geographical distance or speed of diagnosis	*"as long as one of us can drive to get you know to that specialist er treatment it's not important however I do feel sorry for the people that have to travel hundreds of miles, that's a different thing altogether isn't it? "[L7 informant]*
Nurturing and handling uncertainty	Support during the process of diagnostic evaluation	" [my friend who is dealing with cancer] *can call somebody anyway, and the nurturing and the care, and the comfort she is getting is amazing...you keep feeling dropped. You know you've left the radar and then you go and see somebody "[E25 informant]*
Continuity	Organisational coherence e.g. seeing same clinician, records being available	*"Well if they're only seeing you the once, they probably wouldn't be able to diagnose that, but if they saw you a few times, I think they could "[E30 patient]*

5.4.2.5 Applying the analytical framework

The codes outlined in Table 5.3 were used to code the data in all of the transcripts, including the five transcripts which had already been coded.

5.4.2.6 Charting data into the framework matrix

Coded data from each transcript were charted into the framework matrix using the codes outlined in Table 5.3. Data were charted separately (side-by-side) depending on whether data were attributable to a patients or their kin. Short verbatim quotes were included in the chart and were identified by coloured text.

Themes were detected from the printed spreadsheet charts and written onto sticky notes which were then categorised in an iterative process onto sheets of paper. The themes were mapped into a comprehensive but succinct classification, which was then reconsidered and refined against the detected sticky notes and charts to best fit the data. The data were mapped into a classification which was judged to reflect the elements that contribute to the acceptability of a diagnostic process, and which is outlined fully in Section 5.5:

1. Emotional context
2. System resources
3. Own resources
4. Responsive assessment
5. Important condition
6. Diagnostic process
7. Opportunity costs

5.4.2.7 Interpreting the data

The literature was reviewed to identify a theoretical framework to help interpret the findings of the qualitative framework analysis. The Health Belief Model (HBM) [365], Theory of Planned Behavior (TPB) [366], Normalization Process Theory (NPT) [367], and The Theoretical framework of acceptability (TFA) [368] were identified as being potentially relevant. The HBM relates to the uptake of health services and was originally formulated to describe the uptake of screening programs and health promoting behaviours. The TPB posits a function to understand deliberate human behaviour and asserts that this is a function of perceived behavioural control and the intention to perform the behaviour (which is held to be related to attitudes towards the behaviour and subjective). NPT is advocated as being applicable to help understand the process of implementing, embedding and integrating an innovative practice. The TFA is a multi-construct theoretical framework derived from

a systematic review which relates to "the extent to which the person delivering or receiving the healthcare intervention consider it to be appropriate" [368]. Overall the TFA was chosen as being the most appropriate model as this relates specifically to the acceptability of a healthcare intervention.

Based on a systematic review of 43 studies the authors of the TFA propose that acceptability has seven facets: affective attitude, burden, ethicality, intervention coherence, opportunity costs, perceived effectiveness, and self-efficacy. *Affective attitude* is [368] "how an individual feels about the intervention". *Burden* is [368] "the perceived amount of effort that is required to participate in the intervention". *Ethicality* is [368] "the extent to which the intervention has a good fit with the individuals value system". *Intervention coherence* is [368] "the extent to which the participant understands the intervention and how it works". *Opportunity costs* are [368] "the extent to which benefits, profits or values must be given up to engage in the intervention". *Perceived effectiveness* is [368] "the extent to which the intervention is perceived as likely to achieve its purpose". *Self-efficacy* is [368] "the participants confidence that they can perform the behaviour(s) required to participate in the intervention". Section 5.6 describes how the TFA was related to the empirically derived classification for the acceptability of diagnosis of dementia by a GP.

Charted data in the category of effective communication were not re-categorised, or mapped into a classification, but were retained to facilitate understanding of the context of the statements that participants made. Section 5.7 reflects on the challenges that were experienced when interviewing people with cognitive problems, and the impact of this.

5.5 Results: Acceptability of GP diagnosis

A classification comprising seven aspects was judged to comprehensively reflect the data. As a reminder, the research question was: *How acceptable is it to patients and family members for general practitioners to make a diagnosis of dementia independent of specialist input?* The findings for each of the seven aspects are presented below followed by a statement suggesting the implications of the findings on policy and/ or practice. The seven aspects were:

1. Emotional context
2. System resources
3. Own resources
4. Responsive assessment
5. Important condition
6. Diagnostic process
7. Opportunity costs

5.5.1 Emotional context

A number of people framed the process of being evaluated for dementia as an experience that was associated with fear, perhaps even being the most frightening diagnosis that they could imagine: "the big D" (R15 patient). People stated that courage was needed to access help "my step forwards have been going to the doctor, getting the courage to go" (L3 patient). For some people there was a sense of urgency to get the problem addressed, "I think they should be seen quickly" (E19 informant), which was increased by the desire to halt or indeed reverse the cognitive decline, or help reassure "that would give people a lot more reassurance" (M24 patient). Other people did not perceive the speed of access to appointments as important, so long as symptoms were mild and not progressing rapidly "I'm not sure about being seen quickly is any benefit because you've gradually lost your memory over 30 years ... if you can't remember anything it would be different" (G18 patient).

Implications of findings: Some people with cognitive symptoms report fear, especially while in a period of uncertainty about the diagnosis, and may benefit from a process which supports positive emotional and psychological well-being.

5.5.2 System resources

A number of participants experienced fragmented care at their general practice, which some attributed to changes in working patterns of doctors: "GPs [are] all part time these days" (C20 patient). Some people reported a tendency for people with cognitive difficulties to withdraw from society "I've lost the group situation" (Z11 patient), and some thought that a whole community approach might help people feel integrated "I think with memory people have to be sympathetic. I think sympathy has gone out of the window... having neighbourly relationship we learn from each other" (S13 patient). There were a range of views on the extent to which care should be provided locally to home. For people who were able to drive, some were happy to travel though the acceptable distance varied from the local GP surgery, to the major city, or indeed hundreds of miles. For people who could not drive themselves there were a variety of approaches for getting to appointments, namely public transport, taxis, family, or community transport. Some people experienced difficulties with travel arrangement as a structural barrier to care, for instance if "the buses don't go where I want to be" (S14 patient), but for others it was less of an issue: "travelling these days, [is] quite easy, you can go anywhere quite quickly" (E19 informant).

Implications of findings: Some groups of people may require particular assistance, perhaps offered through community or voluntary organisations, to ensure they experience equitable access to diagnostic resources. Continuity may need to be intentionally designed to be understood at the organisation or system level, for example with a consistent visual identity, to mitigate for discontinuity at the level of the individual.

5.5.3 Own resources

For some participants cognitive difficulties made the business of daily life challenging. Advice on how to handle the day-to-day was reported to be as useful as the bio-medical and pharmacological expertise of a consultant "I think the day-to-day thing of having somewhere to go to perhaps to spend a day or do things and people will give things to do that will be helpful, I think that's just as important as a consultant to be perfectly honest" (E23 informant). Some people had developed strategies to aid their memory, or were interested in whether changes to diet and lifestyle might improve memory. There was a desire amongst some people to be able to do something for themselves. Other people felt that "in fact I was overwhelmed with, with information" (E19).

Implications of findings: People undergoing diagnostic evaluation for cognitive disorders should be offered resources to enable effective self care and management, delivered in a self-paced way.

5.5.4 Responsive assessment

A variety of accounts indicated that a diagnostic evaluation for dementia needed to be personalised and consider whether any impairment was attributable to medication, sensory impairment, or stress "I think there's other reasons for not remembering other than something wrong with your brain" (S14 patient). Some people reported that they had not received an assessment that actually evaluated the concerns and problems that they were experiencing "I do feel that a lot of his problems haven't really been addressed" (C20 informant). For other people, the initial assessment that had been done by the GP was not sufficiently in depth or challenging to identify a problem "from where we were coming from it seemed a little bit kind of light and he was feeling that he needed to perhaps that we needed something more" (R15 informant). There were sensitivities around the language that was used in the diagnostic process, with some people being irritated by the idea of needing a carer as this implied a degree of severity that did not correspond to their current difficulties "we have in the last two or three months become irritated by the concept that [wife] is my helper, my carer, I'm not in requirement of a helper or a carer" (R15 patient).

Implications of findings: People with cognitive problems should be offered an opportunity to express their personal concerns and supported to perceive their assessment as personalised to them.

5.5.5 Important condition

A range of people viewed dementia as an important and complex problem that required training to be able to diagnose properly: "It's a very serious thing to be told you've got dementia" (R22 patient). For some interviewees, the job title of the person who was making the diagnosis was not important, so long as they were trained, and competent "I could be persuaded that the right nurse with the right training, the right background and the right understanding could be the mechanism in the general area. If not for anything else other than to sift through them" (R13 patient). Indeed, some people conceptualised diagnosis as "a list of questions which are arranged by experts probably and they

can ask those sorts of questions and ... at the bottom you've got a total" (E27 patient). For others there was a need for the person who was diagnosing to be trained to the same level as a specialist. A number of people held the view that dementia was too serious and specialist to be diagnosed by GPs. For some people the desire to access help was so great that they would be willing to travel long distances, even abroad, to see a someone who might be able to help: "I mean NHS says to me, sorry we can't do any more for you, and – but there is this genius in United States for instance, I would like to go there and see him" (M23 patient). Some interviews indicated that a diagnosis of dementia by a non-specialist would be subject to challenge or suspicion on the part of the patient because they would consider the GP to be "misguided" (C25 patient) if they diagnosed dementia, whereas if a specialist made this diagnosis they would be "more [pause] prone to take notice of what was said and what was diagnosed" (C25 patient). In contrast other people stated they "would accept it [the opinion of a GP] not because I'd be glad to or that I feel I should but I... I don't think I would quarrel with him if that was his opinion" (M25 patient). Various accounts indicated that continuity with a particular GP was perceived as being a necessary condition for a GP based diagnosis of dementia: "Well if they're only seeing you the once, they probably wouldn't be able to diagnose that, but if they saw you a few times, I think they could" (E30 patient).

Implications of findings: Accreditation, re-certification, and quality standards may help patients to be assured of a high quality services underpinned by rigorous training.

5.5.6 Diagnostic process

There was a recognition that it was important to have an explanation for cognitive problems and symptoms, but for some people the diagnostic process was not just about getting a diagnosis or label, but was inseparably linked to a desire for practical advice, an indication of prognosis or what to expect, and an offer of interventions "Because I want to know one way or the other, what it is, what's the diagnosis, what's the outcome or non-outcome?" (Z9 patient). A range of people recognised that there was, or could be, a hierarchical (or step-wise) process to getting diagnosis. Some accounts indicated that the stepwise approach was frustrating and unhelpful "it's been a very long drawn out process and I don't feel anybody's on top of anything. We feel in the wilderness" (E25 informant) whereas others accepted it "you've really got to wait to allow the authorities to sort it out" (G18 patient). A number of people either explicitly or implicitly recognised a series of trade-offs between expertise, access, and continuity "if you're insisting on going right up to the top, it's stupid, because probably you've got to wait an awful long time to get in to see them" (L3 patient). Some interviewees stated that interpersonal factors were an important factor in the extent to which the diagnostic process was considered as acceptable "he was very good actually ... There again if it was the err someone in their 20 something I don't think I would be quite happy about it because they wouldn't have the experience" (M22 patient).

Implications of findings: Diagnostic pathways should be designed so that they can provide not just a diagnostic label, but also advice and interventions. The number of encounters and stages in

the process of reaching a diagnosis should be kept to a minimum.

5.5.7 Opportunity costs

Some participants recognised that resources in the health system were limited, and were stretched because of the demand for healthcare "I mean, we all know the NHS is overstretched and you don't want to overburden it but I think there is a role for a doctor with one in the area that has extra training, different tools" (R13 patient). One view was that GPs did not have the capacity to take on additional work, or would have to drop something else to be able take on the diagnosis of dementia: "I can't see them sort of fulfilling a leading role in it because he wouldn't have the time, even if he had the expertise" (Z11 patient). Some people took the view that if diagnosis was not of practical benefit to patients then perhaps money should be invested in social care instead "If early diagnosis makes no difference then there's a case to be made for a little pot of money to help people get through the first stages and more to put into social care" (R13 patient).

Implications of findings: Organisations should be resourced to provide a service that monitors demand and is able to anticipate and plan changes in the capacity it is required to deliver.

5.6 Interpretation of results

As outlined in subsection 5.4.2.7 the TFA was chosen as the most appropriate theoretical lens through which to view the empirical findings. The TFA defines acceptability as follows:

> *"A multi-faceted construct that reflects the extent to which people delivering or receiving a healthcare intervention consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention "* *[368]*

Table 5.4 outlines how the TFA fits against the the empirically derived categories and describes some possible implications of the findings. If the seven aspects of the TFA are accepted as reflecting the acceptability of a diagnostic evaluation for dementia then to an extent, some of these apply regardless of whether the diagnosis takes place in a generalist or specialist setting, and some are particularly important in one or other setting. Based on the interview accounts, *perceived effectiveness* and *ethicality* were judged to be particularly influential on the acceptability of a diagnosis by a GP. The other components of the TFA were considered be important to consider regardless of where the diagnosis took place and so less specifically relating to acceptability of a diagnosis in general practice.

Perceived effectiveness is "the extent to which the intervention is perceived as likely to achieve its purpose"[368] and corresponds to the concept of responsive assessment, because people who are undergoing evaluation for possible dementia reported that they wanted the process to account for individual differences and the particular individual problems; an unresponsive assessment may leave an unmet need [369]. Diagnostic evaluations by GPs could be responsive by explicitly asking how people would like to be addressed and eliciting the scope of the concerns in an open,

relatively unstructured, and patient focused encounter, which is typically within the expertise of a GP. Conversely it would be harder for GP evaluations to be responsive by addressing the specific concerns with a bespoke assessment that will evaluate the particular concerns that the patient has noticed, because GPs have a limited range of clinical phenotypes (see Section 1.1.3.3) that they can compare a new case against, and limited expertise in using a variety of cognitive tests. These factors make it unlikely that an evaluation by a GP would be totally bespoke and nuanced to the individual problems that were faced by the patient, whereas a specialist assessment might have the opportunity (whether or not it is done in practice) to offered a fully tailored evaluation. One response to this issue would be for a GP to focus initially on taking a clear history about the scope and form of the cognitive problems and then to evaluate only those phenotypes that were within their experience, and refer the others to a specialist. Alternatively, future technology could be used to target the testing of cognition to problematic domains. The crucial point is that the diagnostic evaluation must be responsive to the needs and the problems of the individual and be comprehensive and plausible.

Ethicality is "the extent to which the intervention has a good fit with the individuals value system" [368] which maps to the important condition concept in the empirical classification because some people with memory problems report that dementia is a serious condition and have a level of expectation that they will be evaluated comprehensively by an expert. GPs in the United Kingdom practice within a professional framework set by the General Medical Council, and this requires doctors to [368] "recognise and work within the limits of your competence" [370]. An accreditation and certification process, and evaluation of care provision against approved quality standards may help GP providers of a diagnosis of dementia to meet their professional obligations and the public expectations that were described in the interviews. Any service commissioned from GP providers would need to incorporate the costs of training and accreditation.

Five of the seven components of the TFA were considered as being generally important to consider regardless of whether diagnosis is made by a GP or a specialist, and therefore less specific as factors that would determine the acceptability of a diagnosis made by a GP. However, they are still important factors that determine the acceptability of a diagnostic evaluation, regardless of where that evaluation takes place. The emotional context (TFA: *affective attitude*) that people with cognitive symptoms and their kin perceived while experiencing cognitive symptoms and waiting for a diagnosis or explanation was often one of fear and uncertainty. Acceptability of diagnostic interventions might be enhanced by supporting positive emotional and psychological well-being for patients and kin, encouraging autonomy and self-empowerment by providing information about timescales, appointments and action to take at times of deterioration or crisis. System resources (TFA: *burden*) that could improve acceptability of diagnostic interventions might be the identifying and mitigating barriers to care, especially for groups who may require particular assistance, such as those without a car or with low levels of (health) literacy. Additionally, intentionally designing continuity into the system (e.g. with a consistent visual identity, or a named caseworker) might help to reduce the number of steps or transitions in the diagnostic process. Providing self resources (TFA: *Self efficacy*) or strategies

that could be used by people experiencing cognitive problems to manage their symptoms and enhance their own autonomy and independence might also improve acceptability of diagnostic evaluations, because accounts indicated that these are often desired. A particular finding of the empirical work was that it is important for self care resources to be self paced; this finding fits within a well established literature reporting varying levels of patient activation, "an individual's knowledge, skill, and confidence for managing their health and health care" [371]. Empirical accounts indicated that the diagnostic process (TFA: *intervention coherence*) should address needs for advice and interventions, not a just a diagnosis. Through this lens, diagnostic pathways may be viewed as more acceptable when they are joined-up and structured, keeping encounters to a minimum while remaining comprehensive in scope, and considered in the context of the findings about self efficacy. *Opportunity costs* was an item in both the TFA and in the empirically derived classification, and serves as a helpful reminder that pathways should be resourced not just to be able to meet the current demand but also to monitor and respond to changes in the balance between demand and capacity.

Disregarding the impact of the two other factors in the TFA (perceived effectiveness and ethicality) then there is no clear reason why diagnosis of dementia in general practice would be any more or less acceptable than a diagnosis in secondary care, given consistency of the five factors in the TFA that are generally applicable to the acceptability of a diagnostic evaluation regardless of setting (affective attitude, burden, self-efficacy, intervention coherence, opportunity costs).

Table 5.4: Implications of qualitative findings

Aspect of acceptability	**Example**	**Possible implications**	**Aligns to TFA** ∗
Emotional context	Some people with cognitive symptoms experience fear about getting worse while waiting for an explanation for their problems	At initial encounter provide information about indicative events†, timescales, prognosis, and when to seek help	Affective attitude
System resources	Some report lacking continuity with a particular GP Accessible care and a cohesive community approach to dementia were also important	Engage with local community partners e.g. voluntary organisations, in partnership with local care networks	Burden
Self-resources	Some people want information about self-care for their symptoms; others are overwhelmed by mis-timed information or coping with daily life	Provide evidence-based self-management strategies that are accessible at the time of need e.g. guided handbook or online	Self-efficacy
Responsive assessment	One view was that the cognitive evaluation did not address the patients specific problems. Some people found certain terms distressing.	Perform evaluations that explicitly elicit and assess the patient's concerns using agreed language.	Perceived effectiveness
Serious condition	Some people were concerned about misdiagnosis. Evaluations were said to require trained staff, with sufficient time (or serial assessments) to evaluate the cognitive problem(s) in detail	Evaluations should be done by trained, credible staff who have sufficient time to perform a plausible assessment	Ethicality
Diagnostic process	Some people linked diagnosis to an explanation about prognosis, interventions and practical advice. Some people wanted help to understand their progress in a system they perceived as hierarchical.	The diagnosis - or explanation of symptoms - should be provided with advice on the next steps. Patients should have a consistent single point of access for queries.	Intervention coherence
Opportunity costs	Some people said that GPs had limited capacity to take on extra work.	Any diagnostic pathway should be resourced.	Opportunity costs

∗ *TFA* Theoretical framework of acceptability

† *indicative events* e.g. neuroimaging or first outpatient appointment

5.7 Discussion of qualitative work

This chapter has attempted to answer the following research questions using semi-structured qualitative interviews of people with cognitive symptoms and their kin:

- How acceptable is it to patients and family members for general practitioners to make a diagnosis of dementia independent of specialist input?

5.7.1 Summary of main findings

Acceptability 26 interviews were done with people with symptoms of dementia and (if they wished) a third party (their kin). A range of views was elicited from a group of people who represented a diverse mix of cognitive diagnoses, age and sex, and GP practice. A framework analysis indicated that there are seven aspects which influence the extent to which it is acceptable for general practitioners to make a diagnosis of dementia independent of specialist input. Interpreted through the lens of the TFA the empirical work suggests that the extent to which such an innovation is acceptable will be increased by a process that: takes account of the (often fearful and pressing) emotional context, is accessible and cohesive, recognises the daily challenges that are experienced by those with cognitive problems and the desire of some people to self-care, is responsive to the needs and the problems of the individual, evaluates the problems with a comprehensive and plausible evaluation, holistically relates the desires of diagnosis advice and intervention, and is properly costed and resourced.

5.7.2 Relation to literature

The current study identified that for some people the process of diagnosis occurs within an emotional context of fear and a sense of time urgency. Previous reports have identified the emotional impact that a diagnosis of dementia has on a person, including a loss of identity, anger, uncertainty, and frustration [372]. Similarly the diagnostic process has previously been identified as being frightening and disorganised [373]. Many respondents to the empirical interviews reported a lack of continuity and that having a care service that was accessible and cohesive were important. Similarly, participants in previous studies have reported fragmented routes to a diagnosis and a lack of support and follow-up after the diagnosis [374].

The current study found that dementia was regarded as a serious condition, that sufficient time and resources would be required to provide an acceptable diagnostic service, and that there would be opportunity costs in having the service provided by GPs. Similarly, other investigators have reported the view that some people would rather be told that they have cancer than dementia [375], that GPs had reservations about making an incorrect diagnosis of dementia [340], and that a primary care led dementia service would need to be appropriately resourced and sufficient time made available [340].

The finding that people link diagnosis to an explanation about prognosis, interventions and practical advice has also been reported by other investigators [372–375]. Indeed, more generally a

hypothesised model for a lay understanding of ill-health is that patients typically are seeking answers to six questions [174]:

1. What has happened?
2. Why has it happened?
3. Why to me?
4. Why now?
5. What would happen if nothing was done about it?
6. What should I do about it, or whom should I consult for further help

Some discordant views should be specifically noted. For instance, some people stated that waiting for a diagnosis was not important because cognitive symptoms typically developed slowly, and this is not in keeping with the literature in general, or with the rest of the respondents. Similarly some people viewed diagnosis as a tick-box exercise and stated that it could be done by anybody with the appropriate tools, which again is somewhat discordant with the rest of the literature.

5.7.3 Strengths and limitations

A strength of the work is the diverse range of people that contributed to the interviews, and the systematic and repeatable approach to the analysis using Framework. The interviews regarding acceptability continued until there were no new emerging themes, and the findings of this aspect of the work are cohesive with the TFA model of acceptability.

As described in Section 5.4.1 the Framework Method was chosen because it uses a systematic approach and is particularly appropriate for the analysis of semi-structured interviews [362]. Three key limitations of the Framework approach have been described [362]. Firstly, the risk that for those with a quantitative research background there will be a risk of quantifying qualitative data. Secondly, that it can be time-consuming and resource intensive, like all qualitative research methods. Thirdly, that there is a training component to the use of the method in a multi-disciplinary team. Measures were taken to mitigate the impact of these limitations on the research described here. Firstly, efforts were made to avoid quantifying research findings, other than in the description of the participants. Terms such as *some*, *a number of*, or *a range of* were used in the reporting of the findings to avoid inappropriate quantification. Secondly, dedicated time was allocated to the qualitative research during the investigations. Recruitment to the qualitative interviews started towards the end of the quantitative investigations, when research clinics were occurring only once a week, rather than two or three times a week and this supported an opportunity for reflexivity regarding the qualitative investigations. Thirdly, the Author undertook two formal training courses on Qualitative methods and Framework at the NatCen National Centre for Social Research, in addition to a three-day introductory

course at the University of Bristol. Additionally, the Author had a dedicated qualitative advisor with significant experience of conducting qualitative research, who supported coding and reflexivity.

Although a wide spectrum of people were invited to take part in interviews, it is arguably a limitation that the people who were judged as being most frail were not invited, and the presented classification may not adequately reflect the views of this group. Indeed, by definition, the range of views is restricted to people who were willing to participate in research. To the extent that people who take part in research are different from those who do not, the opinions expressed may not reflect the views of the entire population. In particular, it is plausible that people who take part in research are willing to accept new ideas and suggestions, and so they may have been more open to the idea of a GP diagnosis of dementia that the general population. A second limitation relating to the range of views that were obtained is that all of the participants were recruited from BNSSG and the findings are only transferable to the extent that the views are applicable to people in other settings. It is plausible that people in BNSSG may be more open to the idea of GP diagnosis of dementia than others because of the GP dementia diagnosis service described in Section 3.2.2, which was restricted to practices in Bristol. Of the 26 participants in the qualitative interviews, four were from practices in Bristol and the remaining 22 were from practices which were not part of the GP diagnosis service. It is very likely that the views expressed reflect range of experience in dementia diagnosis services across 10 different practices and three different clinical commissioning groups. Although the interviewees were all participants in research, the interviews elicited a wide range of views, including some views that only a specialist could diagnose dementia (Section 5.5.5), and the interpretation is a synthesis of the views expressed using Framework. Section 5.7.4 explores the impact of conducting triadic rather than dyadic interviews.

A further limitation is that there were difficulties in answering the research question addressing burden of tests. These difficulties are outlined in Section 5.2.1.

5.7.4 Trustworthiness and reflections

The credibility [376] of the findings is particularly supported by prolonged engagement and persistent observation. One interviewer conducted all of the interviews with people. The interviewees had all been encountered previously at the quantitative research clinic, and the researcher had been considering the setting and context of the research for a year before the interviews were done. The researcher was immersed in the data through the process of doing the interviews and reviewing the transcripts, audio and video recording, and coding. As the interview guide was refined during the process of the interviews the most important issues for the research question were crystallised and persistently observed. Data were triangulated by specifically seeking dis-confirming views, and by adding new questions as interviews progressed to explore ideas that had been proposed by earlier participants. Quality assurance was done by having 20% of transcripts double coded.

There are at some important reflections on the qualitative work. Firstly, the interviews and analysis were conducted by a researcher who is also a GP. The investigator therefore undertook the studies

with some pre-conceived ideas about what might be important to people seeking professional advice about cognitive symptoms. In particular the researcher was of the opinion that the *acceptability* of a diagnosis was a function of the *credibility* of the process that had led to that outcome and the person conveying the diagnosis, and that credibility was related to factors such as a person's job title (more credible if made by a Professor), the built environment that the encounter took place in (more credible if polished and clean), the extent to which expectations were met, and the emotional response to the encounter (underpinned by relational and interpersonal factors). Secondly, the process of doing interviews with people with cognitive problems was challenging at times. Specifically, it became clear over the process of the interviews that it was more difficult than anticipated to ask people who had cognitive impairment to talk in relatively abstract terms about the pros and cons of different approaches to diagnosing a significant condition that they (and their kin) were emotionally involved with. Notwithstanding that qualitative research can take a critical realist stance which recognises multiple different perspectives, or a post-modern approach which understands there as being no objective truth, there are particular issues when interviewing people with cognitive impairment.

In interviews and encounters with people who were experiencing cognitive symptoms, or who may have had sensory impairment, there was a risk that both the reception and transmission of information may be partial or misleading, as those with cognitive impairment may not have been able to communicate their full intended meaning, or they may have been distracted. Reflecting on this process, the joint interview with both the patient and their kin may have provided a support to the person with cognitive symptoms and helped to ensure that their views were as well articulated as possible. As described in Section 5.4.2.6 the transcribed data were charted (side-by-side) using a Framework chart separately for patient and informant respondents to indicate the person who had contributed the view. Participants were *offered* the option of taking part in the qualitative interview with their kin, but this was not mandatory. Alternative approaches to data collection would have been to mandate either dyadic or triadic interviews, or to conduct focus groups. It is likely that mandating either dyadic or triadic interviews would have led to a reduced range of views being obtained for two reasons. Firstly, it is plausible that some patient participants would have declined to take part if they had to take part alone, because of concerns from them or their (often protective) kin about their wellbeing. Secondly, it is plausible that in triadic interviews the views of the respondents may have been influenced by one another; arguably this is similar to the use of focus groups and so is a strength because it supports the development of ideas of views and ideas as they are articulated and people respond to the developing discussion. In contrast, it is very likely that many of the patient respondents would have struggled to have participate fully in a focus group because some people had difficulties maintaining focus in a triadic - or even dyadic - interview.

Participants had a range of cognitive engagement with the interviews. Some participants were able to give clear answers to questions, answering questions clearly and recalling events well, including on occasions specific details of the research clinic some months prior to the the qualitative interview. Other people had difficulty staying focused on the interview, for example talking at a

tangent to the question about a previous medical problem, or on the other hand giving a very detailed and drawn out response to a relatively simple question. Some people denied that they had any problem with their memory but then would repeatedly ask for questions to be restated, and not obviously because of sensory impairment. Some people had quite advanced cognitive impairment and struggled to given specific answers to clear focused questions, or to find the right words to convey their intended meaning.

5.7.5 Conclusion and implications

Table 5.4 summaries some of the possible implications that emerged from the interrogation of the classification in Section 5.5. The coherence between the TFA and the classification in Table 5.4 provides support to the empirical findings. If commissioners were to introduce any of the recommendations in Table 5.4 to the health system then the TFA might provide an appropriate framework to evaluate the impact on the acceptability of the diagnostic process.

One view is that the findings described in Section 5.5 are recycling conventional wisdom. Certainly for people who are interested in the delivery of healthcare and in particular the care of people with dementia (such as the Author) some of the findings resonate with accepted views on good practice; indeed the pre-conceptions of the Author prior to conducting the interviews have already been discussed in Section 5.7.4. Likewise, as discussed fully in Section 5.7.2 other investigators have reported many similar findings previously, particularly with respect to the emotional context surrounding evaluation for cognitive symptoms, reports of fragmented services, the perception that dementia is a serious condition, and the desire for practical support not just a diagnosis. Some of these findings, such as the need for coordinated care, that engages with the voluntary sector, and provides responsive information and support at the point of diagnosis are so well accepted by professionals that they are incorporated in national guidelines [377]. To the extent that the views of the participants, who may or may not have read the guidelines (either is plausible), are consistent with accepted wisdom the findings may simply reflect that people are describing an accepted view. An alternative stance is that the consistency between the interviews and accepted wisdom supports the findings.

The Author proposes in Section 5.1.1 that some of the components that define acceptability of a diagnostic evaluation are important regardless of the setting in which that evaluation takes place and arguably this view is supported by the consistency between interviewee opinions on diagnosis (in general practice), and national guidelines on diagnosis (typically in specialist settings). Section 5.6 suggests that the most important components of acceptability of diagnosis in general practice relate to the TFA domains *perceived effectiveness* and *ethicality*. Regarding *ethicality* while current guidelines [377] advise on the need for staff training, they do not clearly recognise the need for *showing* people undergoing evaluation that their care-providers have been specially trained, for example with certificates or quality assurance plaques being displayed in consulting rooms and on correspondence. With respect to *perceived effectiveness*, while recognising the need for person

centred care, current guidelines [377] do not currently explicitly recognise the need for agreeing a shared language (viz. *kin*) when discussing cognitive symptoms and the impact of these on the life of the person undergoing evaluation.

Given the acknowledged emotional context of people undergoing evaluation for people cognitive symptoms these findings may have relevance for delivery of diagnostic services in all settings. Clear display of quality assurance processes may enhance confidence in the diagnostic service. Explicit agreement of a shared language and request by staff to forgive any unintentional "mis-speaking" may help to humanise encounters. Taken together these relatively simple measures may make a stressful and unsettling process less distressing regardless of the setting.

For research Further research might apply the framework to a clinically commissioned pathway to evaluate the extent to which that service was acceptable and met the needs of service users. Additionally there might be merit in exploring the extent to which patient activation measures [371] can be used to target the information given to people using the service and whether giving targeted information results in a more acceptable service.

There is likely to be value in using other additional qualitative methodologies to better understand the experiences of people who are being evaluated for cognitive symptoms. One approach would be a discourse analysis of reports on internet forums and social media to identify novel insights, especially from people who do not participate in focus groups, interviews or quantitative surveys. A second approach would be an ethnographic investigation into the experience of people who are undergoing evaluation for cognitive symptoms which would give valuable insights, especially because for people with cognitive symptoms there may be difficulty in recall at a later date. Because of the emotional context that was identified in this study and has been described previously, an ethnographic investigation would arguably be especially valuable if it could recruit people shortly after they presented to their GP with cognitive symptoms, so the investigators could capture some of the experiences and views of people in the *process* of evaluation and set within the life context, rather than simply collecting data from *events* such as clinic appointments.

For clinical practice The findings of this study would support a GP based diagnosis of dementia for some people with cognitive symptoms, in some circumstances, provided that clinicians were provided with adequate training, resources, and time and so long as the service fully addresses the agendas of people seeking help: not just a diagnosis but also an explanation about why it has happened, what can be done about it (including self care) and what will happen next [174].

Chapter Summary

This Chapter has presented the rationale, methods, characteristics, data analysis, findings and discussion of the qualitative investigations. The qualitative research question was *how acceptable is it to patients and family members for general practitioners to make a diagnosis of dementia independent of specialist input*. Participants were purposively sampled from the main quantitative study for maximum variation in characteristics of referring practice, age and sex. People who were judged to be in an advanced stage of frailty, either physically or cognitively, were not invited. Interviews were arranged on a separate day to the research clinic.

The data were managed and analysed using a framework approach. A combined inductive/ deductive approach was taken using a seven step approach to data analysis: transcription, familiarisation, coding, developing a framework, applying the framework, charting the data, interpreting the data.

The data were mapped into a seven-item classification which was judged to reflect the elements that contribute to the acceptability of a diagnostic process. The findings were interpreted through the theoretical lens of the Theoretical Framework of Acceptability. The seven aspects of acceptability that were judged to comprehensively reflect the data were: the *emotional context* of fear about symptoms getting worse, *system resources* that contributed to a sense of cohesiveness and continuity, the challenge of supporting people with their *own resources* but not overwhelming them, the need for a *responsive assessment* that addressed the specific problems that were reported, the importance of dementia as a *serious condition* and the need for training and time for an assessment, the view of the *diagnostic process* as comprising more than just diagnosis, and the recognition of *opportunity costs* such as GPs having limited capacity for extra work.

The findings indicated that a GP based diagnosis of dementia would be acceptable in some circumstances if there were sufficient resources, training, and quality assurance, and if the service fully addressed the needs of patients, providing not just a diagnosis but holistic care.

CHAPTER 6

DISCUSSION

This Chapter discusses the results of the current work. The main findings are summarised, of the systematic review, the current diagnostic test accuracy study, and the qualitative work. The risk of bias and strengths and limitations of the work are considered. The findings are interpreted in the context of the literature and clinical practice. In the final Section the implications for practice and research are considered.

6.1 Summary of main findings

6.1.1 Systematic review

The systematic review included 10 studies, of which seven could be included in the meta-analysis. The seven studies in the meta-analysis included a total of 2550 people, of whom 694 (27%; 95% CI 26% to 29%) had dementia. The summary accuracy of GP judgement for the diagnosis of dementia was LRP 5.3 (95% CI 2.7 to 10.7) LRN 0.46 (95% CI 0.33 to 0.66). In individual studies sensitivity ranged from 34% (95% CI 17% to 54%) for Cambridge to 91% (95% CI 85% to 96%) for Mannheim, and specificity ranged from 58% (95% CI 51% to 66%) for France to 99% (95% CI 97% to 100%) for Zwolle. The summary accuracy of GP judgement for the diagnosis of cognitive impairment (including dementia) was LRP 3.8 (95% CI 2.6 to 5.8) LRN 0.25 (95% CI 0.09 to 0.72). In individual studies sensitivity ranged from 58% (95% CI 49% to 67%) for Amsterdam to 97% (95% CI 93% to 99%) for Mannheim, and specificity ranged from 51% (95% CI 44% to 57%) for Mannheim to 88% in Amsterdam (95% CI 84% to 91%) and Zwolle (95% CI 84% to 92%). Cognitive impairment was chosen as a target condition for the systematic review because that was commonly used by original studies. For the re-coded target condition normal, the diagnostic accuracy of GPs was sensitivity 79% (95% CI 57% to 92%) specificity 80% (95% CI 45% to 95%). The summary points are overly-simplistic summaries of the data because of the heterogeneity

and the summary accuracy may be better reflected by the summary ROC curve in Chapter 2. The most important limitations to the meta-analysis were the substantial heterogeneity in the data and that five studies were at high risk of bias in the *flow and timing* domain, because of incomplete verification of the target condition.

Investigation of heterogeneity was limited by non-convergence of meta-regression models. It was difficult to draw firm conclusions about heterogeneity but the data were compatible with studies that used ICD-10, or applied retrospective judgement, having higher specificity compared to studies with DSM definitions or using prospective judgement. For sensitivity there was no evidence of heterogeneity by definition, studies that used a prospective index test may have had higher sensitivity than studies that used a retrospective index test, and studies that allowed access to the medical records may have had higher sensitivity than studies that did not or where this was unclear. Studies at high risk of bias in the flow and timing domain appeared to have lower specificity and sensitivity than studies at unclear or low risk of bias in this domain. However, all of these findings could have been due to chance and there was significant uncertainty.

There were five key implications of the systematic review. Firstly, GP judgement results in better classification than random allocation to disease state based on prevalence alone. Secondly, many identified studies were at risk of partial verification bias. Thirdly, none of the identified studies had investigated accuracy of GP judgement in combination with a test. Fourthly, the accuracy of GP judgement alone was insufficient to rule in dementia. Finally the accuracy of GP judgement for the diagnosis of dementia in people with symptoms was reported by only one study, whereas the other eight studies reported the diagnostic accuracy in unselected people presenting to their GP.

6.1.2 Diagnostic test accuracy study

From 21 participating GP surgeries with a total population of 34,956 aged over 70 years, 465 people were referred to a test accuracy study, 241 were seen (126 men and 114 women) and 240 were evaluated. Of the 240 people who were evaluated, 132 (55%; 95% CI 48% to 61%) had dementia and 47 (20%; 95% CI 15% to 25%) were normal. On the ACE3, people with dementia had lowest standardised scores in the memory domain, compared to people with MCI who had lowest scores in fluency, and people who were normal who had lowest standardised in the attention domain. Mean duration of tests varied from one minute (EPSS) to nine minutes (MOCA). People with dementia had lower mean index test scores, and a longer mean index test duration, than people with normal cognition. Informant measures had similar diagnostic accuracy and discrimination, regardless of whether they were completed by an interviewer or self-completed, and had high levels of agreement as assessed with weighted kappa.

GP judgement as a single test had LRP 5 (95% CI 3 to 9) LRN 0.5 (95% CI 0.4 to 0.6) for the target condition dementia and LRP 5 (95% CI 3 to 9) LRN 0.7 (95% CI 0.5 to 0.8) for the target condition normal.

In combination with GP judgement, TAC and MAT both had LRP ∞ for the target condition

dementia (95% CI incalculable). The LRN for GP + TAC was 0.8 (95% CI 0.8 to 0.9), whereas it was 0.5 (95% CI 0.3 to 0.7) for GP + MAT. Confidence intervals overlapped preventing inference about superiority of particular test combinations. At threshold probabilities of 80% to 90% the GP combined tests with the highest net benefit were GP + SPMT, and GP + Eurotest. GP + SPMT had a LRP of 13 (95% CI 5 to 35) and GP + Eurotest had a LRP of 16 (95% CI 5 to 49). At threshold probabilities of 93% to 98% the combination GP + TAC had the largest net benefit. The treat-none approach had the largest net benefit at threshold approaches of more than 98%.

In combination with GP judgement, Sniffin sticks and MOCA both had LRP ∞ for the target condition normal (95% CI incalculable). The LRN of GP + Sniffin sticks was 1 (95% CI 0.9 to 1), whereas it was 0.9 (95% CI 0.8 to 1) for GP + MOCA. Confidence intervals overlapped preventing inference about superiority of particular test combinations. At threshold probabilities of up to 81% the GP combined test with the largest net benefit was GP + IQCODE, which had a LRP of 29 (95% CI 7 to 123). At threshold probabilities of 82% to 93% the GP combined test GP + MOCA had the largest net benefit but at threshold probabilities of more than 93% the treat-none approach had the largest net benefit.

For the target condition dementia, there was some evidence that tests had *lower* sensitivity when they were used as continuous scores in GP combined tests compared to when they were analysed at the published thresholds, whereas specificity was relatively similar. For the target condition normal there was no evidence of a difference when tests were analysed as continuous scores or when analysed at the published thresholds. For the target condition dementia there was some evidence of heterogeneity in diagnostic accuracy by age, with GP combined tests frequently having *lower* sensitivity in people aged < 80 years than those aged ≥ 80 years, whereas specificity was often *higher* in younger people; there was also some evidence that GP combined tests were *more* sensitive in women than in men without loss of specificity. For the target condition normal there was some evidence that GP combined tests often had *higher* sensitivity in people aged < 80 years than those aged ≥ 80 years whereas specificity was often slightly lower; there was some evidence that GP combined tests had *lower* sensitivity in women compared to men.

6.1.3 Qualitative interviews

From the 241 participants in the quantitative study, 26 people were recruited to semi-structured interviews to investigate the acceptability of GPs making a diagnosis of dementia independently of specialist input. Interviews were done with 10 women and 16 men, aged 70 to 89 years, with ACE3 scores of 62 to 98 out of 100, and of whom nine had dementia, 11 had MCI and six had normal cognition. A framework analysis was used to map the data into a classification which was interpreted through the lens of the Theoretical framework of acceptability (TFA) and judged to reflect the elements that contribute to the acceptability of a diagnostic process.

Seven aspects of acceptability were identified: the *emotional context* of fear while waiting for a diagnosis, the *system resources* that could lead to lack of cohesion and continuity, *self-resources* which could support self-care, the need for a *responsive assessment* that addressed the specific problems

that were faced by the patient, the importance of dementia as a *serious condition* that required training and time to evaluate the problem(s) in detail, the importance of coherence in the *diagnostic process* and need for information on prognosis and interventions, and challenges of addressing *opportunity costs* such as GPs limited capacity to take on additional work.

The key findings of the qualitative interviews were that a diagnosis of dementia would be more likely to be acceptable if it were quality assured, addressed the specific concerns that were raised by the patient including their emotional needs, and provided an opportunity for interventions including self-paced resources for self-care.

6.2 Risk of bias, strengths and limitations

6.2.1 Introduction

The strengths and weaknesses of the review process are discussed fully in Section 2.5 and the qualitative interviews are discussed in Section 5.7.

The current study reported in Chapter 4 was judged at low risk of bias in all QUADAS-2 domains, but nevertheless there are some strengths and limitations of the work, and it can be considered in light of the existing literature. Additionally because the empirical test accuracy study was a comparative accuracy study there are additional methodological considerations that are discussed in Section 6.2.6. Table 6.1 compares the current study to the six (of ten) studies from Chapter 2 that used a prospective GP judgement about dementia.

6.2.2 Patient selection

As shown in Table 6.1, of the six studies in Chapter 2 that used a prospective GP judgement as the index test, only one recruited people who were concerned about symptoms of dementia. While the current study had the fewest number of people undergoing the index test, it was also one of only two studies with complete verification by the reference standard. In other studies the numbers that were verified by the reference standard ranged from 10 (1%) to 2294 (70%), and based on the number verified the current study is the fifth largest study of the seven.

The results can be generalised to people presenting to general practice with symptoms of possible dementia to the extent that the referred sample reflects the range of people presenting to general practice, but it is difficult to fully quantify this. Efforts were made to fully recruit a consecutive sample of patients, by emphasising to GPs that people could (and should) be referred even if the GP thought they were normal, or had cognitive impairment that was not sufficiently bad to be dementia, so long as someone was concerned about dementia. Efforts were made to reduce barriers to referral, such as having the research team confirm eligibility and consent patients, and utilising electronic prompts at the point of care to remind busy clinicians about the study. The generalisability of the findings in the current study is supported by the finding that GPs judged 40% of participants to have cognitive impairment that was not dementia, 36% as having dementia and 14% of participants to

be unimpaired, and so were not restricting referrals to people who they believed to be impaired. In contrast, in the only other study of people with symptoms GPs judged 48% of the participants as being normal (of the 5% of these who were seen by a specialist, 32% had dementia) and 26% as having dementia. In the four other studies that reported data on the level of GP judged impairment the proportion judged to have dementia ranged from 6% in AgeCoDe to 33% in Hawaii.

In the current study, participants had a mean age of 80 years, 47% of participants were women, and 55% had dementia. In contrast, as shown in Table 6.1, the mean age in the six most comparable existing studies was 75 years to 85 years, and the proportion of female participants ranged from 63% to 84%. In the five existing most comparable studies that recruited people regardless of symptoms the proportion with dementia ranged from 2% to 29% whereas in the one other study that recruited only people with cognitive symptoms the proportion with dementia was 50%. Overall, the proportion of people with dementia and the age of participants is comparable to existing studies.

A notable finding shown in Table 6.1 is the difference in the proportion of women participants between the current study and the six other studies. Despite this, the proportion of men who had dementia, 54% (95% CI 45% to 63%) was similar to the proportion of women who had dementia, 56% (95% CI 47% to 65%). The apparently high proportion of men in the sample relative to other studies could be a chance finding, however, the probability of this is low because the confidence interval for the proportion of women in the current study (40% to 54%) does not overlap with any other study (Mannheim: 68% to 71%; Sydney 80% to 87%; Hawaii 57% to 68%; Antwerp 60% to 66%; AgeCoDe 64% to 68%; France 69% to 73%). A second explanation is linked to improved survival of men between study dates. The six comparable studies were all conducted before 2005, nearly 10 years prior to recruitment for the current study, when life expectancy for men was shorter [378]: if men have died, they cannot participate in research, so the smaller proportion of men in earlier studies could be due to smaller denominator of surviving men. For the current study, data from NHS England indicates that in the recruiting practices at the time of the study there were 15,517 men and 19,439 women aged over 70 [310]. A third possible explanation relates to the sampling of participants. This study was specifically recruiting people in whom there were concerns about dementia, and these may be more likely to be *noted* in men. People living alone may be less likely to have symptoms of dementia noted by a spouse; women are more likely to outlive their partners than men, and live alone more often than men [379]. The higher proportion of men in this study may reflect concern amongst a spouse or other relative about symptoms or behaviours that do not meet criteria for dementia or MCI "the worried-spouse well". Against this, the proportion of men and women with MCI and normal cognition was similar (see Section 4.7.3). A fourth possibility is that the higher proportion of men in this study reflects concerns about cognition which do not reach formal definitions for MCI or dementia. Published incidence rates for dementia are similar in men and women up to age 80 years [17], and in this study the average age in women (81 years σ 6 years) was similar to that in men (80 years σ 6 years). Heterogeneity in diagnostic accuracy by sex in the current study is discussed in Section 6.2.7.

Excluded participants An important limitation is that the findings of this study cannot be generalised to people who were excluded, as outlined in Section 3.2. In the judgement of the author, none of these exclusions are inappropriate, as they reflect clinical practice. A further limitation is that despite providing translation services the population were largely Caucasian native English speakers. The extent to which heterogeneity may affect the diagnostic accuracy of tests is unknown, but it it is plausible that there may be some differences in the diagnostic accuracy between different ethnic groups, especially for informant completed items which on activities of daily living, which may be at least partly culturally determined [380].

Key strengths relating to patient selection The patient selection in the current study closely reflects clinical practice in the United Kingdom, with efforts to avoid people being excluded on the basis of language, transport, or appointment availability. The age of participants is comparable with the existing studies in the literature. The proportion of people with dementia is comparable to the existing literature and lower than the 75% figure from NHS memory clinics in Bristol (Section 3.7.7) which reflects the broader inclusion criteria compared to the NHS clinic. Perhaps most importantly participants were included with a range of GP opinions about the presence of cognitive impairment in people who had presented with symptoms. Most previous studies detailed in Table 6.1 had not considered the presence of cognitive symptoms, and this is a major strength of this study as it makes it much more applicable to clinical practice than the situation where index tests, be they GP judgement or cognitive tests, are done in people attending their GP regardless of symptoms.

Table 6.1: Comparison of prospective diagnostic accuracy studies

Design aspect	**Mannheim**	**Sydney**	**Hawaii**	**Antwerp**	**AgeCoDe**	**France**	**Chapter 4**
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 & verification with reference standard							
Target condition ∗	D	D	D	D	D	D	D
Verified N	407	105	303	10	2294	385	240
Verified (%) ∗	(11)	(24)	(100)	(1)	(70)	(26)	100
GP judgement (%)							
Not impaired	36	76	33	-	94	48	14
CIND	41	-	-	-	-	-	40
Dementia	23	19	33	-	6	26	36
Uncertain	-	5	33	-	-	22	-
Diagnostic accuracy of GP judgement for dementia							
Sensitivity (%)	91	42	-	100	51	73	56
Specificity (%)	76	89	-	100	96	58	89

∗ C consecutive R random D dementia (see Table 2.2 for definitions)
Symptomatic: symptoms *required* for participation
% verified = $\frac{\text{number underwent reference test}}{\text{number underwent index test}}$ % with dementia = $\frac{\text{number with dementia}}{\text{number verified}}$
† participants were not presenting with symptoms but GPs were asked to maximise the inclusion of people with suspected dementia
-: not reported

6.2.3 Index tests

GP judgement It is possible that GPs are only making a diagnosis of dementia in people with the most manifest impairment. A test which only identifies people with the most serious illness would generally have low sensitivity and high specificity. GP judgement for dementia follows this pattern with sensitivity 56% specificity 89%, but the sensitivity is comparable to some other single index tests, such as Phototest which had sensitivity 57%, and MAT which had sensitivity 63% specificity 80%. As described in Chapter 1 Section 1.1.3.3, dementia is a rare condition in general practice, with approximately 21 new cases per GP per year, out of perhaps 6,300 consultations per GP per year, and the range of clinical exemplars available to GPs for pattern recognition is relatively low. Despite the acknowledged benefits of a diagnosis of dementia, such as an explanation for symptoms and the opportunity to plan, there is no disease modifying therapy and some clinicians may view the harm of a diagnosis as outweighing the potential benefit [381].

However, as Figure 4.3 shows, people that GPs judged as having dementia had a total ACE3 score IQR of 60 to 74, with a 90th centile of 81/100 and a highest score of 95/100. This compares to published ACE3 thresholds of <82 for dementia [382], and suggests that in this study a GP judgement of dementia is not restricted only to people with severe impairment. Similarly, people that GPs judged as having CIND (cognitive impairment, but not dementia) had an ACE3 score IQR 71 to 89, which compares to the published ACE3 threshold for MCI of < 88 [382], and indicates that GPs are not being overly liberal in their identification of CIND. The lower end of the IQR is similar for GP judgement and four other exemplar tests (within five units)[1], while the upper end is slightly lower (1 - 11 units); this suggests that compared to the four exemplar tests (MOCA, Minicog, IQCODE and AD8) GPs are identifying a similarly impaired group of people as test positive at the lower end of impairment, and not identifying the less impaired people, this is reflected in the lower sensitivity for GP judgement reported in Table 4.10 compared to the other tests.

The GP judgement about cognitive impairment was made by the referring clinician. Clinicians were asked:

> *"Is your gut feeling that this person*
>
> - *Has dementia*
> - *Has cognitive impairment but not dementia*
> - *Has normal cognition for age"*

One limitation relating to conduct of the index test is that GPs were not asked to give a continuous probability of dementia, because in a pilot version of the referral form the GPs unanimously expressed concern about judging diagnosis as a probability. The consequence of categorising GP judgement rather than using a continuous probability of dementia is that it is difficult to understand whether there is heterogeneity in diagnostic accuracy of GP judgement by assessed probability of the target

[1] see Section 4.2

condition. Intuitively one might expect greater diagnostic accuracy of GP judgement when they report higher probabilities (approaching 100%) of the target condition.

Another aspect relating to conduct of the GP judgement index test is that it reflects the judgement of the referring GP, rather than the person conducting the index tests. Referrals were received from 21 different GP practices, each of which had more than one clinician. Indeed based on NHS England data which indicate an average 1754 patients per whole time equivalent GP in the BNSSG region [383] there would be an estimated 142 whole time equivalent GPs for the total 249,556 patients in the 21 practices (Table 3.2), and nearly all GPs work less than whole time [383]. It is difficult to quantify the number of different referring clinicians as mail-merged details on the referral forms are not reliable because the form may be completed at a later date by an administrator or other clinician rather than the GP making the judgement. However, the correct interpretation of the index test GP judgement in this study is that it reflects an average measure of diagnostic accuracy for many different GPs working in different settings.

A final aspect regarding the conduct of GP judgement as an index test is that it is difficult to operationalise exactly what aspects of the encounter the referring GP was integrating to form their opinion. Referring GPs were asked to:

Please write a few words about what you think led you to form your gut feeling

and responses typically comprised comments such as *"face to face presentation"*. As described in Section 3.3 study administrators were instructed to ask the practice to supply any missing data, but were not asked to make subjective judgements about the quality of information supplied. A particular limitation is that GPs were not asked explicitly whether they had or had not used any formal test to inform their judgement, though they were explicitly instructed that this *was not* necessary. Based on previous studies, GP judgement is likely to be formulated by rules of thumb [201] and is unlikely to be based on formal tests [270].

Section 6.4 outlines some possible implications for future research to better understand the thought process behind GP judgement for dementia. It is important to note that all of the above limitations described for the work in this thesis would also apply to all of the published studies on GP judgement for dementia diagnosis detailed in Table 6.1.

Non GP judgement index tests All index tests were conducted without knowledge of GP opinion or the reference standard, and were analysed at at published thresholds. Therefore the accuracy of the single index tests, the GP combined tests, and the GP 360 tests may be interpreted as reflecting the accuracy in clinical practice when a GP consults with a patient and then arranges for a cognitive test to be completed by a different clinician. It is conceivable that diagnostic accuracy of the patient-completed index tests could vary depending on whether they are completed by the consulting GP or a third party. On one hand, accuracy could be higher if completed by the consulting GP because they would be able to see how the patient completes the test and intuitively use heuristics [201] to adjust the score, e.g. for sensory impairment. On the other hand, this additional random variation

may lead to increased heterogeneity in test performance and indeed overall lower accuracy than if the test were completed in a standardised form by an independent clinician.

Key strengths relating to index tests The GP opinion and the other index tests were conducted independently which makes the findings applicable to practice when the tests are done by separate clinician to the consulting GP. GP judgement was made prospectively, at the time of presentation, in people who were presenting with cognitive symptoms, and very few studies had investigated this previously. GP judgement reflected an average of GP diagnostic accuracy about dementia, rather than the accuracy of a single clinician. A wide range of index tests were evaluated using published thresholds so that it is possible to compare the accuracy of these tests, which is a novel and particular strength.

6.2.4 Reference standard

As described in Chapter 3 Section 3.5, the original intention had been to have a consensus panel form a reference standard diagnosis about each participant, based on a specialist assessment at a research clinic, however this was not possible. In comparison with the six most similar studies outlined in Table 6.1 most (Mannheim, Hawaii, Antwerp, France) used only a single specialist to evaluate each person and apply the reference standard, whereas Sydney and AgeCoDe used a consensus classification. Consensus classification is arguably most useful for borderline cases i.e. those on the cusp of dementia | MCI and MCI | normal. In the present study, a second independent specialist was used to adjudicate borderline cases and reached a diagnosis after having access to the full reference standard evaluation at the research clinic, as well as the primary medical record six months either side of the research clinic visit.

While all participants had a consistent reference standard applied, there was a divergence in how the primary medical record was reviewed. As described in Figure 4.1 and Section 3.5 for the 220 people where a single classification could be reached according to the reference standard the primary medical record was reviewed by the author for any diagnosis that would contradict the judgement of the specialist at clinic. For 20 people where a single classification could not be reached the medical record was reviewed by the second specialist who was adjudicating borderline cases. Given unlimited time and resource it would be ideal for the reference standard assessment, including the primary medical record extract, for all cases to be reviewed by a second specialist, as had been originally planned, but despite best efforts this was not possible within the time constraints of the thesis. While acknowledging the difference in *process* between the 220 people whose primary medical record was reviewed by the author and the 20 people whose record was reviewed by the adjudicating specialist it is judged that the probability of a different *outcome* regarding diagnostic classification is low.

Key strengths relating to reference standard With respect to QUADAS-2 quality appraisal it is likely that the reference standard as applied will correctly classify the target condition according to a recognised definition of dementia, which was determined without knowledge of the index tests, and using an approach which could be repeated in a future study; therefore in line with published anchoring statements to assess the risk of bias relating to the reference standard [384] it is at low risk of bias. The reference standard was based on a detailed, standardised evaluation which included clinical history, informant interview, and standardised cognitive testing, and was performed by a single clinician with significant expertise in the diagnosis and treatment of cognitive disorders. Adjudication was used for borderline cases and follow-up data from the primary care record were reviewed six months after the research clinic to ensure there were not alternative diagnoses that would explain the symptoms.

6.2.5 Flow and timing

A particular challenge for diagnostic test accuracy studies in general practice is that it is is a low prevalence setting for many conditions. This means that to identify one person with the target condition, many people must be evaluated, which is potentially expensive. As shown in Table 6.1, in the six most comparable previous studies of GP judgement for dementia described in Chapter 2, only one verified the whole sample of people who underwent the index test (GP judgement) with the reference standard (though diagnostic accuracy data was not available), and only one other verified more than 26% of the sample (but 18 months after the GP formed their opinion). In comparison, in the current test accuracy study everybody who was evaluated by the index tests and who consented to participating in research also underwent the reference standard.

Figure 4.1 shows that of the 456 people who were referred to the clinic, 155 people (33%) declined to take part. In contrast, in [272] 40% of people who were seen by their GP with suspected dementia declined further assessment. Table 4.3 shows the known characteristics for people who declined compared to people who participated. There may be a variety of reasons why people declined to participate. Some reasons for declining are unlikely to impact the diagnostic accuracy of tests, such as inability or lack of willingness to travel, a role as a carer, or being too busy. There was weak evidence of an association between declining and a GP judgement of CIND and slightly stronger evidence for an association between declining and age or female sex (Section 4.2). It is not clear why people who are older or female would be less likely to take part, but this may reflect the greater frequency with which older people, especially women, live alone [378]. As attending with an informant was necessary, some of these people living alone may have been reluctant to ask an informant (perhaps an adult son or daughter) to attend (who would perhaps be taking time off work or from other caring responsibilities). If the proportion of people with dementia according to the reference standard, or the spectrum of disease, was different in people who declined to those who participated then the diagnostic accuracy of the tests might be affected, and it is possible that this could affect some tests more than others. If people who declined were more impaired than those who participated then it is likely that the diagnostic accuracy of most tests, especially GP judgement, would be better than reported. The number of people declining should be considered in the context of the design, to make it easy for GPs to refer they did not have to fully discuss the study or consent the patient, only obtain consent to pass their contact details to the research team. Therefore some people may have been given only very scant information about the study from their GP, and may have been reluctant to agree to something over the telephone, especially as older adults are often victims of scams using the telephone [385].

It is plausible that in some cases people with suspected dementia were not referred to the study, based on the potentially eligible figure of 1735 shown in Figure 4.1 which is derived from the age standardised incidence applied to the study population using person years at risk. Therefore it remains a limitation that GPs would have formulated a judgement about the presence of the target conditions in people who were not evaluated by the study team. To standardise the administration

of the index tests in the current test accuracy study and ensure full verification with the reference standard, it was necessary for GPs to refer people to the research study. The alternative, to have had GPs complete the full range of index tests in their practices, would have been unfeasible. Section 6.2.2 describes the potential impact of this on the results and generalisability.

The median time between referral (the date of GP judgement) and the clinic appointment (reference standard) was 47 days (IQR 30 to 72 days), the longest interval was 177 days, approximately six months, which was due to difficulties with attending earlier appointments. Dementia is a slowly progressive condition and it is unlikely that the condition would have progressed significantly in a six-month period. However, it is possible that for borderline cases impairment became more manifest over time, and that the diagnostic accuracy of GP judgement would be biased towards being lower for people who were seen with a longer delay. Compared to the studies in Table 6.1, France, Sydney, Antwerp, Hawaii and Mannheim did not specify the interval between GP judgement and the reference standard, AgeCoDe had an interval of 1.5 years. All other index tests other than GP judgement were done on the same day as the reference standard.

Key strengths relating to flow and timing All index tests, other than GP judgement were completed on the same day as the reference standard. The interval between GP judgement and the reference standard was relatively short, and unlikely to be associated with a significant progression in cognitive impairment. GP judgement was fully verified against the reference standard for all consenting people who were referred and there was weak evidence of selective participation by cognitive status.

6.2.6 Comparisons between tests

In addition to the items outlined in QUADAS-2 there are further methodological aspects that deserve consideration, arising from the comparative test accuracy design.

Firstly, the intention was that all of the participants would receive all of the index tests. Other than as described in Chapter 3 Section 3.6.2 tests were not done by participants because they were declined. Therefore the selection of participants is unlikely to have led to bias in the comparison between index tests.

Secondly, participants received multiple index tests which were (except GP judgement) conducted by the same examiner. Therefore it was impossible to blind the examiner to the results of prior tests. To mitigate this, as described in Section 3.4.1 and reported in Appendix K, the order of the index test battery was randomised to minimise order effect. The order of Sniffin Sticks was not randomised, for two reasons: firstly this test was only given to some of the participants because, as described in Section 3.4.1 it was imported on special order; secondly it was judged that it would minimise the cognitive burden on participants to do Sniffin sticks last because it was a test of smell rather than cognitive functioning. All of the index tests were interpreted without knowledge of the reference standard and at prespecified thresholds. Overall the conduct of the index tests is unlikely to have introduced systematic bias in the comparison of tests, but the performance of Sniffin Sticks may have

been different (either better or worse) if it had been done first; which would have been an alternative approach.

Thirdly, all of the index tests were verified against the same reference standard and none were incorporated in the reference standard. The reference standard is unlikely to have led to bias in comparisons between tests.

Fourthly, all tests had the same interval between index test and reference standard (with the exception of clinical judgement). Because dementia is a progressive disorder, it is plausible that over time the clinical scenario may have been more obvious (either normal, or cognitive impairment) with a greater interval between clinical judgement and reference standard. However, as reported in Chapter 4 Section 4.2 there was no evidence of an association between time from referral and cognitive status. Undergoing multiple index tests may have been burdensome for participants and towards the end of the index battery people may have made errors due to cognitive fatigue, but the randomisation of the index tests should mean that no one test is systematically advantaged or disadvantaged in the comparison. Missing data was due to participants declining or change to the case report form as described previously and is unlikely to have led to systematic bias. Overall participant flow is unlikely to have led to bias in comparisons between tests

Fifthly, because of the uncertainty around the point estimates of each single test and combination of tests (both GP combined and GP 360) it is entirely possible that apparent differences in test accuracy are due to chance (type 1 error). This is especially important because confidence intervals provided are not adjusted for multiple comparisons (i.e. they are 95% CI rather than a 99% CI or Bayesian-derived credible interval) and in view of the multiple comparisons and combinations this may underestimate the impact of chance findings falsely implying superior test accuracy. In particular, the uncertainty in the estimates indicated by the confidence intervals mean that some tests which appear at face value to have superior diagnostic accuracy than others may in fact have very little diagnostic value. Consider Table 4.11, Sniffin sticks for the target condition normal has a point estimate of LRP 11 compared to MOCA LRP 8, but the 95% CI for Sniffin sticks is 1 to 101, indicating a high level of uncertainty and including 1, whereas for MOCA the 95% CI is 3 to 23, indicating stronger evidence that MOCA has some diagnostic value but (because of the overlapping 95% CIs) no real evidence to support use of one test rather than another.

Finally, the power of the sample size to make comparisons between different candidate index tests is low. While the sample size calculation was based on published tables [354], methodological advances since the design of the empirical test accuracy study have shown that conditional dependence between comparator index tests requires a larger sample size than would be calculated using less robust approaches [386] and that the traditional 10 events per variable rule [355] is misleading and should be revised because sample size calculation is context specific [387].

Key matters relating to comparisons of tests In summary, while randomisation of test order was used to partially mitigate the effect of a single examiner conducting all of the candidate index tests, the small sample size, overlapping confidence intervals, and multiple comparisons mean that there

is no real evidence to support the hypothesis of any one test being clearly superior to others and the results should be interpreted with caution. The uncertainty due to sample size is important for single and combined tests, but is especially important in considering both the GP combined and GP 360 tests.

6.2.7 Heterogeneity

Heterogeneity is described in Chapter 4 Section 4.7, and summarised in Table 6.2.

Age There was some evidence that for the target condition dementia GP combined tests had higher sensitivity and lower specificity in people aged ≥ 80 years than those aged < 80 years, with a reversed pattern for the target condition normal. This pattern of heterogeneity could reflect increasing sensory impairment with age. People with visual or hearing impairment, which may be not even be serious enough to have been noted by the patient, kin, or clinician, would be more likely to give incorrect answers to questions which utilise these faculties. For example, TAC asks the candidate to make change with coins, and other tests require drawing, or recognition of objects, all of which would be very difficult for visually impaired people regardless of cognitive ability. Similarly, tests which utilise delayed recall of an item which has been registered aurally would be difficult for a person with hearing impairment, regardless of cognitive functioning. The idea of tests being less specific with age due to sensory impairment is supported by the tests which were particularly affected: TAC, SPMT, MOCA and Phototest all require sensory skills. Other studies have reported influences on diagnostic accuracy of GP judgement for dementia. A false positive GP diagnosis of dementia is more likely with increasing age, impaired mobility or hearing, subjective memory impairment, or GP diagnosed depression [271]. Furthermore, less familiarity with the patient was associated with a higher probability of being rated as impaired in one study [201] but a better recognition of dementia in another study [270].

Heterogeneity by age also particularly affected FAQ but this is an informant completed test, so is not explained by increasing sensory impairment with age. However, this heterogeneity could reflect difficulty doing activities independently which is due to physical limitations rather than cognitive problems. Of the items on FAQ, at least three (shopping alone, travelling, working on a hobby) would be made more difficult by physical impairments

Table 6.2: Summary of heterogeneity

Characteristic	**Sensitivity**	**Specificity**
Target condition dementia		
Age ≥ 80 years ∗	higher	lower
Female ∗∗	higher	similar
Target condition normal		
Age ≥ 80 years ∗	lower	higher
Female ∗∗	lower	similar

∗ compared to people aged < 80 years
∗∗ compared to men

Sex There was some evidence that for the target condition dementia tests had higher sensitivity and similar specificity for female participants than male participants, with a reversed pattern for the target condition normal. Other studies have found that GPs label men as having dementia more often than women [265, 273]. For tests which measure change in the performance of activities of daily living, this may reflect a tendency for some of these items (such as doing the laundry, or cooking a meal) to be things which have traditionally been a "female role" in some households, and so the loss of the skill is more remarkable (and therefore more readily noted) in a woman than a man. Other possible explanations are that women had more severe cognitive impairment than men, but mean ACE3 scores were similar. Another possible explanation is that women (who are in general more likely to live alone) were more likely than men to attend with a grown-up child as an informant, who scored them more critically than a spouse would have done, but this cannot be evaluated with the available data. A final possibility is that despite similar ACE3 scores female participants were more frail, because frailty is a significant predictor of dementia and frail women may have a higher risk of developing incident dementia than men [388].

Setting As detailed in Section 3.2.2, referring GP surgeries were able to participate in the research study, as well as a funded programme of activity to encourage the diagnosis of dementia in general practice. There were two separate programmes of funded activity in general practice: first a scheme to identify people with cognitive impairment that was available to all practices in England, including referring practices (case finding); secondly a specific funded programme to improve the diagnosis and management of dementia in general practice which was restricted to Bristol practices (diagnosis).

It is likely that the net effect of the programmes was to encourage uptake of the study by referring GPs. On the other hand, it is possible that in the context of *some* of the GPs in Bristol referring practices (it is important to note, not all GPs) having received additional training on dementia and how to make a diagnosis some GPs may have been less likely to refer people to the study than would normally be the case, because they perceived themselves as being sufficiently capable to manage the situation themselves. It is difficult to quantify the effect of this, if there is any. Of the 240 people who were evaluated in the research clinic, 68 were referred by Bristol practices, and 86 by each of North Somerset and South Gloucestershire practices; compared to an total population aged 70 years and over of 10304 in Bristol, 14122 in North Somerset and 10530 in South Gloucestershire. However, differing start dates for practices to recruit (Table 3.2) makes it difficult to compare recruitment between practices and areas. A further possible impact of the NHS programmes of activity is that they contributed to the relatively high recruitment of men in this study in comparison to other studies (which is discussed in Section 6.2.2), because 53% of people with diabetes are men [389] and under the case-finding programme people with diabetes would have typically been asked about their memory at their annual diabetes review.

The NHS diagnosis programme of activity, which was restricted to Bristol, may also have had an impact on the diagnostic accuracy of GP judgement. It is important to note that typically training on the diagnosis and management of dementia for the programme would have been given to some,

rather than all, of the GPs in a practice. However, it is conceivable that diagnostic accuracy of GP judgement could be improved by a training programme which aimed to enhance the skills of GPs. Diagnostic accuracy of GP judgement for dementia in South Gloucestershire and North Somerset practices was lower than in Bristol practices (Section 4.7.5). In contrast the diagnostic accuracy for normal in South Gloucestershire and North Somerset practices was very slightly better than in Bristol practices. At face value, this would support the idea that the additional training that Bristol GPs received enhanced their diagnostic accuracy for a positive diagnosis of dementia but impaired their accuracy for diagnosing normal. An alternative possibility is that Bristol GPs were being more selective in only referring people who they thought had dementia, for confirmation of the diagnosis, to gain payment under the NHS scheme; this is plausible but a similar proportion of people from Bristol practices had dementia according to the reference standard compared to South Gloucestershire and North Somerset practices (Section 4.7.5). A further possibility, especially given the overlapping confidence intervals is that this is a chance finding, because changing behaviours and practice of GPs regarding dementia is very difficult: other tailored educational packages for GPs did not improve case identification of dementia even when there was reimbursement under the Quality and Outcomes Framework (see Section 3.2.2) [390].

A limitation is that because the systematic review and the empirical work were done simultaneously rather than in sequence it was not possible to use the empirical work to address the area for future work that were identified in 2.5.4.

6.3 Interpretation

6.3.1 GP judgement

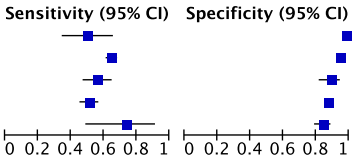
The finding from Chapter 4 of this thesis is that a positive GP judgement of dementia is often correct: PPV 86% LRP 5, whereas a positive GP judgement of normal is less often correct: PPV 56% LRP 5; the difference in PPV with a constant LRP is partly attributable to the prevalence of dementia being higher than the prevalence of normal. These results fit well with the findings of the systematic review described in Chapter 2 which were that GP judgement for dementia had a PPV of 61% LRP 11 whereas GP judgement for normal had a PPV of 26% LRP 2. This means that GP judgement is better at ruling in dementia, than at ruling in normal.

Dementia can be a complex clinical presentation, with physical symptoms and signs which can be challenging for the patient to process and explain, as well as affective, psychological, and cognitive symptoms, and of course an important impact on on social and role functioning. As discussed in Section 1.1.3.3 dementia is a low incidence condition in general practice, with potentially serious consequences, and this is likely to lead to a higher treatment threshold than for a more trivial condition like a sore throat, which is more frequently encountered. Figure 6.1 summarises the sensitivity and specificity of GP judgement for different conditions and shows that the accuracy for a diagnosis of dementia is comparable to other conditions, with a similar sensitivity and specificity

to a GP judgement of depression. Compared to dementia, depression can also have a variety of presentations, is also predominantly a mental health condition, but is more commonly encountered than dementia. For the diagnosis of depression GP judgement had LRP for GP judgement of 4 and LRN of 0.6. Pneumonia is an acute respiratory infection which is fairly common in general practice and often treated successfully by a course of antibiotics, and so with a relatively low treatment threshold (see Section 1.1.2.1) in general practice. For the target condition pneumonia as defined by chest radiograph, the index test GP judgement, based on history and examination alone, has LRP 5 (95% CI 3 to 7) LRN 0.3 (95% CI 0.2 to 0.7) [391]. For acute coronary syndrome, GP judgement had LRP 25 LRN 0.5 [392]. These findings indicate that GP judgement for dementia is comparable to the

Figure 6.1: Comparison of diagnostic accuracy of GP judgement for different target conditions

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Acute Coronary Syndrome	22	23	22	1145	0.50 [0.35, 0.65]	0.98 [0.97, 0.99]
Dementia (systematic review)	494	285	271	4797	0.65 [0.61, 0.68]	0.94 [0.94, 0.95]
Dementia (this study)	74	12	58	96	0.56 [0.47, 0.65]	0.89 [0.81, 0.94]
Depression	159	162	153	1084	0.51 [0.45, 0.57]	0.87 [0.85, 0.89]
Pneumonia	14	37	5	194	0.74 [0.49, 0.91]	0.84 [0.79, 0.88]



The Forest plot shows the sensitivity and specificity of GP judgement for different conditions, Depression [393] Pneumonia [391] Acute Coronary Syndrome [392] **TP** true positive **FP** false positive **FN** false negative **TN** true negative **95% CI** 95% confidence interval

accuracy of clinicians for other similar conditions, and suggest that GP judgement may be higher for more frequently encountered (i.e. more common) conditions with a more typical clinical phenotype, such as pneumonia, than for rarer conditions with a more varied presentation, such as dementia or depression.

6.3.2 What do we mean by dementia?

As described in Section 1.1.2 diagnosis can be conceptualised as a method of classification that is fit for purpose in the clinical setting, and therefore GP judgement may often use heuristics and system one cognitive processes. In this study, there was a signal that specificity of GP combined tests is higher in younger people than older people (Section 6.2.7) which could reflect increasing sensory impairment with age; making a false positive GP diagnosis of dementia more likely with increasing age, impaired mobility, GP diagnosed depression and subjective memory impairment [271] (see Section 6.2.7).

One lens to interpreting these findings is that the GP diagnosis paradigm for dementia and cognitive impairment is different to formal definitions, and that just as ICD-10 and DSM-IV-TR definitions of dementia identify different groups of people (Section 1.1.1.1), it is possible that the GP heuristic for dementia is selecting a group of people who are systematically different to those who are identified as having dementia by formal definitions. In particular 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. In the

possible relationship between GP judgement and ICD-10 dementia (see also Section 4.4.1 Table 4.6) some people with ICD-10 dementia are in the set of people with GP CIND, others are in the set with GP dementia and still others are not identified by the GP as being impaired.

If diagnosis is conceptualised as a classification procedure (Section 1.1.2) with importance for patient outcomes [175] and models of illness [174] (Section 1.1.2.8), then it may be that the GP paradigm of dementia as described above better serves the needs of patients in that setting regarding appropriate goals of care, support, and advance care planning than more formal definitions. An important limitation of this GP paradigm of dementia is that it is a less clearly defined disease for research, with a less certain disease course and for which the benefit of interventions is uncertain.

Further research using qualitative methods such as ethnography could explore further what factors influence a GP heuristic of dementia and cognitive impairment (see also Section 6.4.1). When the GP heuristic is better understood, quantitative methods could then explore whether the GP paradigm is more valuable as a prognostic definition of disease [175] in a primary care population than more formal definition of dementia.

6.3.3 Use of tests in practice

This thesis has demonstrated that a combination of tests including GP judgement can have high diagnostic accuracy for the target conditions dementia and normal in a group of people who are presenting with cognitive symptoms. How might these findings be applied? Often clinical practice guidelines are used to inform clinical practice. A review of guidelines relating to BCAs (brief cognitive assessments) for the diagnosis of dementia found four relevant guidelines, two in Canada, one in Australia and one in the United Kingdom [308]. The United Kingdom guidelines did not recommend any specific BCA but did indicate that at the initial assessment a history should be taken, including from an informant if possible, and then appropriate investigations should be done along with cognitive testing [377]. Of the tests in the GP 360 tests that were identified in this thesis, FAQ, 6CIT and IQCODE are all listed in the United Kingdom guideline, but Eurotest is not [377]. Instead, the United Kingdom guidelines include the six-item screener (which asks year, month, day, and a three item recall) [394], Test your memory [395], Mini-cog; and two tests which are described in Section 3.4.1: Memory impairment screen [331] and 10-point cognitive screener [332]. Test your memory is scored out of 50 and is designed to be self-completed by the patient, it tests a range of domains including memory, orientation, verbal fluency, abstraction, visual coordination, recall, and clock-draw; it is advised that the a person completing the test is supervised by a clinician. However, there are many validated BCA and as illustrated in the review of clinical guidelines, there is very little consensus about what test should be used: the Australian guidelines recommend ten further tests, with no overlap with the United Kingdom guideline [308].

If there is little consensus from guidelines about what test to use in practice, then there is also little consensus amongst clinicians. In a survey of 52 GPs, who indicated doing a mean four cognitive assessments per month (range 0.24 - 30) the most commonly used tests were the MMSE and GPCOG,

other tests that were used included clock draw, Minicog, ACE3, IQCODE and MOCA [308]. Factors that were reported to influence test choice included: time (35%), local factors (21%), familiarity (19%), patient factors (10%), other (9%), guidelines (6%) and cost (1%) [308]. Regarding time, 74% indicated that an acceptable test time if a GP were doing the test was less than five minutes whereas if a non-GP clinician were doing the test it would be acceptable to spend longer completing the test [308].

One problem for applying the findings of the thesis in practice is that health professionals struggle to understand commonly used measures of test accuracy [396]. Similarly, systematic reviews of test accuracy can also be reported to be difficult to understand, in part due to problems with background knowledge, but also due to problems with layout and presentation [397]. In both cases, the use of natural frequencies was helpful in improving understanding [396, 397]. A further problem for both clinicians and policy makers to date has been that very few studies have made direct comparisons between BCA and evaluated their accuracy in a clinical setting that is applicable to general practitioners: that is, people presenting with symptoms to their GP [308]. Therefore, the ability to compare accuracy between BCA, and GP combined BCA in a symptomatic general practice population is a major strength and novel finding of the thesis.

Even with a perfect test, it might still be difficult to make a diagnosis of dementia in general practice. Barriers to diagnosis of dementia by GPs have been reported to fit within four main categories: organisational, clinician-related, patient-related, and societal [398]. Organisation factors included (among other things) time constraints and doubts about using BCA to diagnose dementia rather than to highlight important symptoms that required a specialist opinion, especially in the context of performance being subject to cultural or language factors [398]. Other potential barriers to the diagnosis of dementia have been identified as being lack of support, stigma, diagnostic uncertainty, and disclosing the diagnosis [399]. Because of resource constraints people with severe symptoms have been reported to be prioritised over people with more mild problems [398]. On the other hand, continuity has been reported to facilitate a GP recognising cognitive problems [398]. Concerns about resources and lack of specialist expertise for GPs to diagnose and manage dementia well have been reported for many years [400, 401], and GPs have been reported to frame dementia care as a specialist activity [400]. Despite this, the priorities of patients and their kin are to get a prompt diagnosis, in an emotionally safe setting, in a personalised way [402], and patients report that they are generally satisfied with either primary or secondary care approaches to diagnosis [340]. General practice based approaches to diagnosis have been used effectively in the United Kingdom [340, 403] and Canada [404], and follow-up of people with dementia can be done as effectively by general practice as by specialist [405]. However, regrettably, a well designed intervention to improve practice was not effective in improving documentation or increasing case identification, and it may be that rather than training up GPs, practices need to be provided with additional trained human resource [390] to facilitate the diagnosis and management of dementia in general practice; in England this could be through Primary Care Networks [406].

6.4 Implications

6.4.1 For research

This thesis has made an important contribution to the literature. The especially novel findings are the data on the proportion of people presenting with cognitive symptoms who have dementia MCI and normal cognition, and the comparative accuracy of a range of index tests including GP judgement for diagnosing two target conditions. A further novel contribution has been the distinction of two target conditions, normal and dementia, rather than just one (dementia). The thesis has not considered cost effectiveness, specific dementia aetiologies, or evaluated the combinations of tests prospectively. An important limitation is that the sample size limits the conclusions that can be reached about the comparative accuracy of tests.

One future priority for research should be to determine the the proportion of people who have a rare or atypical dementia out of those with cognitive symptoms, and the clinical features which are markers for this. This is because these people would need specialist evaluation in any diagnostic pathway and knowing the proportion of people aged over 70 years who would require specialist evaluation is an important part of designing a diagnostic pathway.

A second area for future research is in people who are judged by their GP to have CIND. MCI was not a target condition, but half of participants with cognitive symptoms were classified by referring GPs as having CIND. Future research should explore what contributes to GPs making a diagnosis of CIND rather than normal or dementia, perhaps using qualitative or ethnographic techniques.

Thirdly, it would be useful to identify the most discriminative tests to distinguish dementia and normal from MCI in people who are judged by a GP as being cognitively impaired but not having dementia. Because of the difficulty in distinguishing borderline cases, this group of people is likely to require more detailed testing than can be done with a BCA, such as the ACE3 and perhaps also bio-markers [407]. Electronic approaches to cognitive evaluation [408] may be especially informative for people in this group. Finally, Chapter 5 describes a brief research agenda for further approaches to investigate the experience and views of people being evaluated for cognitive symptoms.

These three further investigations should inform subsequent research to approach the design of a diagnostic pathway through the lens of guidance to develop complex interventions [409, 410]. It is likely to be valuable to consider the pathway in the context of a proposed theoretical framework to determine whether a trial is necessary [411] and consider whether randomisation should be at the level of the individual or in clusters (e.g. practices, primary care networks, or even Clinical Commissioning Groups). Evaluating all of the combinations of GP combined and GP 360 tests in a prospective diagnostic trial incorporating cost effectiveness, and powered to determine heterogeneity by aetiology, is likely to require a very large sample size and therefore be very expensive. However, given the workload associated with dementia diagnosis in the western world it may be still be cost-effective for health services. An alternative approach would be to use the three proposed initial investigations, together with the qualitative research agenda and the integrated development of

a complex intervention to select a number of (perhaps four) candidate GP combined and GP 360 tests that could be evaluated in a large number of people. Possible outcomes for a future trial could include numbers diagnosed, cost, and time to diagnosis.

6.4.2 For practice

Based on the qualitative results in Chapter 5, when a GP encounters a patient who has consulted about possible dementia the GP should aim to focus the initial consultation on identifying the patients specific concerns, quantify the impact on daily life, determine safety (e.g. from the stove or driving) and ask about red flags for rare or atypical illness. At the end of the initial encounter the GP should aim to make an initial determination as to whether the person has dementia, or is normal. People with rare or atypical presentations may need referral to a specialist who could provide a bespoke assessment to address the problems they are facing, rather than a more generic evaluation for the relatively limited range of clinical exemplars that are known to the GP.

Regarding the diagnostic process, based on the quantitative results and the qualitative interviews, the initial clinical judgement from the GP should guide further tests to ensure that the tests evaluate the symptoms that are reported by the patient. Based on the qualitative findings, these could be conducted by a trained health worker such as a health care assistant. Based on the qualitative findings, the second encounter could also offer advice about lifestyle measures to promote cognition, and identifying any carers who might also need support. Further evaluation with formal measures of mood, and investigations to exclude metabolic, endocrine or neoplastic disorders should also be considered, as part of good practice and in line with guidelines [377]. The patient should then be reviewed by a GP together with the results of tests.

A multidisciplinary approach is likely to be cheaper than a fully GP delivered model, but there are three additional advantages that based on Chapter 5 would help to improve the acceptability of diagnosis of dementia in general practice. Firstly, having formal cognitive testing done by an allied professional would allow the GP time to explore the specific concerns of the patient and their family and meet their needs. Secondly, within appropriate boundaries both professionals could provide advice on interventions; the allied professional could advise on self care, and the GP could advise on appropriate medication as the cost effectiveness of this has evolved due to lower acquisition costs [412]. Thirdly a multi-disciplinary approach could facilitate the integration of community services to providing holistic and integrated care for people with dementia and their kin after a diagnosis and would fit within the framework of the NHS Long Term plan [413]. Finally the multi-disciplinary approach could be readily quality assured by having regular certificated training for the allied professional (perhaps relating to testing and self care) and the GP (perhaps relating to prescribing). A limitation of the work for practice is that prescribed pharmacological interventions for dementia are specifically licensed in the United Kingdom only for Alzheimer's disease, rather than all-cause dementia. Drugs are recommended for off-license use in Dementia with Lewy Bodies and Frontotemporal dementia by the National Institute of Clinical Excellence [377].

6.4.3 Accuracy as prevalence varies

A positive GP judgement for dementia is sufficiently accurate that if FAQ and Eurotest indicate no dementia (as in the GP 360 test described in Section 4.6.4.2), the probability of dementia is still 63%; though the uncertainty in this prediction limits the conclusions that can be reached at this stage. Conversely, if a GP thinks a patient is normal and the 6CIT and FAQ indicate that cognition is not normal, the probability of normal cognition is only 5%. Because, as described in 1.1.2.3 predictive values are dependent on prevalence the high prevalence of dementia in this study inflates the PPV. If the prevalence was lower, such as 30% of people having dementia [414] then the diagnostic accuracy of GP + Eurotest + FAQ would be LRP 14 LRN 0.5 PPV 85% (95% CI 69% to 94%) i.e. the PPV would be lower than reported in the thesis. The prevalence of normal cognition was not reported in [414] but if it were 30% instead of 19% as in the thesis then the diagnostic accuracy of GP + 6CIT + IQCODE would be LRP 59 LRN 0.7 PPV 96% (95% CI 77% to 100%) i.e. the PPV would be higher than reported in the thesis.

6.5 Conclusions

To return to the aims and objectives of the thesis: The aim of this thesis was to investigate approaches to diagnosing dementia syndrome in general practice. The objectives have been addressed:

- In a group of people presenting with symptoms of dementia to their general practitioner, the prevalence of dementia was $\frac{131}{240}$, and the prevalence of MCI was $\frac{59}{240}$.
- The accuracy of a range of tests, including GP judgement, for diagnosing dementia syndrome in symptomatic people was investigated. LRP of single tests ranged from 1 (many tests) to 10 (FAQ).
- Combination of tests with good discrimination, calibration, and high LRP for identifying people with dementia and normal cognition were identified. GP judgement + Eurotest + FAQ had the largest net benefit for diagnosing dementia and GP + 6CIT + IQCODE had the largest net benefit for confirming normal cognition. Uncertainty in the estimates and small sample size restricted the ability to draw any firm conclusions about comparative test accuracy especially for the combinations of tests.
- Based on interviews with patients and their kin it was judged that it would be acceptable for GPs to diagnose dementia if clinicians were adequately trained and resourced, and if the service fully addressed the needs of the patient for information and interventions, not just a diagnosis.

Diagnosis of dementia is possible in primary care. Implementing this is likely to be a challenge, and may require additional human resources, not just training of existing staff.

"Men must endure
Their going hence, even as their coming hither " *Edgar, Act 5 Scene 2, King Lear*

APPENDIX A

SEARCH REPORT

This appendix includes the search report for the systematic review.

Search Report
Time frame: all dates (to present: April 2019)

Contact Person: Sam Creavin

Searches by: AN-S
First assess by:
Search results sent:

Results: TOTAL: April 2016: 10154 // April 2019:
TOTAL after de-dupe: April 2019: 7060 // April 2019: 1058

Source	Version/Platform/url	Date of Search	Filter applied as overall limiter	Records retrieved
1. MEDLINE In-process and other non-indexed citations and MEDLINE 1946-present	(Ovid SP)	26/04/16 29/04/19	No	4493 927
2. EMBASE 1974-2016 April 25	(Ovid SP)	26/04/16 29/04/19	No	2730 1010
3. PSYCINFO 1806-April week 3 2016	(Ovid SP)	26/04/16 29/04/19	No	1970 367
4. Web of Science Core Collection	ISI Web of Science	26/04/16 29/04/19	Yes	928 210
6. LILACS	BIREME	26/04/16 29/04/19	No	33 13
TOTAL before de-duplication				10154 2527
TOTAL after de-duplication				7060 1058

Search Strategies:

Copy and paste into search appendix of RevMan file

Source	Search strategy	Hits retrieved
1. MEDLINE In-process and other non-indexed citations and MEDLINE 1946-present (Ovid SP)	1. exp "sensitivity and specificity"/ 2. "reproducibility of results"/ 3. diagnos*.ti. 4. di.fs. 5. sensitivit*.ab. 6. specificit*.ab. 7. (ROC or "receiver operat*").ab. 8. Area under curve/ 9. ("Area under curve" or AUC).ab. 10. sROC.ab. 11. accura*.ti,ab. 12. (likelihood adj3 (ratio* or function*)).ab. 13. ((true or false) adj3 (positive* or negative*)).ab.	Apr 2016: 4493 Apr 2019: 927

	14. ((positive* or negative* or false or true) adj3 rate*).ti,ab. 15. or/1-14 16. exp Dementia/ 17. Delirium, Dementia, Amnestic, Cognitive Disorders/ 18. dement*.mp. 19. alzheimer*.mp. 20. (lewy* adj2 bod*).mp. 21. (chronic adj2 cerebrovascular).mp. 22. ("organic brain disease" or "organic brain syndrome").mp. 23. ("normal pressure hydrocephalus" and "shunt*").mp. 24. "benign senescent forgetfulness".mp. 25. (cerebr* adj2 deteriorat*).mp. 26. (cerebral* adj2 insufficient*).mp. 27. (pick* adj2 disease).mp. 28. (creutzfeldt or jcd or cjd).mp. 29. huntington*.mp. 30. binswanger*.mp. 31. korsako*.mp. 32. "cognit* impair*".mp. 33. exp *Cognition Disorders/ 34. MCI.ti,ab. 35. ACMI.ti,ab. 36. ARCD.ti,ab. 37. SMC.ti,ab. 38. CIND.ti,ab. 39. BSF.ti,ab. 40. AAMI.ti,ab. 41. MD.ti,ab. 42. LCD.ti,ab. 43. QD.ti,ab. 44. AACD.ti,ab. 45. MNCD.ti,ab. 46. MCD.ti,ab. 47. ("N-MCI" or "A-MCI" or "M-MCI").ti,ab. 48. ((cognit* or memory or cerebr* or mental*) adj3 (declin* or impair* or los* or deteriorat* or degenerat* or complain* or disturb* or disorder*)).ti,ab. 49. "preclinical AD".mp. 50. "pre-clinical AD".mp. 51. ("preclinical alzheimer*" or "pre-clinical alzheimer*").mp. 52. (aMCI or MCIa).ti,ab. 53. ("CDR 0.5" or "clinical dementia rating scale 0.5").ti,ab. 54. ("GDS 3" or "stage 3 GDS").ti,ab. 55. ("global deterioration scale" and "stage 3").mp. 56. "Benign senescent forgetfulness".ti,ab. 57. "mild neurocognit* disorder*".ti,ab. 58. (prodrom* adj2 dement*).ti,ab. 59. (episodic* adj2 memory).mp. 60. ("preclinical dementia" or "pre-clinical dementia").mp. 61. or/16-60 62. Family Practice/ or Ambulatory Care/ 63. Physicians, Family/ or Physicians, Primary Care/ 64. Primary Health Care/ 65. "family practice".ti,ab. 66. "general practi*".ti,ab. 67. *General Practice/ 68. "family practices".ti,ab. 69. "family practitioner*".ti,ab. 70. "general practitioner*".ti,ab. 71. "primary care".ti,ab. 72. Physician Assistants/ 73. "physician assistant*".ti,ab. 74. Nurse Practitioners/ 75. "nurse practitioner*".ti,ab. 76. or/62-75 77. 61 and 76	

	78. 15 and 77 79. "clinical judgement*".ti,ab. 80. "practitioner* judgement*".ti,ab. 81. ((clinician* or GP* or physician* or doctor*) adj3 (intuit* or recognis* or detect* or diagnos*)).ti,ab. 82. "gut feeling*".ti,ab. 83. gestalt.ti,ab. 84. "GP judgement*".ti,ab. 85. ((clinician* or GP* or physician* or doctor*) adj3 accura*).ti,ab. 86. *Practice Patterns, Physicians'/ 87. or/79-86 88. 61 and 87 89. 78 or 88	
2. EMBASE 1974-2016 April 25 (Ovid SP)	1. *diagnostic accuracy/ 2. reproducibility/ 3. diagnos*.ti. 4. sensitivit*.ab. 5. specificit*.ab. 6. (ROC or "receiver operat*").ab. 7. area under the curve/ 8. ("Area under curve" or AUC).ab. 9. sROC.ab. 10. accura*.ti,ab. 11. (likelihood adj3 (ratio* or function*)).ab. 12. ((true or false) adj3 (positive* or negative*)).ab. 13. ((positive* or negative* or false or true) adj3 rate*).ti,ab. 14. "sensitivity and specificity"/ 15. or/1-14 16. exp dementia/ 17. dement*.ti,ab. 18. alzheimer*.ti,ab. 19. (lewy* adj2 bod*).ti,ab. 20. (chronic adj2 cerebrovascular).ti,ab. 21. ("organic brain disease" or "organic brain syndrome").ti,ab. 22. ("normal pressure hydrocephalus" and "shunt*").ti,ab. 23. "benign senescent forgetfulness".ti,ab. 24. (cerebr* adj2 deteriorat*).ti,ab. 25. (cerebral* adj2 insufficient*).ti,ab. 26. (pick* adj2 disease).ti,ab. 27. (creutzfeldt or jcd or cid).ti,ab. 28. "cognit* impair*".ti,ab. 29. huntington*.ti,ab. 30. binswanger*.ti,ab. 31. korsako*.ti,ab. 32. MCI.ti,ab. 33. ACMI.ti,ab. 34. ARCD.ti,ab. 35. SMC.ti,ab. 36. CIND.ti,ab. 37. BSF.ti,ab. 38. AAMI.ti,ab. 39. LCD.ti,ab. 40. QD.ti,ab. 41. AACD.ti,ab. 42. MNCD.ti,ab. 43. MCD.ti,ab. 44. ("N-MCI" or "A-MCI" or "M-MCI").ti,ab. 45. ((cognit* or memory or cerebr* or mental*) adj3 (declin* or impair* or los* or deteriorat* or degenerat* or complain* or disturb* or disorder*)).ti,ab. 46. "preclinical AD".ti,ab. 47. "pre-clinical AD".ti,ab. 48. ("preclinical alzheimer*" or "pre-clinical alzheimer*").ti,ab.	Apr 2016: 2730 Apr 2019: 1010

	49. (aMCI or MCIa).ti,ab. 50. ("CDR 0.5" or "clinical dementia rating scale 0.5").ti,ab. 51. ("GDS 3" or "stage 3 GDS").ti,ab. 52. ("global deterioration scale" and "stage 3").ti,ab. 53. "Benign senescent forgetfulness".ti,ab. 54. "mild neurocognit* disorder*".ti,ab. 55. (prodrom* adj2 dement*).ti,ab. 56. (episodic* adj2 memory).ti,ab. 57. ("preclinical dementia" or "pre-clinical dementia").ti,ab. 58. or/16-57 59. general practice/ 60. ambulatory care/ 61. primary medical care/ 62. "family practice".ti,ab. 63. "general practi*".ti,ab. 64. "family practice*".ti,ab. 65. "family practitioner*".ti,ab. 66. "general practitioner*".ti,ab. 67. "primary care".ti,ab. 68. physician assistant/ 69. "physician assistant*".ti,ab. 70. nurse practitioner/ 71. "nurse practitioner*".ti,ab. 72. or/59-71 73. 15 and 58 and 72 74. "clinical judgement*".ti,ab. 75. "practitioner* judgement*".ti,ab. 76. ((clinician* or GP* or physician* or doctor*) adj3 (intuit* or recognis* or detect* or diagnos*)).ti,ab. 77. "gut feeling*".ti,ab. 78. gestalt.ti,ab. 79. "GP judgement*".ti,ab. 80. ((clinician* or GP* or physician* or doctor*) adj3 accura*).ti,ab. 81. or/74-80 82. 58 and 81 83. 73 or 82	
3. PSYCINFO 1806-April week 3 2016 (Ovid SP)	1. diagnos*.ti. 2. sensitivit*.ab. 3. specificit*.ab. 4. (ROC or "receiver operat*").ab. 5. area under the curve/ 6. sROC.ab. 7. accura*.ti,ab. 8. (likelihood adj3 (ratio* or function*)).ab. 9. ((true or false) adj3 (positive* or negative*)).ab. 10. ((positive* or negative* or false or true) adj3 rate*).ti,ab. 11. "sensitivity and specificity"/ 12. exp Test Reliability/ or exp Diagnosis/ or exp Medical Diagnosis/ 13. or/1-12 14. exp DEMENTIA/ 15. dement*.ti,ab. 16. alzheimer*.ti,ab. 17. (lewy* adj2 bod*).ti,ab. 18. (chronic adj2 cerebrovascular).ti,ab. 19. ("organic brain disease" or "organic brain syndrome").ti,ab. 20. ("normal pressure hydrocephalus" and "shunt*").ti,ab. 21. "benign senescent forgetfulness".ti,ab. 22. (cerebr* adj2 deteriorat*).ti,ab. 23. (cerebral* adj2 insufficient*).ti,ab. 24. (pick* adj2 disease).ti,ab. 25. (creutzfeldt or jcd or cid).ti,ab. 26. huntington*.ti,ab. 27. binswanger*.ti,ab.	Apr 2016: 1970 Apr 2019: 367

	28. korsako*.ti,ab. 29. "cognit* impair*".ti,ab. 30. MCI.ti,ab. 31. ACMI.ti,ab. 32. ARCD.ti,ab. 33. SMC.ti,ab. 34. CIND.ti,ab. 35. BSF.ti,ab. 36. AAMI.ti,ab. 37. MD.ti,ab. 38. LCD.ti,ab. 39. QD.ti,ab. 40. AACD.ti,ab. 41. MNCD.ti,ab. 42. MCD.ti,ab. 43. ("N-MCI" or "A-MCI" or "M-MCI").ti,ab. 44. ((cognit* or memory or cerebr* or mental*) adj3 (declin* or impair* or los* or deteriorat* or degenerat* or complain* or disturb* or disorder*)).ti,ab. 45. "preclinical AD".ti,ab. 46. "pre-clinical AD".ti,ab. 47. ("preclinical alzheimer*" or "pre-clinical alzheimer*").ti,ab. 48. (aMCI or MCIa).ti,ab. 49. ("CDR 0.5" or "clinical dementia rating scale 0.5").ti,ab. 50. ("GDS 3" or "stage 3 GDS").ti,ab. 51. ("global deterioration scale" and "stage 3").ti,ab. 52. "Benign senescent forgetfulness".ti,ab. 53. "mild neurocognit* disorder*".ti,ab. 54. (prodrom* adj2 dement*).ti,ab. 55. (episodic* adj2 memory).ti,ab. 56. ("preclinical dementia" or "pre-clinical dementia").ti,ab. 57. or/14-56 58. exp Family Physicians/ or exp Primary Health Care/ 59. exp General Practitioners/ or exp Clinical Practice/ 60. "family practice".ti,ab. 61. "general practi*".ti,ab. 62. "family practices".ti,ab. 63. "family practitioner*".ti,ab. 64. "general practitioner*".ti,ab. 65. "primary care".ti,ab. 66. "physician assistant*".ti,ab. 67. "nurse practitioner*".ti,ab. 68. or/58-67 69. 13 and 57 and 68 70. "clinical judgement*".ti,ab. 71. "practitioner* judgement*".ti,ab. 72. ((clinician* or GP* or physician* or doctor*) adj3 (intuit* or recognis* or detect* or diagnos*)).ti,ab. 73. "gut feeling*".ti,ab. 74. gestalt.ti,ab. 75. "GP judgement*".ti,ab. 76. ((clinician* or GP* or physician* or doctor*) adj3 accura*).ti,ab. 77. or/70-76 78. 57 and 77 79. 69 or 78	
4. Web of Science core collection	TOPIC: (dement* OR alzheimer* OR "lewy bod*" OR DLB OR "vascular cognitive impairment*" OR FTD OF FTLD OR "cerebrovascular insufficienc*" OR "mild cognitive impairment" OR MCI) AND TOPIC: ("primary care" OR "general practi*" OR GP OR "doctor* surgery" OR "family practi*" OR "ambulatory care") AND TOPIC: (diagnosis OR sensitiv* OR specificit* OR ROC OR "receiver operat*" OR "Area under curve" or AUC OR sROC OR accura* OR "follow*-up" OR "positive predictive value*" OR "negative predictive value*" OR longitudinal OR longitudinally)	Apr 2016: 928 Apr 2019: 210

	AND TOPIC: (((GP OR practitioner* OR clinician*) AND (judgement* OR assessment* OR diagnosis)) OR "gut feeling*" OR gestalt) Timespan: All years. Search language=Auto	
6. LILACS (BIREME)	dementia OR demencia OR demência OR alzheimer OR alzheimers OR alzheimer's OR cognition OR "mild cognitive impairment" [Words] and "primary care" OR "general practice" OR "atención primaria" OR "Práctica general" OR "Prática geral" [Words]	Apr 2016: 33 Apr 2019: 13
TOTAL before de-duplication		Apr 2016: 10154 Apr 2019: 2527
TOTAL after de-dupe and first-assess		Apr 2016: 7060 Apr 2019: 1058

APPENDIX B

ADMINISTRATIVE SOP

This appendix includes the SOP that were used by the administrative team prior to the research clinic. A SOP was provided for the initial telephone call to the referred person.

STANDARD OPERATING PROCEDURE : INITIAL CALL TO PATIENTS

CHECK FOR MISSING DATA IN REFERRAL FORM

1. Review the referral form for missing data.
2. If there is no missing data, continue
3. **If there is missing data**
    a. Update Timeli referrals and practices spreadsheet on google drive | GP missing info tab
    b. Send letter Letter to GP re Missing info 19_06_2015 template to GP (saved in Studies/Timeli/GP Referral queries
    c. When sending email follow Standard Operating Procedure emails to surgeries TIMeLI 19_06_2015
    d. Body of email is saved in Text for email for missing information to GPs 25062015_AK

CHECK CONTACT DETAILS AVAILABLE

1. Check there are contact details, if not, email surgery and ask for a phone number. Try on three different days on at least two different parts of the day (e.g. morning, afternoon). If unable to make contact after one week from referral send SOP Ineligible or declined letter to surgery
2. If the surgery only has the number for a relative and has documented permission to permission / consent to discuss with that relative on the phone then discuss with Sam (who will contact relative)
    a. Usually this will mean discussing with the relative.
3. Document responses in study database call log and then if would like to participate transfer to contacts database.
4. If the patient is unable to use the telephone at all they may be ineligible, discuss with Sam Creavin in all cases.

INTRODUCTIONS

5. Introduce self – I am XX from the university of Bristol. May I speak to YY?
6. I'm calling about the timeli memory research study, I think you have seen Dr ZZ recently and they may have mentioned it to you. The research is looking at the best tests for GPs to use to diagnose memory problems.
7. Do you have five minutes or so for me to discuss it with you?
8. The first thing I need to do, if I may, is just check if you are suitable to take part in the study, is that alright? The first thing is:
9. Can I check your details? (date of birth, name, address)

INTRODUCE THE STUDY

1. The study is to try to find the best tests for people with memory problems, to see whether or not they have a memory problem.

[*If referral form from GP indicates that the patient does not perceive a memory problem remember that they might deny it.*]

CHECK ELIGIBILITY

[In general, assume eligibility (as GP should have checked). If appears to be not eligible, need to double check before deciding this is the case and discuss every case with Sam Creavin within 24 hours.]

2. I need to ask you some yes or no questions. Just to remind you, I'm a researcher at the university, not a doctor or nurse.
3. Had you noticed any problems with your memory?
    a. Thinking about those problems with your memory, how long have they been there for? At least six months or so?
    b. Are your symptoms relatively stable, not getting worse week-by-week?
    c. Since you've noticed the change, have you noticed any new nerve problems (change in speech, tremor, weakness of hands or legs)?
    d. If patient says no problems with memory see separate paragraph below
4. There are a few medical problems that would mean you can't take part, I just need to check whether you've got any of those or not.
    a. Has your doctor ever told you you've got:
        i. Parkinsons disease
        ii. Multiple sclerosis
        iii. Learning Difficulty
        iv. Huntingdon Disease
        v. Sensory impairment (blind, profound deafness)
        vi. Motor Neurone Disease
        vii. Under care of a psychiatrist
5. And just to be sure, you haven't already been given a diagnosis of dementia or memory problems from a specialist in the past? **An existing diagnosis of dementia is an exclusion criteria.**
    a. Also check referral letter from GP
6. Do not post information out offer participation or an appointment date until eligibility confirmed

OFFER PARTICIPATION IF ELIGIBLE

7. We obviously can't say for sure if you have memory problems or not but it sounds as if you would be suitable to take part in the study. If you wanted to take part we would see you at a research clinic in either Hanham or Shirehampton. We can arrange transport for you and you will spend 2 hours at the clinic and see two different doctors. One is a GP in his final year of training and the other is a memory specialist doctor.
8. You would need to be able to come with someone that knows you well, will that be possible? (If no – see "IF APPEARS TO BE NOT ELIGIBLE OR DECLINES")

9. Would you like for me to book an appointment?
10. If Yes
    a. I'll put some information in the post for your today, there'll be a consent form and information sheet, but you don't need to sign them until you are seen at the research clinic.
    b. So I've booked the appointment provisionally for AAAA. We won't be able to give you a diagnosis on the day but we will send a letter back to Dr ZZZ. If anything worries us, like your care or if we are worried about you driving home, we will tell you on the day.
        i. If unable to make an appointment within three weeks see Booking Clinics SOP | CHANGING ASSIGNED CLINIC
    c. Would you like me to arrange transport for you?
        i. If so, arrange transport with Bristol Community Transport – see TIMeLi transport booking email version 1.0 2015_03_31.
        ii. Also seek permission to obtain quotes from Taxi company as in some cases this is better.
            1. Email traveldesk@v-cars.com to get a quote – without giving any identifiable information just postcode to postcode and time / date.
            2. If need wheelchair then BCT needs to be used
            3. If using Taxi Sam Creavin will need to pay driver directly on the day in cash.
    d. Shall I give you the study mobile phone number, 07773 472 622? Or the study email address?
    e. Arrange a time to call to confirm everything is OK, 2-7 days before appointment.
        i. Book time on timeli calendar and add task to Trello but not using identifiable information
    f. Complete SOP letter in pack 13_08_2015 v1 and enclose in envelope.
    g. Is English your first language?
        i. If YES Would you like me to book an interpreter? [we will pay]
            1. If No – confirm, and document on ACCESS
        ii. Discuss with sam Creavin in all cases. Appointment will need 3 hours rather than 2. Need to reduce slots per clinic accordingly.
11. If No – see "IF APPEARS TO BE NOT ELIGIBLE OR DECLINES"

IF APPEARS TO BE NOT ELIGIBLE OR DECLINES

[Only dispose someone as ineligible if you can confirm that they do not meet the inclusion criteria or do meet exclusion criteria with a **second** person (e.g. GP referral letter, relative, husband); you will need their permission to discuss with someone else.

If appears to be not eligible, need to double check before deciding this is the case and discuss every case with Sam Creavin within 24 hours.

If there is no possibility of speaking to a second person then send GP letter SOP Ineligible or declined letter to surgery.

If you are unsure, go to "If not sure" and obtain permission for Sam Creavin to call to discuss.]

1. If person has stated either "no memory problem or duration less than six months"
    a. Ask if there is anyone else that you can talk to there at the moment, e.g. a wife friend or other relative, obtain permission to discuss matter with them, and then recheck eligibility.
    b. If no one available at present see if can call back another time and book this in
    c. If not possible to talk to an informant on the phone, thank patient for their time and send letter to GP SOP Ineligible or declined letter to surgery
2. If person has stated "no informant available"
    a. Check and confirm no-one available
    b. Send letter to GP SOP Ineligible or declined letter to surgery
3. If person has declined
    a. Thank them for their time
    b. Confirm their decision, would they like posted materials first and then an opportunity to discuss again?
    c. Ask, I'm not trying to change your mind, but if you feel able, could you say why you felt it wasn't right for you, just so we can try to improve things for next time
        i. Document response
    d. Send letter to GP SOP Ineligible or declined letter to surgery
4. If person wants more time to decide
    a. Book time on timeli calendar and add task to Trello but not using identifiable information
5. If person has stated "already been diagnosed with dementia by a specialist"
    a. Check letter from GP to see if the diagnosis is on the record – it should be.
        i. If not on medical record, obtain permission to discuss with GP, document this on Access, and contact Sam Creavin.
        ii. An email will need to be sent to the GP to confirm the situation.
    b. If diagnosis of dementia is confirmed on medical record then patient is not eligible. Explain this to patient
    c. Send letter to GP SOP Ineligible or declined letter to surgery {exclusion reason 2 – give details}

IF NOT SURE

1. Would it be OK for me to arrange a time for Dr Creavin, the study lead, to give you a call to discuss the study with you?
    a. If Yes – Email Sam Creavin and add task on Trello but do not give identifiable information – only store identifiable information on Access.

SOP were also provided for how to address missing information in the referral forms.

MISSING INFORMATION PROCEEDURE

ON RECEIVING REFERRAL

1. Check original referral letter when reviewed for missing information
2. If any information missing, send missing information letter back urgently to practice

WHEN MISSING INFORMATION RETURNED

1. Check against original referral letter to confirm that missing information has been provided
2. Check for consistency between reply to missing information letter from GP and the original referral letter, particularly regarding their clinical impression (dementia, normal, CIND)
3. If any information is still missing, or there are inconsistencies, write back to GP surgery to request further information or clarification. If writing a second time, send SECOND Letter to GP re Missing info 14_08_2015 template (example on next page).
    a. Delete parts of letter which are not applicable to the situation – e.g if no discrepancies or they have given their confidence.

Delete parts of letter which are not applicable to the situation – e.g if no discrepancies or they have given their confidence.

Dear Doctor,

Thank you for your reply regarding this patient. We are sorry but you still haven't quite addressed the missing areas fully. It is very important for the research that we know your clinical opinion about the diagnosis, how confident you were, and what led you to form your judgement because this enables us to analyse how good GPs judgements are compared to tests, which is a key part of this work.

Regarding confidence about the we need to know this on a scale from 0 – 100, so if you cant get the word file to work, or can't print, scan and email it to us, you could just say a percentage confidence.

Bristol and South Gloucestershire CCGs, patient groups and our main funder the Wellcome Trust recognise this as an important aspect of our work – and we are grateful your help.

We also noticed that there was a discrepancy between your clinical opinion in your original referral letter and your reply to our earlier email about missing information. Please could you confirm what your opinion was when you referred the patient?

With best wishes, and thank you for your help,

TIMeLi Memory Study timelimemorystudy@nhs.net

b) Is your gut feeling that this person

☐ Has dementia

☐ Has cognitive impairment but not dementia

☐ Has normal cognition for age

c) How confident are you in your opinion for (b)

Please move the x along in the box using the space bar to indicate your confidence from guess to certain.

No idea (guess)	Certain
x	

d) Please write a few words about what you think led you to form your gut feeling

MISSING INFORMATION LETTER 2

NAME: **NHS**: **DOB**:

Dear Dr ,

Delete parts of letter which are not applicable to the situation – e.g if no discrepancies or they have given their confidence.

Thank you for your reply regarding this patient. We are sorry for this second letter about missing information but you still haven't quite addressed the missing areas fully. It is very important for the research that we know your clinical opinion about the diagnosis, how confident you were, and what led you to form your judgement because this enables us to analyse how good GPs judgements are compared to tests, which is a key part of this work.

Regarding confidence about the we need to know this on a scale from 0 – 100, so if you can't get the word file to work, or can't print, scan and email it to us, you could just say a percentage confidence (0-100).

Bristol and South Gloucestershire CCGs, patient groups and our main funder the Wellcome Trust recognise this as an important aspect of our work – and we are grateful your help.

We also noticed that there was a discrepancy between your clinical opinion in your original referral letter and your reply to our earlier email about missing information. Please could you confirm what your opinion was when you referred the patient?

With best wishes, and thank you for your help,

TIMeLi Memory Study timelimemorystudy@nhs.net

b) Is your gut feeling that this person

☐ Has dementia

☐ Has cognitive impairment but not dementia

☐ Has normal cognition for age

c) How confident are you in your opinion for (b)

Please move the x along in the box using the space bar to indicate your confidence from guess to certain.

No idea (guess)	Certain
x	

d) Please write a few words about what you think led you to form your gut feeling

A SOP was also provided on how to deal with issues with appointments and eligibility.

STANDARD OPERATING PROCEDURE

FAQS REGARDING ELIGIBILITY AND APPOINTMENTS

1. **Participant has been prepped for clinic and has study ID assigned and pack made but now can't come at all or has declined so will not be participating.**

Example: Y6 was booked to come on 1 June 2013 but now doesn't want to take part.

Principle: Study IDs and packs should be uniquely linked to participants or potential participants.

Solution: The next person to be prepped for clinic from surgery Y should be assigned the next sequential ID e.g. Y7 if this is not assigned. Pack Y6 should be securely discarded as it will contain patient identifiable information on it. Stata code referring to the random order of the tests should not be altered and should allow for the (now non-participating) Y6. Participant Y6 should have an ineligible or declined letter sent in accordance with the SOP.

2. Participant had been prepped for clinic and had study ID assigned and pack made but had to change appointment.

Example: Y3 was booked to come on 1 June 2013 but is now coming on 1 September 2013.

Principle: Study IDs and packs should be uniquely linked to participants or potential participants.

Solution: It is fine to use the previously prepared pack and study ID for this unique participant. Send the letter to GP surgery to confirm still appropriate to attend:

> Dear Doctor,
>
> Thank you very much for referring XXX to the Timeli memory study. They were supposed to attend our clinic on XXX but unfortunately they were unable to attend because they were unwell. We would normally arrange a new appointment for participants who failed to attend but we are not sure whether this is still appropriate given this episode.
>
> Could you please let us know whether you think it is still reasonable for us to see them via return email within the next few days. If we do not hear from you, we will assume it is still reasonable to see them. Thank you very much for your help. Best Wishes YOUR NAME

3. Who to email if main contact away

Example: Surgery G have an out-of-office to say they are not in this week.

Principle: Communication with surgeries should be prompt and contemporary. Risks of data breach should be minimised.

Solution: Phone the surgery and let them know there is an email waiting for the person who is away and that due to data protection we would prefer not to send it to someone else - can they open it as a delegated mailbox? If this isn't possible then ask the surgery to email you from the email address that they would like you to email.

4. Informant has already been

Example: Mr Jones came along with participant Y3 and now is planning to come with participant W2.

Principle: Confidentiality should be preserved.

Solution: It is acceptable for the same informant to come along with a different informant. Get permission from the new participant (W2 in example) to talk to the participant and emphasise the importance of confidentiality at the clinic. Alert Sam Creavin to the situation before clinic and on the case report forms so that confidentiality can be emphasised on the day as well.

5. Changed mind about informant

Example: Fictional participant GG had decided that he couldn't find a suitable informant and a letter sent to the surgery to say he was ineligible but he has now found someone to come with him.

Principle: Participation should be offered if appropriate.

Solution: A new referral is NOT required but the GP should be updated. Get verbal permission from GG to write to the GP and inform them of this change and that he is now eligible. Use following wording, using wording based on email for cancelled appointment below:

> Dear Doctor,
>
> Thank you very much for referring XXX to the Timeli memory study. You had referred them but XXX was unable to find an informant and was therefore not suitable. However, XX has now found an informant to come along and given us verbal permission to inform you of this change in circumstances. We have arranged a new appointment but we are not sure whether this is still appropriate.
>
> Could you please let us know whether you think it is still reasonable for us to see them via return email within the next few days. If we do not hear from you, we will assume it is still reasonable to see them. Thank you very much for your help. Best Wishes YOUR NAME

6. How long to wait between referral and attendance at clinic

Example: Fictional participant HH had been referred on 1 April 2015 but was unable to attend clinic until 1 July 2015.

Principle: Participation should be offered if appropriate.

Solution: Appointments at clinic can continue to be offered up to six months from the original referral. A new referral is NOT required but the GP should be updated. Get verbal permission from GG to write to the GP and inform them of this change and that he is now eligible. Use following wording, using wording based on email for cancelled appointment below:

> Dear Doctor,
>
> Thank you very much for referring XXX to the Timeli memory study. They were supposed to attend our clinic on XXX but unfortunately they have been unable to attend appointments until now. Their appointment is booked for XXX but we are not sure whether this is still appropriate given the time interval.
>
> Could you please let us know whether you think it is still reasonable for us to see them via return email within the next few days. If we do not hear from you, we will assume it is still reasonable to see them. Thank you very much for your help. Best Wishes YOUR NAME

7. Re-referral

Example: Fictional participant JJ had been referred on 1 April 2015 but declined to take part and a declined letter was sent to the GP. Now JJ has been re-referred as he has changed his mind and would like to take part.

Principle: Participation should be offered if appropriate.

Solution: It is fine to offer participation to JJ if he would like to take part. Both the original referral date and the new referral date should be kept on the Spread sheet that is used to track the interval between referral dates, appointments and letters back to surgeries.

8. Prior memory clinic appointment

Example: Fictional participant QQ had been seen in a memory clinic 10 years ago but not diagnosed with dementia.

Principle: Participation should be offered if appropriate.

Solution: Attendance at a memory clinic in the past is not an exclusion criterion for the study. However, people who have already been given a diagnosis of dementia are not suitable. People who have been seen at a memory clinic in the past are suitable, and it does not matter when they were seen. We do need to be careful about being used as a second opinion service: if appointment at memory clinic was within the last six months then advise that participation will probably be OK but need to confirm.

9. Spouse or relative has already been as a patient.

Example: Fictional participant BB had been seen at TIMeli with his daughter. Now BB's wife would like to come along with the same daughter.

Principle: Participation should be offered if appropriate. Confidentiality should be preserved.

Solution: See FAQ 4. The spouse should be offered participation but confidentiality should be emphasised to the informant – the daughter in the example. Eligibility should be checked carefully.

A standard letter was sent out to people to confirm the details of their appointment, together with a map and a consent form and participant information leaflet.

CONFIRMATION OF APPOINTMENT DETAILS

Your appointment with the timeli memory study is on 06/08/2015 at in the morning.

The appointment will be at
Hanham Surgery, 33 Whittucks Road, Hanham BS15 3HY.

Transport to the appointment
will be provided by you or a helper / carer

Someone will contact you in the week of the appointment to confirm everything is still OK.

If you have any concerns or questions you can contact us on the study mobile phone number which is 0777 347 2622.

Dr Sam Creavin will see you at your appointment.

Best Wishes,

Timeli memory study

If there appeared to be no concern about possible dementia a standard email was sent back to the GP.

STANDARD OPERATING PROCEDURE

APPEARS THERE IS NO CONCERN ABOUT POSSIBLE DEMENTIA

1. In addition to sending file ineligible or declined also put following text in email to practice:

```
Hi XXXX,

Thanks for sending this referral back to me so quickly. Dr. XXXX
said in his referral letter that noone (neither him, the patient
or  a relative) is concerned about possible dementia. It is a
core inclusion criteria that someone (one of the three mentioned
above) is concerned about possible dementia.

Please could you clarify with Dr XXXX and ask him to re-refer the
patient if appropriate..

Many Thanks,

Your Name
```

Standard correspondence templates were provided for communication with GP surgeries and for booking transport.

STANDARD OPERATING PROCEDURE : EMAILS TO SURGERIES

BASICS

1. All emails are sent from secure nhs.net mail to a nhs.net account
2. The timelimemorystudy@nhs.net should be used for all emails to surgeries
3. Always replying to the original referral email to ensure the correct surgery email address is used
4. When sending any attachment confirm with three forms of ID (Nhs number DOB and name) that the letter is going to the correct surgery.
    a. Check against the referral letter saved in the relevant folder on the G drive for the appropriate surgery

LETTER IF INELIGIBLE OR DECLINED

1. Reply to the original referral email to ensure the correct surgery email address is used
2. When letter is attached on nhs.net mail view the attachment to confirm
    a. The attached letter relates to the patient in the email
    b. The surgery email address is the correct one to use for the patient (check that the person in the attachment was referred by the surgery in the TO email address)
    c. The last 2 digits of the letter filename should refer to the same surgery in the email field
3. Use the following wording for the body of the email:
    a. `Dear XXX Please find attached a letter for the attention of Dr XXX regarding XXX. Best Wishes YOUR NAME`

LETTER RE PARTICIPATION

1. Reply to the original referral email to ensure the correct surgery email address is used
2. When letter is attached on nhs.net mail view the attachment to confirm
    a. The attached letter relates to the patient in the email
    b. The surgery email address is the correct one to use for the patient (check that the person in the attachment was referred by the surgery in the TO email address)
    c. The last 2 digits of the letter filename should refer to the same surgery in the email field
3. Use the following wording for the body of the email:
    a. `Dear XXX Please find attached a letter for the attention of Dr XXX regarding XXX. A further letter will follow in due course with a clinical opinion. Best wishes YOUR NAME`

LETTER SENT TO WRONG SURGERY

If identifying that a letter has been sent to the wrong surgery

1. Contact Sam Creavin immediately who will
    a. If email sent to an nhs.net account

i. Contact the receiving surgery by phone and ask them to delete the email and sent written confirmation this has been done
ii. Document this adverse event in a secure file
iii. Inform Supervisors by routine email

b. If email sent to a non secure account (non nhs)
i. Contact receiver immediately by phone or email and ask them to delete the email and sent written confirmation this has been done
ii. Document this adverse event in a secure file
iii. Inform Supervisors urgently same day (or next working day)
iv. Follow AE policy and
1. Discuss with RED and UH Bristol urgently same day (or next working day)

STANDARD REPLY TO SEND IF EMAILED PATIENT TEST RESULTS

```
Thank you so much for your email. Actually we don't need to be
emailed a copy of the blood results or tests - we will request
access to these as and when it is necessary but don't need to be
sent them unless asked.

Hope this makes sense, no reply needed.

Thanks again, so much, for all your help with the study.
YOUR NAME
```

TEXT TO ACCOMPANY NEW CLINICAL OPINION LETTERS FROM 5 OCTOBER 2015

1. Reply to the original referral email to ensure the correct surgery email address is used
2. When letter is attached on nhs.net mail view the attachment to confirm
   a. The attached letter relates to the patient in the email
   b. The surgery email address is the correct one to use for the patient (check that the person in the attachment was referred by the surgery in the TO email address)
   c. The last 2 digits of the letter filename should refer to the same surgery in the email field
3. Use the following wording for the body of the email:

```
Dear XXX Please find attached a letter for the
attention of Dr XXX regarding a clinical opinion for
XXX.
Due to feedback from the NHS memory service we have
changed the format of our clinical opinion letter
slightly and all future letters will be sent in this
format.
Thanks for your help with the study.
Best wishes
YOUR NAME
```

LETTER RE CLINICAL OPINION PRE 5 OCTOBER 2015

4. Reply to the original referral email to ensure the correct surgery email address is used

5. When letter is attached on nhs.net mail view the attachment to confirm
    a. The attached letter relates to the patient in the email
    b. The surgery email address is the correct one to use for the patient (check that the person in the attachment was referred by the surgery in the TO email address)
    c. The last 2 digits of the letter filename should refer to the same surgery in the email field
6. Use the following wording for the body of the email:
    a. Dear XXX Please find attached a letter for the attention of Dr XXX regarding a clinical opinion for XXX. Best wishes YOUR NAME

EMAIL TO SEND IF REFERRAL LETTER SENT TO BRISTOL.AC.UK ACCOUNT

1. Reply to the original referral email to ensure the correct surgery email address is used
2. Ensure you remove any attachements – sometimes they are embedded in the text so if you just reply you will inadvertently include all of the patient identifiable information. It is best to select Control +A and delete and clear the text of the email before composing the reply

Dear XXXX,
Thanks for your referral re XXX.. Please can you arrange for this referral to be resent to timelimemorystudy@nhs.net so that we can process it.
Please note referrals containing patient identifiable information must be sent to and from an nhs.net account i.e. not this Bristol account
Once we have this in the nhs.net account we can process from there. Thanks for your help,
YOUR NAME

TRANSPORT EMAIL

Dear BCT,

I have details of another booking for transport for the TIMeLi Memory diagnosis study.

Date:

Pickup:

From:

Going to:

Arrive time:

Pick up time:

Number of people travelling:

Mobility issues? i.e. need wheelchair? ☐

Please can you call them on to confirm the booking and email me?

Many thanks,

Your Name

Send to bctoffice@hctgroup.org

Document request in Google spreadsheet "Timeli referrals and practices" | Transport sheet

Document also in Access and on Trello (with no identifiable information)

A standard process was used for booking clinics and confirming them with the research site.

STANDARD OPERATING PROCEDURE : BOOKING CLINIC SITES

ASSIGNING CLINICS

1. When clinics are put on calendar, determine ideal dates from TIMeli perspective
    a. Ideally we want alternating weeks at each site: Hanham / Shirehampton so take account of the current bookings
    b. Ensure all bookings that are not confirmed are in as "Hanham or Shirehampton"
2. Email Hanham (mailto:Lorraine.Dodimead@gp-l81079.nhs.uk cc mailto:Darren.Maslen@gp-l81079.nhs.uk and Kim.Elmes@gp-l81079.nhs.uk) and Shirehampton (mailto:sue.buckley@gp-l81008.nhs.uk) to request desired dates
    a. Shirehampton can usually do most dates, though Friday is more difficult for them.
    b. Hanham can often only confirm with 3 weeks notice
3. When availability confirmed by practice update calendar

CHANGING ASSIGNED CLINIC

1. If participant prefers a particular site, and cannot make an appointment at that site within three weeks, then explore possibility of changing study site.
    a. Explain to participant you will try to find an earlier slot for them and ask permission to phone them back – arrange a time and ensure if not same day that it is booked on Google calendar and notes made on Trello (with no identifiable information) and Access
    b. Phone Shirehampton- Sue Buckley - 0117-9162230 and Darren Maslen (or Kim Elmes or Lorraine Dodimead) 0117 9352318 to request desired dates
    c. If preferred surgery can swap and cover the clinic date then:
        i. Update calendar
        ii. Email surgery to confirm in writing (details above) and Sam Creavin
        iii. Phone participant back and offer new date
        iv. Return to follow Initial phone call SOP and document Reponses on Access; add tasks (non identifiable) to Trello and calendar and update Google Spreadsheet
    d. If preferred surgery cannot swap then
        i. Phone participant back and explain no earlier date at preferred site available and confirm whether they would rather (a) wait for their preferred site or (b) be seen earlier at alternative site.
        ii. Document response on Access

If the appointment was cancelled for medical reasons a standard process was followed.

STANDARD OPERATING PROCEDURE : CANCELLED APPOINTMENT FOR MEDICAL REASONS

SCOPE

To be used when a participant has phoned in advance of clinic – or on the day – to cancel the appointment for medical reasons.

In event of non-arrival at clinic (DNA) where no phone call is made in advance refer to TIMeLi fire and DNA SOP version 2.0 2015_03_26 for managing the DNA then refer back to this SOP for cancelling transport and arranging re-booking.

FIRST STEP

1. If the patient is still unwell, ensure that the patient has or is seeking local medical advice. Ask carer or helper are they able to arrange medical care – advise to call own GP, 111 or 999 depending on the severity of the problem.
    a. If carer unable to contact appropriate assistance then researcher should phone appropriate help through either via GP, 111 or 999 if urgent.
2. Contact transport and inform them of cancelled booking

WHEN PATIENT SAFE

1. Arrange with carer to contact them again in three weeks with a view to arranging another appointment.
2. Document date in Access database and put on Google calendar

WHEN CARER RE-CONTACTED

1. If now able to re-arrange an appointment do so.
    a. Explain will contact surgery to confirm they still feel appropriate for patient to take part
2. If not, rearrange another time to call and book a slot on the Google calendar and update Access.

WHEN APPOINTMENT REARRANGED

1. Contact referring surgery with following email:
    a. Dear Doctor,
    b. Thank you very much for referring XXX to the Timeli memory study. They were supposed to attend our clinic on XXX but unfortunately they were unable to attend because they were unwell. We would normally arrange a new appointment for participants who failed to attend but we are not sure whether this is still appropriate given this episode. Could you please let us know whether you think it is still reasonable for us to see them via return email within the next few days. If we do not hear from you, we will assume it is still reasonable to see them. Thank you very much for your help. Best Wishes YOUR NAME

APPENDIX C

RESEARCH CLINIC SOP

This appendix includes the SOP that were used on the day of the research clinic. A SOP was available to check off the tasks that were required on the day of clinic.

STANDARD OPERATING PROCEDURE

CHECKLIST TO DO BEFORE EVERY CLINIC

1. Assigned participants study ID
2. Randomised order of test for each participant
3. Printed Case Report Form for participants
4. Arrange index test battery CRF in correct order for each participant and document at back of CRF
5. Ensure have correct coins: 2x £1; 1x 50p; 3x 20p; 7 x 10p; 7x 5p available
6. Need tape measure that can measure up to 2m
7. Arrange refreshments
8. Ensure contact details are available for participants and researchers and clinic
9. Check contact details for participants
10. Print signs

ON ARRIVAL AT THE CLINIC

1. Confirm receptionists have been briefed
2. Put up direction signs
3. Confirm details of when surgery must close
4. Confirm where fire muster point and action to be taken in event of fire
5. Prepare rooms for research clinic
6. Check case report forms

A standard procedure was used for performing the index tests, in addition to following the recommended process by the original authors.

STANDARD OPERATING PROCEDURE

RUNNING THE INDEX TEST CASE REPORT FORM CLINIC

1. Indicate that the testing will begin shortly: ***"we're almost ready to start, do you want to ask anything?"***
2. Complete the details on page 3 of Timeli case report form regarding date of birth and name.
3. Give the informant case report booklet to the informant and state: "***There are some questions for you too, about X. Please can you work through this booklet whilst I talk to X. It should be self-explanatory, but please do ask me at the end if you have any questions. If anything is particularly troubling then please feel free to interrupt me as we go along.***
   Is the text and font clear enough for you to read?
   Please enter your details at the front of the book.
   If you make a mistake, please cross it out with a single line and put your initials.
   If you don't know the answer to a question, please don't mark a box but put "don't know" at the side.
   Is that OK?
4. Tell the participant ***"OK, we're ready to start"***

COMPLETING THE BOOKLET

1. Smile at the participant and be encouraging and gentle in body language and tone throughout the interview
2. Work through the case report form **in the order that it is presented: this may not result in correct page sequencing** (order is randomized and booklet will have been sorted prior to clinic).
3. Enter the time that each test begins at the top of the page in the appropriate box.
4. For questions that ask for recall of items, only allow the exact term e.g. if participant is asked to remember AXE and they state HAMMER this scores zero (wrong).
5. If participant asks for clarification, repeat the question, but do not re-word it.
6. If participant asks a question, for example "is it a type of nut?" leave a moments silence and then state the question again in a warm tone.
7. If asked to list names or items and the question is timed, and participant stops listing items, say "***you have a bit more time"*** or "***you have another ten seconds"***
8. For Scenery Picture Memory Test, when stating numbers for repeating, state each number at a time, but the whole string in one go e.g. "***1635" – repeat " 79267" – repeat " "303265" – repeat " 1801774" – repeat.***
9. When asked to name time to nearest hour, write down the stated time and score if correct to within the nearest hour.
10. When asking informant questions for GPCOG it is acceptable to replace the term "the patient" with the patient's name e.g. "***does bob have more trouble remembering things that have happened recently?"***
11. When scoring TUG allow one practice and then three time-trials. Start the timer when they start to rise and end when they have turned to sit down and placed their bottom on the seat.

Eurotest

12. When asking questions about money in Eurotest and asked to list coins, when the participant appears to have finished listing coins, ask "***any more?".*** This can be repeated once (two occasions in total) if they then list some more.
    a. Do not tell participant what they have already said
    b. If they list notes (when asked to list coins, or vice-versa) then say "***I asked for coins and not bank-notes***" and score it an error.
13. When scoring Eurotest for splitting money into piles, be sure the participant has finished sorting and is submitting an answer. They may do this by presenting the piles to you or indicating that they have finished, but if not ask: "***Just so I can be sure – is that answer?".***
14. If participant offers a wrong answer say: "***No, it is not correct, try again.*** "
15. If wrong twice, say "***Okay, now I want you to tell me … [the next item]"***
16. If not offered an answer after 1 minute say ***"Let's try something else, now I want you to tell me [the next item]"***
17. When asked to name animals in Eurotest, if a long pause before any response repeat the statement and offer dog as an example
18. When asked to name how many coins there were (2.9) and how much there was (2.10), it is common to start listing the types of coins, in this case say: "***no, I just want you to tell me how many coins there were overall, not what they were"*** or, for 2.10 ***"no just tell me how much money there was in total, not what the coins were".*** If despite this, they carry on consider as if responding to 2.11 and when they finish say "***OK, how many coins where there in total"***

WHEN THE INDEX TEST BOOK IS COMPLETED

1. Tell the participant "***well done, that is the end of the questions. How was that for you?"***
2. ***"I'm just going to check I haven't missed anything, now would be a good time to go to the loo if you need to"***
3. Check informant booklet through with informant and clarify any questions
4. Check back through the index test CRFs to confirm that all questions have been completed
5. Write study ID at the top of every page and on the front page of the booklets.
6. Confirm that the order of test administration to the patient is written in the CRFs and documented, and that all time boxes are complete
7. Re-sort the patient CRF into the correct page order for scanning and coding.
8. If there are any discrepancies in the informant booklet, for example questions are answered very differently to expected then confirm with informant that was what was intended – "***Just so I can be sure, you have put this here, have I understood that right?"***

INDICATE THAT IS THE END OF THIS SECTION

1. Arrange for refreshments (if the reference test is next) or ask for feedback (if there are no more tests).
    a. The information in the post
    b. The transport
    c. The process of doing the tests
2. Confirm the next steps with participant
    a. Have they had a scan and blood tests yet, if not they might need one and we will write to the surgery

b. They should see their GP in 3 weeks (if scan requested) or 6 weeks (if scan not done yet) to go through the letter from the research team and make a plan.
c. If in the next 24 hours they need to be admitted to hospital or have any problems, please could they let us know (give study mobile phone number)
d. We will write to your GP with a diagnosis once we have looked over your results.

3. If asked "how did I do" or similar:
   a. "***You did well to get through so many of the questions."*** It may be appropriate to also say: "***Maybe today was a bad day. Don't worry, the two doctors make two separate assessments so the other doctor won't take account of any of these tests in reaching her decision".***

A standard process was used to consent participants and to judge the capacity to consent.

STANDARD OPERATING PROCEDURE

ARRIVAL AT RESEARCH CLINIC

1. On arrival at research clinic, show participants to the research waiting area
2. Introduce self and role.
3. Ask them to introduce themselves and what they would like to be called
4. Show participants where toilet is
5. Explain procedures for fire alarm (see fire SOP)

IN THE RESEARCH CLINIC ROOM

1. Explain what is going to happen: "***First I'm going to check you have read through the information sheets, then, if it's OK, I'll see you alone for a moment to ask you a few questions to check you have understood the most important bits. Then we will all go through the consent forms together, both for you (patient) and for you (informant). Then we will start doing the study.***"
2. Confirm they understand what they are there for: "***can you briefly tell me why you are here?***"
3. Acknowledge the prior telephone conversation
4. Ask if there is anything in particular that they would like to discuss: "***is there anything particular about the study that you would like to ask, or like to discuss?***"
5. ***You will see two doctors here today, one is me, and the other is a memory specialist. Both doctors will do tests of memory and brain functioning. It is important today that the two doctors don't know anything at all about what happened in the other consultation, so some of the questions might be repeated, and please try not to tell the doctors anything about what happened with the other doctor. The order you will see the doctors in has been decided by random chance.***
6. ***All the questions we will ask today are taken from tests that have been used before by other people, we have to use the exact wording that that they have suggested. We are sorry if some are hard to understand but do please tell us what you think of the questions as this will help us to make new questions better and easier for the future.***

INFORMATION SHEETS

1. Confirm they have had a chance to read through the information sheets and ask if there is anything they would like to ask: "***Did you receive a copy of the information leaflet in the post? Have you had a chance to read through it? Is there anything in particular you would like to ask?***"

ASSESSING CAPACITY

1. Ask the informant to step outside for a moment "***is it OK if I see X by him/herself for a moment***?"
2. Confirm they understand what they are there for: "***can you briefly tell me why you are here?***"
3. If they are not sure, offer a prompt: "***was it something to do with your memory?***"
4. If they are still not sure consider carefully whether the person has capacity.
5. The Mental Capacity Act (2005) guides the decision-making about capacity.

6. If the person appears to lack capacity, call the informant back in, and explain: "***X, I'm sorry but I'm not completely sure that you can remember enough about the study to make an informed decision to take part. It is important you are able to decide for yourself. Perhaps Y can help me to check how much you understand"***
7. ***"X, can you tell me briefly why you are here today?"***
8. ***"Y, you can help X if you like"***
9. If X still appears to lack capacity then say: ***"OK, don't worry about it. You will still be able to help us out a lot by being here today. You will see the memory specialist first, and you don't need to worry about seeing the second doctor"***

HELP WITH ASSESSING CAPACITY*

*Church & Watts(2007). Assessment of mental capacity: a flow chart guide. DOI: 10.1192/pb.bp.106.011353

A process was prepared in the event that a participant wanted to withdraw part of the way through the research clinic.

STANDARD OPERATING PROCEDURE

EXPLICIT PATIENT WITHDRAWAL

To be used when, for example, a participant says, "I have had enough" or "can we stop".

1. Clarify that the participant would like to completely withdraw from this stage of the study, or would they like to try again, either after a short break, or on another day: "***Of course, I understand we are asking a lot of you and we are grateful for your help. Just so I can be sure I understand, would you like to stop completely for good, or would you like to come back to this another time, either later on today, or on a separate occasion?"***
    a. If they reply that they would like to have a break***, agree how long the break will be***
    b. If they would like to come back another day, ***agree a day and time to phone to arrange another appointment.***
2. If they would like to withdraw completely: "I ***understand, and of course, that's fine. Thank you for your help. As we discussed at the start, we will use the anonymised data that we have collected so far when we analyse our results."***
    a. Discuss any response. If participant wants data not to be used, explain ***this would mean completely withdrawing from the study, are they sure this is what they want? Would they like to think about it and then discuss another time (arrange a time)***
3. Clarify are they happy to be electronically followed: ***"At the start of the day, when we did the consent forms, I said that we would plan to follow up your electronic records so we can see what has happened to your health. You said Yes that was OK. Is that still OK?"***
4. Document responses on notes field in case report form.

EXAMINER PERCEIVES PARTICIPANT TO BE STRUGGLING

To be used when, for example, a participant appears to be taking longer over questions than other participants.

1. Acknowledge the issue "***I can see some of these questions are causing you some bother".***
2. Ask: ***"are you happy to carry on or would you like to have a short break?***"
    a. If wanting to have a short break clarify whether the participant would like to try again after a short break today, or on another day: "***Of course, I understand we are asking a lot of you and we are grateful for your help. Just so I can be sure I understand, would you like would you like to come back to this later on today, or on a separate occasion?"***
    b. If they reply that they would like to have a break***, agree how long the break will be***
    c. If they would like to come back another day, ***agree a day and time to phone to arrange another appointment.***
3. If happy to continue, "***OK, thank you. If you do feel you want to have a break at any point, let me know, and if happen to think again that questions are bothering you I will check with you again.***"
4. Document intervention and responses on notes field in case report form.

A standardised process was prepared in the event of non attendance or fire.

STANDARD OPERATING PROCEDURE

FIRE

1. Find muster point on arrival at the surgery
2. Follow consent SOP to ensure that the fire exits have been explained on arrival at the clinic
3. If the fire alarm sounds, assist the participant to leave the building and go to the muster point
4. If possible, take study documentation when leaving the building in the study site file, but do not stop or delay to find the documents.
5. Once in safety, make a note of what question had been reached and document the time.
6. Once "all clear" is given, if able to access study documentation, make note of events in notes box on case report form.
7. Ask participants if they would like to continue, rearrange appointment, or withdraw: see withdrawal SOP.

LATE ATTENDANCE

1. Clarify on arrival at the clinic what time that the surgery must close at, and what time that researchers can stay until.
2. If a participant is more than 20 minutes late for the booked appointment then attempt to contact them; if contact is made, introduce self and explain reason for call, ***to check everything is OK.***
3. If participants have not arrived after 40 minutes, attempt to call them again (step 1.)
    a. If participants arrive more than 40 minutes late, offer them refreshments, a loo break and five minutes to sit without interruption.
    b. Assess whether continuing with the appointment that day is viable (accounting for time, surgery factors, researcher factors and circumstances).
    c. **Once they are comfortable** (having used the loo, with a cup of tea if they wish). If appropriate ask whether they would like to continue with their appointment today or rearrange for another time. If it is not possible to complete the appointment in the remaining time, ***apologise, explain it will not be possible to have the appointment in the remaining time, and explain the reasons for this. Offer to call to arrange another time and agree a day and time to call.***
    d. If participant would not like to continue with appointment on this occasion, see withdrawal SOP
    e. Document discussion
4. If participants have not arrived after 60 minutes, attempt to call them again (step 1.)
    a. Usually at this stage participants will be unable to continue with the appointment if they arrive late on this day. See steps a-e above.
5. If participants do note attend, follow up the following day on the telephone, and if contact is not made, in writing.

A standardised process was prepared in the event of serious adverse events or safeguarding concerns.

STANDARD OPERATING PROCEDURE

DEFINITIONS

For this study, we define an adverse event as an unexpected effect of an untoward clinical event affecting the participant that

1. Occurs during or within one hour of attendance at the single visit to the research clinic

   and

   a) Results in death;
   b) Requires hospitalisation or prolongation of existing hospitalisation;
   c) Results in persistent or significant disability or incapacity;
   d) Is otherwise considered medically significant by the investigator.

DETECTING AND RECORDING ADVERSE EVENTS

Adverse event detection will continue until 24 hours after the research clinic visit, as from this point onwards the research process will consist only of monitoring the patient's electronic medical record.

Adverse events may be reported by several methods:

1. Directly by the participant (i.e. by email, phone call or voice mail message)
2. Indirectly from family members, carers, guardians or representatives
3. From the participants GP practice

REPORTING ADVERSE EVENTS TO STUDY TEAM

Participants and GP practice staff will be asked to notify any adverse event that they believe may have occurred as a result of the research process.

ON NOTIFICATION OF AN ADVERSE EVENT

On notification of such an adverse event which may be related to the research process or intervention, a researcher should complete an adverse event reporting form within 5 working days, paying specific attention to information regarding the timescale of events i.e. when the event started, were there any specific changes to medication or behaviour preceding the event. Further information should be requested from the participant or GP/practice nurse as necessary.

A completed form should be securely sent to the Chief Investigator.

Reporting of adverse events

Adverse events will be reported as follows:

1. To the sponsor immediately,
2. To the UH Bristol contact (fax 0117 3420239 or research@uhbristol.nhs.uk) by investigational staff within 24 hours of their knowledge of the event,

3. To the REC within 7 days of the Chief Investigator becoming aware of the event.

Any relevant further information will be subsequently communicated within 8 days.

ADVERSE EVENTS REPORTING FORM

1	Study ID	
2	Date of Birth	
3	Research site ID	
4	Research site name	
5	Description of adverse event	
6	Date of onset	
7	Has the event resolved	Yes
		No
		Ongoing
8	Date resolved	
9	If resolved and the SAE involved admission to hospital please give a summary of the discharge diagnosis	
10	Which serious category did the event match? (one box only)	Resulted in death
		Life Threatening
		Required hospitalisation, or prolongation of existing hospitalisation
		Persistent or significant disability /incapacity
		Other important medical condition
11	Was the event related to study participation?	Unrelated
		Unlikely to be related
		Possibly related
		Probably related
		Definitely related
12	Was the event expected	Expected
		Unexpected
13	Date SAE form completed	
14	Person completing form (signature)	
15	Please print name	
16	Please print position	

STANDARD OPERATING PROCEDURE

DRIVING

1. Remind the participant that safeguarding concerns were discussed on the telephone and on arrival at the clinic e.g.: "***Do you remember talking earlier about what would happen if we had concerns about your welfare?"***
2. Explain that the examiner has concerns about the participants ability to drive, including the ability to drive home after the research clinic.
3. Explain that we will write to the GP to inform them of this.
4. Explain that the participant has a duty to inform the DVLA
5. If they would like to undergo an assessment of their driving skills, this is possible.
6. Document discussion

CONCERN ABOUT MEDICAL DISORDER

1. Remind the participant that safeguarding concerns were discussed on the telephone and on arrival at the clinic e.g.: "***Do you remember talking earlier about what would happen if we had concerns about your welfare?"***
2. Explain that the examiner has concerns that the participant may have a medical disorder – briefly explain what.
3. Ask permission to write to the GP to inform them of this.
4. Explain we will follow up in writing to confirm what has been discussed
5. Document discussion

CONCERN ABOUT HEALTH AND SOCIAL CARE

1. Remind the participant that safeguarding concerns were discussed on the telephone and on arrival at the clinic e.g.: "***Do you remember talking earlier about what would happen if we had concerns about your welfare?"***
2. Explain that the examiner has concerns about the care arrangements for the participant – briefly explain what.
3. If there are concerns about the care that is provided by the informant, ask to see the patient on their own "***Can I have a few words with xx by themselves for a moment?***"
4. Ask permission to write to the GP to inform them of this.
5. If immediate action needs to be taken then ask permission to contact the appropriate authority directly. The appropriate authority may include:
    a. The duty social worker for issues of social care provision.
        i. South Gloucestershire: 01454 868007 or EDT 01454 615165.
        ii. Bristol: 0117 922 2700 or EDT 01454 615165.
        iii. North Somerset: or EDT 01454 615 165.)
    b. NHS England for concerns about GP practices (0300 311 22 33).
    c. The GMC for issues of fitness to practice for medical practitioner (0161 923 6399). This requires a discussion with a senior study supervisor (Prof Sarah Purdy or Prof Yoav Ben Shlomo first).
    d. The CQC for issues of quality and safety in healthcare settings (03000 61 61 61). This requires a discussion with a senior study supervisor (Prof Sarah Purdy or Prof Yoav Ben Shlomo first).
    e. The police if there is concern that a crime may have been committed.

6. If permission is not given, and life is deemed to be at risk, then explain a disclosure must be made anyway.
7. Document discussion

APPENDIX D

SOP FOR ISSUES AFTER RESEARCH CLINIC

This appendix includes the SOP that were used after the research clinic. A SOP was available to in the event that a participant requested a copy of their notes. Sometimes this was done if they were attending the NHS memory clinic.

TIMeLi memory diagnosis study

University of BRISTOL

TimeLi Memory Study
Dr Sam Creavin
Centre for Academic Primary Care,
School of Social and Community Medicine,
University of Bristol,
39 Whatley Road, Bristol,
BS8 2PS

Dear ,

I believe you would like to request a copy of your notes from your consultation with Dr Haworth the specialist at the TIMeLi Memory study. If this is correct, please could you complete the enclosed letter and return to us in the self-addressed envelope. You will need to complete the letter with the address you wish us to post the information to and sign it at the bottom.

If this is not correct then no further action is required though you may wish to contact us by telephone, the study mobile phone is 07537 167 260 though you may need to leave a message.

Best wishes,

TIMeLi Memory Study

Letter for me to send back to participant

Dear XXX,

I have enclosed a copy of your notes from the specialist from the Timeli memory clinic on **XXXX.** This information resulted purely from your attendance at the TIMeli Memory clinic and included a standardised evaluation that is used in everyday clinical practice, including the Addenbrooke's Cognitive Examination.

With best wishes for your appointment and thanks for your participation in TIMeli.

Sam Creavin

TO:

TimeLi Memory Study
Dr Sam Creavin
Centre for Academic Primary Care,
School of Social and Community Medicine,
University of Bristol,
39 Whatley Road, Bristol,
BS8 2PS

Dear Doctor Creavin,

I am a participant in the Timeli memory study. I am due to attend an NHS memory clinic in the near future. To save time I would like to take a copy of the notes from my consultation with the specialist at the Timeli memory study with me. Please could you post me a copy of the notes using the royal mail to _____?

Yours Faithfully,

A standard procedure was used for scanning the notes to secure electronic archive storage.

STANDARD OPERATING PROCEDURE

SCANNING

1. Remove the participant identifiable information sheet from the informant booklet and the index test CRF and scan these separately
2. Then ensure that every page of the CRF booklets have a participant study ID on them
3. Scan pages – ensure that if 2-sided that the 2 sided option is selected on the scanner
4. Move the files immediately from the P drive to the G| Studies| Timeli | Case Report Forms
5. Rename the files using the scheme as follows:
    a. Study ID_type_person_date
    b. Where type is (consent | CRF|details) and person is (inf = informant | p = patient | reference = reference test) and date takes format ddmmyyyy
6. Ensure all pages have scanned and are legible – cross check against the paper form
7. Ensure paper forms are securely stored again

APPENDIX E

CASE REPORT FORM AND REDCAP VERSIONS

This appendix describes the revisions to the case report form and the REDCap data capture instrument referred to in Chapter 3 Section 3.6.3.
The case report form had the following revisions:

1. Revision 1 [n=35] used from 17-04-2015
2. Revision 2 [n=15] used from 16-10-2015 Added order of assessments boxes on front page (index test assessment first or reference test assessment first); added study site box on front page; removed MAT; added MOCA
3. Revision 3 [n=191] used from 31-05-2016 Added Sniffin sticks test

The REDCap data collection instrument had the following revisions:

1. Revision 1 15-01-2016 12:07 Removed "don't know" options from the informant section, as these were not available on the case report form and had been included in error
2. Revision 2 02-06-2016 09:40 Added additional new coder
3. Revision 3 27-03-2017 08:24 Added additional new coders and made coder notes required field on each page
4. Revision 4 03-07-2017 18:09 Added item for Eurotest recall errors, which had been erroneously been missed in previous iterations. All cases were re-coded to enter this item
5. Revision 5 01-08-2017 09:42 Added additional GP surgeries, on referral form 1
6. Revision 6 01-08-2017 09:44 Added additional GP surgeries, on referral form 2

The code in RStudio to randomly select a 10% sample of cases coded by each coder for duplicate data entry was:

```
>sample(c(list)),  size = number
```

Where list refers to the list of study ids for each coder and number refers to the number of cases reviewed for each coder.

APPENDIX F

GUIDES FOR QUALITATIVE INTERVIEWS

This appendix includes the guide for the qualitative interviews. The original version, and the final version, with all changes, is included. The intermediate versions are not available, because the guide was iteratively refined between interviews some of which took place on the same day.

Original topic guide for qualitative interviews with patients and families

1. Can you tell me a bit about your GP Practice and what your recent experience been of dealing with them?
2. What are you expecting for today?

Thinking now about memory problems in general

3. Can you tell me about when your memory last let you down?
4. When someone is worried about their memory who do you think should take the lead in diagnosing memory problems?
5. What do you think about a GP making a diagnosis of dementia for a patient without getting a specialist involved?
6. Some people have suggested the idea of a specialist GP for a local area who would take referrals from other GPs and act as a triage service, what do you think of that idea?
    a. Any further thoughts about how that might work?
    b. What about having a specialist nurse instead of a GP?
7. What matters most to you when trying to find out what is causing memory problems?
8. How do you think we could make things better for people in the future?

Thinking now about your experience of dealing with your GP about memory issues [removed after first few interviews due to interview perception of low relevance of answers to research question and therefore unnecessary burden]

9. What was your initial experience
10. How was the topic of memory trouble first raised?
11. How well did you know your GP before you spoke to them about this issue?
12. Did you talk to anyone else about possible difficulties with the memory?
    a. Friends and family
    b. Carers
    c. Hospital staff
    d. Nurses

- Participants' experience of talking to their GP about concerns about possible dementia
- Contrast between being diagnosed by GP compared to memory nurse or hospital
- Benefits of some GPs being able to make a confident diagnosis of dementia in some people

- Disadvantages and barriers to GPs diagnosing dementia
- Potential solutions to perceived barriers

Stop video [if videoing]

Do

1. MoCA [removed after first five interviews, other tests removed after first 2-3 interviews due to excessive burden]
2. Eurotest
3. GP COG
4. TUG
5. IQCODE

At the end of each test ask:

[to patient]

1. How did you feel when I was asking you those questions?
2. What did you notice about that test?
3. What do you think [carer] thought about that test?

[To carer]

4. How did you feel when I was asking you those questions?
5. What did you notice about that test?

[to both]

6. [to both] Does it seem like a good test for diagnosing memory problems?

Explain, "*I'm just going to make some brief notes about my thoughts while you were doing that test and then we will do the next one*".

Final topic guide for qualitative interviews with patients and families

1. Can you tell me a bit about your GP Practice and what your recent experience been of dealing with them?

Thinking now about memory problems in general

2. Can you tell me about when your memory last let you down?
3. When someone is worried about their memory who do you think should take the lead in diagnosing memory problems?
4. What do you think about a GP making a diagnosis of dementia for a patient without getting a specialist involved?
5. Some people have suggested the idea of a specialist GP for a local area who would take referrals from other GPs and act as a triage service, what do you think of that idea?
    a. Any further thoughts about how that might work?
    b. What about having a specialist nurse instead of a GP?
    c. *Some people say this wouldn't make any real difference as if you have to get in the car you have to travel anyway, what do you think? [for interviews from 14-6-17]*
6. What matters most to you when trying to find out what is causing memory problems?
7. How do you think we could make things better for people in the future?
8. Some people have suggested having "brain exercises" to try to help, what do you think about that? *[for interviews from 14-6-17]*
9. Some people have said that it doesn't matter if you have to wait a bit, as the memory has declined over months - year and so it shouldn't be a rushed process to try to sort it all out. What do you think*? [for interviews from 14-6-17]*

- Participants' experience of talking to their GP about concerns about possible dementia
- Contrast between being diagnosed by GP compared to memory nurse or hospital
- Benefits of some GPs being able to make a confident diagnosis of dementia in some people
- Disadvantages and barriers to GPs diagnosing dementia
- Potential solutions to perceived barriers

APPENDIX G

ETHNOGRAPHIC NOTES

This appendix includes the ethnographic notes from the qualitative interviews.

2017 24-02-2017.

In a fairly affluent suburb. House set approximately 5 minutes drive from the surgery or around 30-45 minutes walk for me. I guess middle - upper middleclass. In almost an entirely residential setting. Sun was shining, lovely day outside.

Arrived 10 minutes late as had some problems setting up computer before I left. Had tried to call to confirm time with them on the day but no answer, so wasn't sure they would be there. Had just been speaking to my wife on the telephone about a hospital appointment for her. His wife was leaving to go out to "buy a newspaper" as I arrived. I introduced myself, she remembered me from main clinic, and I went inside. Inside the house was comfortable and clean. Pictures on the wall that he had painted in younger life. Decoration tired but very presentable. We sat in the dining room [there was also a flight of stairs up, a small kitchen and a sitting room downstairs with a sofa and a couple of seats]. The wooden table in the dining had space for six and a big picture window that let the light in well, though not into our eyes.

Shook patient's hand and got kit set up. I quickly realised I had forgotten to print the interview topic guide, so tried to access it through 4g to no avail. I went down to the surgery in X and printed it off there. On my return they were both at home.

We started the interview and they sat on one side of the table and me on the end. They wanted me to talk to the daughter as well, which I did at the end, as it seemed this was important to them, but I said I wouldn't be able to do any recording of what she said as she wouldn't be able to sign a consent form [she is in the states]. She told me that mum and dad had been to the GP before she visited in October following concerns from mum, and that the main concern seemed to be around the delays, it took a long time to have a scan [i perceived this was a key milestone for them] and then a long time to get the results.

Interviews progressed fairly steadily; I think he could have gone on for a short while longer but not much more. I wouldn't have wanted to do many more tests with him, maybe just TUG.

Left around 13.15

2017 22-93-2017.

In a relatively deprived small town outside of the main city. Bungalows at approximately five minutes from the surgery or 10 minutes from the centre of town. A rainy wet drizzly day. Probably lower middleclass.

Arrived on time I went inside to see them both. they recognised me from the main clinic. Inside the house was relatively clean and tidy but somewhat smaller than the previous interview and certainly more tired and in terms of decoration. We both sat in the front room his wife joined us and it was a little warm inside as there was a hot radiator on wall.

We made small talk about the weather and started the interviews. Having changed the order of the topic guide I was able to talk first about the views and experiences of the practice followed by reflections on being diagnosed with dementia by GP. They were keen to share their experiences of GP's at the practice and my overriding sense was that transport was a key limiting factor for them they would struggle to get into the centre of town it would involve two buses and they were keen to see people close to home. They had a good relationship with a particular GP and would be happy for that person that they trusted to effectively tell them what to do. They haven't seen anyone them other than the team at the practice and the timeli team. After I turned off the recorder at the end he was telling me about how laptops and mobile phones have now been banned from airlines which is a very recent topic in the news yesterday .

Left after around one hour

2017 22-03-2017.

Second interview of the day. In an affluent suburb in the outer main part of the city, the weather had improved and although it was still raining and drizzly at times there were intermittent periods of sun. I should think middle - upper middleclass. I arrived in plenty of time despite it being the second interview and was able to wait outside in the car for 15 minutes before going in 15 minutes early.

The house was very large and set back from the road in a quiet spot with lovely birds outside which were a very pleasant to hear their bird song. The patient greeted me from the window when I rang on the doorbell as he was helping his wife to get dressed as she had recently had a hip operation. He showed me in and we sat down in the main living area with them on two armchairs and me on the sofa. They were happy for me to video them and we started the interview following the new topic guide.

They shared their experiences of dealing with a GP at the practice but my overriding sense here was that they were usually seeking to get the expert opinion of a specialist. In particular they had concerns that a nurse might not have sufficient training to deal with what they perceived as being a particularly difficult problem with the "mind" rather than a physical or medical issue. The wife shared concerns that she felt that there was something wrong but the cause for this had not yet been identified as he's been discharged from the memory service with a diagnosis of no dementia. They showed me a letter which had been provided to him from the memory service which was signed off by a senior nurse. It was not clear that he'd been assessed by a memory consultant and it looked as if the diagnosis of no dementia had been made in the context of him scoring higher on the Ace3 the second time he did it compared to the first time that he did it. As they both identified, this may have been due to a learning effect. The wife was concerned as she noticed particular issues with sequencing and thought that he would be likely to get a recipe wrong. They were both pleasant to talk to and I noticed that both the wife and the patient failed to complete the consent form correctly. The wife wants to have a go at doing one of the money questions and I allowed her to do this: I did not have the correct change for the money question neither had I for the first interview of today but was able to make up a similar question which tested the principle that I had wanted to examine. The patient was not keen to actually do the tests himself and so we talked through them and what it would mean to do them and how effective they were perceived as being as tests of the memory and brain functioning.

No additional information of relevance are important was provided after the tape recorder was turned off. I did not notice any signs of distress or discomfort when doing the specific tests which were provided i.e. MoCA and eurotest

I left after around 45 minutes

2017 28-03-2017.

First interview of the day, had been a sunny day with intermittent episodes of rain. Bungalow situated off the main road up a set of fairly steep steps on a small row of similar properties. In an economically middling area I should think. Perhaps lower - middle middleclass. Beautiful view of the [relatively calm] sea.

Patient greeted me at the door and showed me inside, where her husband was also waiting. House was well decorated and maintained, clean tidy and comfortable. They both sat on the sofa facing out to the sea and I sat on the floor facing towards them.

My main reflection on this interview was the extent to which the patient [lady] dominated the conversation while her partner sat relatively quietly to her side. Her main concern was about finding out what was going on and wanting an explanation for symptoms that she had noticed and perceived were abnormal. She was less concerned about who she saw - though she liked seeing a particular doctor at her own surgery - and more concerned about knowing what was

causing her symptoms, even if nothing could be done about them. She mentioned on a couple of occasions that her memory wasn't a particular concern for her.

She seemed to cope well with the interview and probably could have gone on longer but I perceived I had the information that I needed.

No additional information of relevance are important was provided after the tape recorder was turned off. I did not notice any signs of distress or discomfort when doing the specific tests which were provided i.e. MoCA and eurotest. My overriding sense was that there was no real difference in how these tests were perceived and that they would be happy to have had whatever test was most useful and helpful.

I left after around 45 minutes

2017 28-03-2017.

Second interview of the day. Had been a pleasant drive from X to X along the coast road. Arrived at the house in a residential area in town. Small garage outside that I was able to park outside. In a relatively economically depressed area for the town, I would think. Perhaps lower - middle middleclass. Patient was at home by herself as her friend had not yet arrived. Inside I set up and after 10 minutes or so asked if it might be worth phoning her friend to check she was coming. The friend arrived a few minutes after this as she had in fact just been pulling up outside.

I sat on the floor and they sat on chairs next to one another. She seemed to be doing fine with the interview but I noticed at around 20 minutes that she seemed to start to tire. The main issue that they both seemed to be expressing here was that they didn't feel she [the patient] was too impaired at the moment, so it was fine to be managed by a GP. There seemed to be a sense that seeing a specialist would imply a more serious or advanced problem and that might not come in the future.

No additional information of relevance are important was provided after the tape recorder was turned off. I did not notice any signs of distress or discomfort when doing the specific tests which were provided i.e. MoCA and eurotest. They commented that MoCA seemed to test a wider variety of domains than eurotest [though of course they may not have appreciated that in fact a large number of cognitive skills are tested with eurotest, even though it is relatively brief and appears simple].

I left after around 50 minutes.

2017 04-04-2017.

First interview of the day, four in total. Arrived in plenty of time, a very beautiful day outside. A detached house in a suburb, in a relatively affluent part of the [fairly middling] town I should think. So middle middleclass overall. She opened the door and let me in, she was at home by herself as she lives at home. Inside the house was notably clean and tidy, everything was put away and very clean. I took up her invitation to have a glass of water. I asked about recording the interview, which she was happy to do, and about videoing, which she was less happy to do, so we quickly agreed that we would not video.

I set up the kit and we made some small talk about the pleasant weather, her move from X some years ago and her desire to finish at around 30 minutes as she was going out to Lunch. At one point I asked her to repeat what she said as she had a noisy clock that chimed. This did not seem to be a problem. I sat on a sofa near to her and then moved next to her at the end to show her the tests. At her request we did not do the MoCA or Eurotest but I just asked her about the questions. She seemed to find the Eurotest and MoCa of similar difficulty and neither of them was particularly burdensome for her. I think she would have been able to have been interviewed for longer if necessary or time was available, but I think we covered everything that we needed

to do. No additional information of relevance was provided after the tape was switched off. At the end we made some small talk about her plans to go out and eat at a local canteen.

2017 04-04-2017.

Second interview of the day, four in total, a short drive from interview #1 of the day. I arrived and the husband let me in. I sat on the floor opposite them. House in an economically more constrained part of town than interview #1 of the day, but still middle middleclass overall. Inside the house was clean and tidy and they were both able to take part in the interview and were happy to be videoed. I declined the offer of tea / coffee.

They wanted to hear the outcome of the main clinic appointment as they still hadn't heard, I apologised that I wasn't able to tell them as I didn't have the information to hand, and that I could confirm that the information had been emailed back to the surgery 2 months ago. They were happy to take this up with the practice. Before the consent forms were signed and therefore before the tape was switched on she spent some time talking about how she had had a series of MRI scans for a ongoing problem at the base of her skull that was not yet fully evaluated or determined. This did not seem to be causing her particular distress and I was rather struck by the calm way that she talked about it; she said that her view was that if it was something serious or worrying then things would have moved along more quickly so it was likely to be nothing to worry about. I had no sense that either of the tests was especially burdensome or troublesome.

After the tape was switched off we had a conversation about how the tests hadn't really seemed to capture her problems and that whatever was shown up she remained sure that something was wrong, even if an explanation hadn't been found. I said that the problem was that for most GPs they might not refer her as she would be normal on the tests, my approach might have been to refer her as she was so intelligent it was likely that specialist input would be needed in any event.

2017 04-04-2017.

Third interview of the day in probably the most affluent part of town so far. The gardener was outside and the husband then came out to greet me. The house had a beautiful garden out the back. I went inside, everything was clean and well decorated, though perhaps a little tired. I took up the offer of a cup of tea. A range of wildlife DVDs on the side cabinet. Probably middle - upper middleclass. Wife seemed a bit apprehensive at first but definitely warmed up.

They talked to me about their experiences of the memory clinic and the appointments, and that it seemed to take an age to have and get the results of the CT scan. They were both interested to look over the memory tests together but I had no sense that either of the tests was especially burdensome or troublesome.

We went on a little longer than I had anticipated as they [especially he] seemed to go off track, at times, from the questions that I asked, and I didn't want to interrupt them too quickly.

Nothing of relevance was said after the tape was switched off.

2017 04-04-2017.

Last interview of the day in a similarly affluent suburb and in an outskirts residential location. Probably middle - upper middleclass. I parked outside on the road and walked up the drive. There was a big dog inside which I found a bit scary, but they quickly put it inside a room and locked it there. We sat in the conservatory with a view of the sunny garden. They sat in chairs next to each other, which we had to move around a bit. There had been some confusion / muddle with the time of the interview as he had thought I had said 11.30 though I had definitely

said - and written down - 1330. I had offered the opportunity to reschedule or rearrange or cancel but they were both happy to continue. I arrived on time.

I declined the offer of refreshments as I had had plenty.

They asked about the outcome of the main clinic visit as they had not yet had the letter. I apologised for this delay and said I was fairly sure that the letter would have gone back to the surgery but that I would check as soon as I got back to base, and that if the surgery said they still did not have the letter by the end of week then to let me know as it may have gone to their junk email.

I thought the interview was about the right length and any longer might have been difficult. I noticed that he seemed to struggle more with the Eurotest than with the MoCA but neither seemed especially burdensome. They did not seem to perceive that the Eurotest on its own would be sufficient to make a diagnosis of dementia. The dog was barking a little bit during the test but was locked away and this was not especially distracting.

After the tape was switched off we spoke briefly about how there had been an episode with nhs.net mail a few months ago - I couldn't remember the precise date - when the system had melted down and around 4000 emails were received by users in a single day.

2017 03-05-2017.

First interview of two today in a lower-middle >middle-middle class part of town. Small bungalows and residential location. I went on my bicycle as the car was in the garage. I arrived five minutes early and was just sorting out my bike when a middle aged woman arrived; I asked if she was the daughter but she was in fact the cleaner. I went in with her. Inside the house was quite disorganised but tidy and clean. There was a large amount of stuff spread out in most of the rooms with the exception of the front room which is where we sat to do the interview. On my way in the patient was talking to the cleaner and remarking on how she had forgotten that the cleaner was coming, again. She offered me a glass of water and I took her up on the offer. The cleaners mobile phone rang a few times during the interview, though she was in the kitchen it was still possible for me to hear her talking and answering the phone, but this was not especially distracting and I don't think the patient could hear it.

We sat in the front room which was clean tidy and well kept. There was a large picture of a house on the facing wall that the lady sat in and I asked about this: it emerged that it was her former house in Saltford - quite a large [10 bedrooms she said] house that she had formerly run as a hotel with her daughter. I was interviewing her by herself. As we made a start in the interview she seemed to forget something she had said earlier on and to be struggling with some of the questions more than I had anticipated. I made sure to focus on the key areas that I wanted to get her views on. She did not seem to think she had any problem with her memory and so I chose not to ask her about when her memory last let her down and about her originally consulting her GP about her memory, as I did not want to cause her distress, especially as she was on her own. [I note, on return to the university, that the impression at her main clinic appointment was normal]. She was very clear that if there was a problem with her memory at all she would want to see a top specialist and would be happy to pay for this if necessary to go to London. After the tape was switched off she told me a bit more about her former house. As I went out the door she remarked that she had thought my bicycle had a light on it, I explained that it did but it was not visible as it folded up.

2017 03-05-2017.

Second interview of two today, I cycled between the interviews on my bicycle as the car was in the garage. Relatively small but modern house in a fairly deprived part of town, probably lower-middle to working social class. I arrived 30 minutes early but they were happy to start sooner than we had planned. I folded the bicycle up and went in, the patient opened the door to me as

his wife was upstairs. They were both happy to take part and bought me a glass of water to drink. I sat on the floor in front of them.

The house was clean and tidy but not large or spacious. There were a number of photos in the house of their grandchildren and I remarked on one of X which he told me he had taken himself. We got on with the interview. I got the impression that the wife was more verbose and dominating in the conversation but I tried to encourage him to speak and share his views as well.

Unfortunately the video recording did not work. I tried - and failed - to connect via the internet using 3G on my phone but it was no good. At the end we spoke briefly about the garden and range of bird feeders that they had outside.

2017 09-05-2017.

Only one interview today, but the end of the day after a research clinic. In a beautiful period [? Victorian] house on a quiet road near the edge of town, though not in an especially wealthy looking area. I would say middle middle class. There was a large, handsome, dog outside but they kindly kept it there without me having to say that I'm quite scared! We went it and got straight on with the interview, they had been outside in the garden as it was a beautiful day. Weather was absolutely splendid.

I sat on the floor in front of them in the sitting room and they both participated. It was interesting to hear their views about getting access to a specialist. The wife seemed to repeatedly emphasise the sense of being dropped or abandoned to get on with it and seemed to value having someone that she could talk to and access easily. But expertise also seemed to be perceived as being very important too.

They closed the encounter by emphasising that they would be happy to be contacted again in the future.

2017 19-05-2017.

A beautiful sunny day. I was wearing shorts it was so hot. First interview of the day in a small bungalow in a town outside of the main city. He greeted me at the door, wearing clothes that looked as if he has been painting. Inside the house looked as though it was being redecorated, but was comfortable and homely. There was a large open plan living area with a small garden out the back. Lots of painting from the middle east on the walls, including one which he said he had painted himself. He said a number of times that his financial situation was "not good". My sense was that this was probably a lower-middle class house though as he referred to getting some money from rental income, but also that he needed his daughter to support him financially at times.

I sat opposite him, there were just the two of us there as his daughter was not available. He told me about his time as a X of X.

At the end, after the microphone was turned off, he told me about an X restaurant in X and showed me the picture in a book that he had copied the wall painting from.

2017 19-05-2017.

Second interview of the day, I drove straight between the two. This was with a X X, and I knew he was a XX which was one of the reasons for asking him to take part in interviews. I arrived and he greeted me at the door. There were two fairly old, "unfashionable" cars on the drive. There was a large garden outside. He remembered that I was a GP: mostly participants have not remembered this though they may recall having seen me at the main clinic. He wanted to show me a series of photographs from X, which seemed important for him to do and so I let him do this. We had not signed the consent forms yet. He showed me X and told me about how X. There was a large photo album which we looked through together while he talked about this for about 10 minutes. We then signed the consent forms and made a start, his wife was happy to take part

as well.

This interview was interesting because it struck me that he was keen to return to talking about X and it was a delicate balance to try to move him back to answering the questions that I had for him. He was keen to talk and I tried to make space to get his wife's view as well.

At the end his wife offered me some coffee, but I was not able to stay as I had to collect my daughter from preschool, so she gave me some poppadoms to take away.

2017 13-06-2017.

First interview of the day, four in total. It had been a someone hectic morning having got up before 6am to do data analysis prior to a meeting with supervisors, driven to that meeting and then back to X to get the stuff for the interviews. A very beautiful day. A semi rural area but a small bungalow that looked like it might have been part of a housing association at one point. The lady greeted me at the door and showed me inside, she was at home by herself and lived alone. She seemed to be managing OK thought things were not spotless inside for example the carpet was a little dirty in places and there were some old bits of paper around. Though she described herself as having gone to a private school she told me afterwards that her father was a tenant farmer on a council farm. Certainly on the basis of the home as it was presented at the time I visited I would describe her situation as working class.

She spoke about her farming background and how she was a country girl. She told me that she had lifelong problems with her hearing but this did not seem a particular problem during the encounter. I deliberately kept the interview fairly brief as I did not want to overburden her and was especially mindful of this as she was home by herself and struck me as fairly frail physically, regardless of her cognitive state.

2017 13-06-2017

Second interview of four today. In a modern detached house on a modern estate. The daughter was present for our encounter. I would describe as middle-middle class. She told me she had been X for her main job since her thirties.

Inside the house was very clean and tidy and well maintained. They offered me some water which I was happy to have as by this point I was quite thirsty. There was a small conservatory outside with a view onto the well maintained garden with water feature.

The interview progressed fairly well. Of note she lip read.

At the end I explained I might send out a postcard to all participants at some point in the future to invite them to a public meeting to share the results of TIMeLI with them, though this may not be for a few months or even a year or so.

2017 13-06-2017.

Third interview of four today. The timetable was progressing well and to time, and my journeys between sites were not burdensome as I had planned well geographically. Things seemed to be going OK and I was not feeling hassled or stressed.

I arrived and it was slightly difficult to find the house as it was on a lane that broke in the middle with a no through road due to it being so narrow; it was possible though to locate the house without a great deal of difficulty.

A modest size bungalow that was clean and tidy inside and decorated in a modern fashion. All on one level as I think the patient's wife has some mobility issues [there was a frame inside to help her get around]. A well maintained garden outside and a clear interest in plants. I would describe as lower-middle middle class.

The conversation progressed well and we established I felt a good rapport. At the end he asked about what the next steps were and I explained that he should contact his own GP surgery to check that they had received our correspondence from the main clinic and they would then be able to progress things as needed from there. This seemed to cover things.

At the end I explained I might send out a postcard to all participants at some point in the future to invite them to a public meeting to share the results of TIMeLI with them, though this may not be for a few months or even a year or so.

2017 13-06-2017.

Last interview of four today. He greeted me outside the front door, there were some very old but elegant classic cars parked on the drive outside. A fairly large house in a suburb, but not flashy or grand. Inside well decorated though tired and with some clutter around. The back garden looked reasonably well maintained. I would say middle middle class.There was a variety of calligraphy in the home that he had done himself.

His wife joined us for the encounter. The challenge here was that he didn't seem to have any insight at all that he had a memory problem. The interview seemed to progress well.

Afterwards, outside, the wife rushed out to X This seemed to be acceptable to the wife and we parted on a positive note.

At the end I explained I might send out a postcard to all participants at some point in the future to invite them to a public meeting to share the results of TIMeLI with them, though this may not be for a few months or even a year or so.

2017 14-06-2017.

First interview of two today. I had a fairly hectic morning having dropped my daughter at childcare and then tried to rush to be in the office for an early morning phone call. The house was further away than I anticipated and sat nav had directed me down a wrong road but I arrived unflustered. I walked up to the gate as it was difficult to park outside, and he greeted me at the gate. A beautiful sunny day. It was a modest bungalow in a small rural conurbation just outside the main town. Inside relatively well maintained and clean and tidy. I would describe as lower-middle class.

I thought there were two interesting things about this interview. Firstly he made the point that if he had to get in the car anyway to see a doctor it didn't really matter how far he had to go so seeing a GP or nurse locally wasn't really adding anything to him going to X to see a specialist. Though he could see some merit in this if the appointment wait was less, and wouldn't in principle have a problem with seeing a nurse for a diagnosis, providing they were appropriately trained, it sounded as if travelling to X instead of for example X was not really a priority for him. He mentioned his practice is now merged and he implied he may have to travel to X on occasions to see his GP, though whether this has had any impact on his views about the acceptability of travelling for healthcare or whether he has always found it acceptable to travel where needed was not clear and I did not especially want to explore this in great detail as it was not the main focus of the interview. Additionally the waiting time to see someone was not regarded as being especially important.

The other point of note was that he related the memory to his recent X problem. He said that he had been given exercises for his X and it would be nice to have something to do to try to prevent the decline of his memory perhaps some brain exercises or something like that.

At the end I explained I might send out a postcard to all participants at some point in the future to invite them to a public meeting to share the results of TIMeLI with them, though this may not be for a few months or even a year or so.

2017 14-06-2017.

Second interview of two today. He was waiting outside his house when I arrived and greeted me, I parked on the drive. He was home alone as his wife was out for the day. The house was a little dated in decoration but was fair sized. I would describe as lower-middle class.

I took the opportunity to follow up on some of the ideas that the previous interview of the day had raised. To what extent was it important to see someone quickly, to see a consultant, was there any value in seeing someone close to home, and were "brain exercises" perceived as useful. I have modified my interview topic guide as a result. The interview seemed to progress well. We closed by him telling me that he was planning to spend the afternoon in the garden doing some work and trimming a hedge.

At the end I explained I might send out a postcard to all participants at some point in the future to invite them to a public meeting to share the results of TIMeLI with them, though this may not be for a few months or even a year or so.

2017 21-06-2017

Only one interview today: and a very hot day at that. Traffic in X in the hot weather was not good and I arrived about 20 minutes late, but I had called on the way and they were fine when I arrived. I bungalow in a semi-rural suburb I would describe as middle middle class. They both greeted me at the door. Inside things were very clean and tidy and smartly / contemporarily decorated. We made some small talk about the weather as I didn't want them to be nervous and I had a slight sense they might be. His wife joined us for the interview and bought me a glass of water as she had one.

I sat on the floor in front of them and they sat on the sofa together.

An interesting thing from this interview was that he seemed to think there was something particular about the mind that made it difficult to diagnose as a GP. I recalled this as being something that has come up before so I tried to explore it a bit further. We also explored the idea and consequences of travelling further to see a more expert specialist, or whether it didn't really matter how far a distance was travelled because they had a car anyway. Though his wife recognised that the situation may change if it was more difficult for them to drive in the future.

At the end we spoke about the next steps and I outlined that the GP would take things forward. They showed me a letter from the memory service and I explained that we could send on some of our information to them *if* he requested me to do that because we took his data protection very seriously. I explained I would wait to hear from him, mentioned the possibility of further contact with a postcard to explain the results in the future, and wished the best.

2017 28-07-2017.

Only one interview today. I arrived on time and went in. She lived in a large modern retirement complex of warden monitored flats in a small town outside of the main city. Inside the flat was small but well maintained and clean and tidy with modern fittings. There was quite a lot of stuff in a small space. An entrance hall, open plan kitchen – sitting area, bathroom and bedroom.

She couldn't get her head around how she had ending up taking part in the study. My impression as the interview went on was that she was modestly impaired cognitively and so I tried to keep the interview as short as possible. It was interesting though some things that she was able to remember.

2017 01-08-2017.

First of four interviews todays. I arrived on time. A reasonably large house on a hill in the older part of the town with a large vegetable patch at the front and a large garden at the back. We conducted the interview in the new modern wood framed conservatory with beautiful views across the estuary. I would say definitely middle middle-class. His wife bought us coffee and biscuits which I was pleased to have and we all sat down together. He was a bit tearful as we

started the interview, my impression was that this was because he was concerned about the impact that his symptoms had on his wife.

He spoke well and I did not need to ask his wife much. He had the letter that we had sent to his surgery printed out and in his file, which he showed to me. He was keen to know what could be done to try to prevent his memory getting worse and he shared the activities that he did with me, such as trying to stay active, going the gym and for walks. He told how he was used to being in a leading role and found it difficult when in situations he perceived he was no longer able to do that, for example following a map when leading people on a walk.

He liked the idea of having exercises that might help so that he could retain some degree of control over things.

2017 01-08-2017.

Second of four interviews today I drove from nearby and arrived in time having grabbed a bite to eat on the way. Outside was a Tesla. They lived in a large bungalow in the middle of a very rural area down a lane with only a few houses on it. Difficult to gauge the socioeconomic circumstances as the house itself did not seem especially large or prosperous to the extent that the car would. He greeted me at the door and showed me inside. We sat at a small table in the living room and all spoke together.

2017 01-08-2017.

Third of four interviews. A small house in the centre of X, the house was organised and tidy. Three of us spoke together, me, the husband and the patient. She seemed to have at least modest impairment and would answer in short answers and seemed to struggle to answer the more abstract questions. They had moved to the area from elsewhere, in the recent past. Their current house was in a less affluent part of the town. A fire alarm was ticking for most of the interview. Husband wanted to talk and said more at length. She had a bit of a cold and a cough.

2017 01-08-2017.

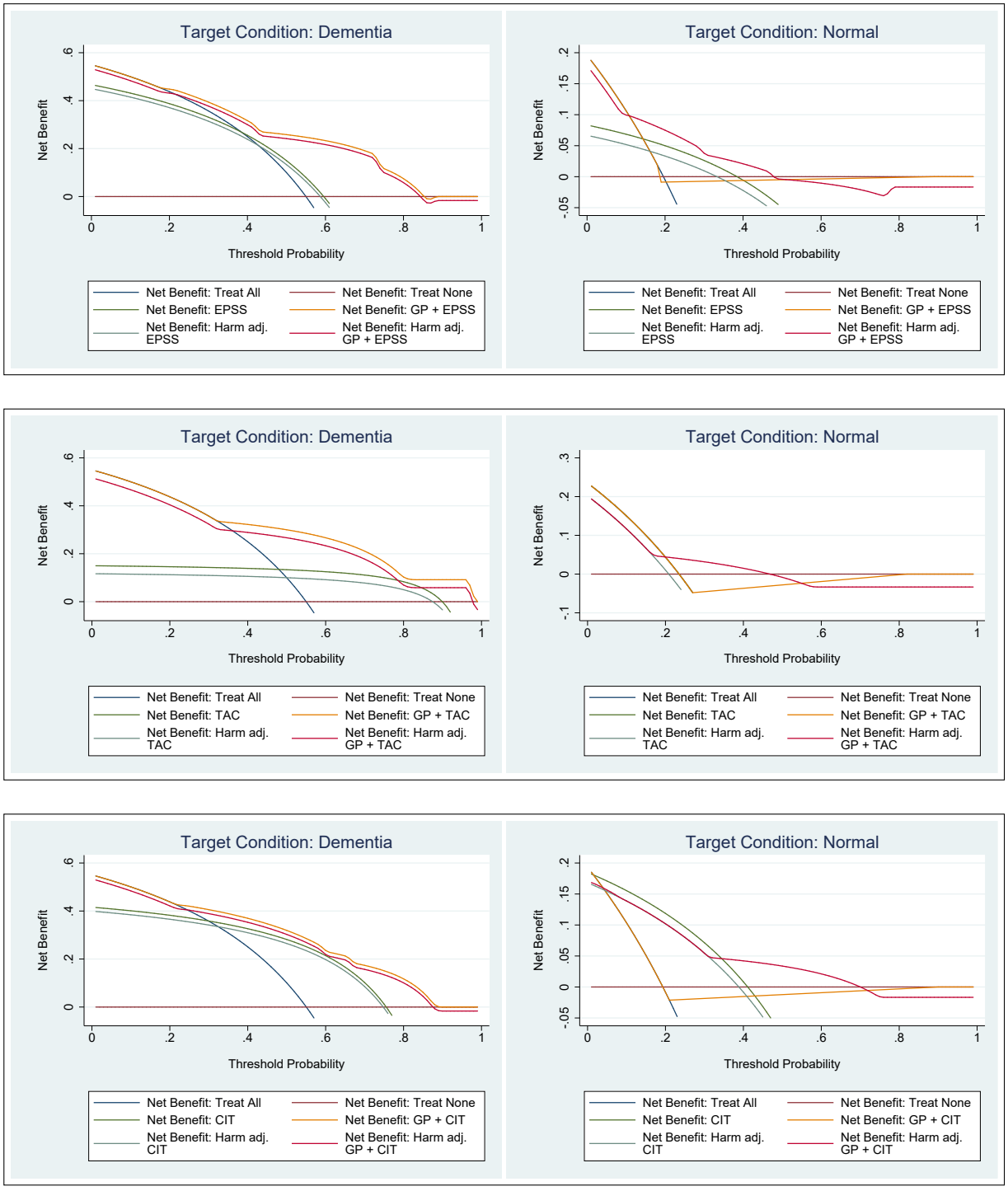
Last interview of the day. A house in a hamlet down a country lane, approximately a mile from the main town. There was a garage, but no car parked outside. Fields out the back of the house. I saw her with her husband. The house was tidy and well organised, and clean. Would appear to be middle class regarding socioeconomic status but feels a bit isolated from the main town. She seemed to speak clearly and articulately in response to the questions. Could articulate reasons for things. We sat at a small table in the kitchen and spoke together, but generally she was able to answer questions without needing input from her husband. The extent to which people talk in the community about health issues was important to them both.

APPENDIX H

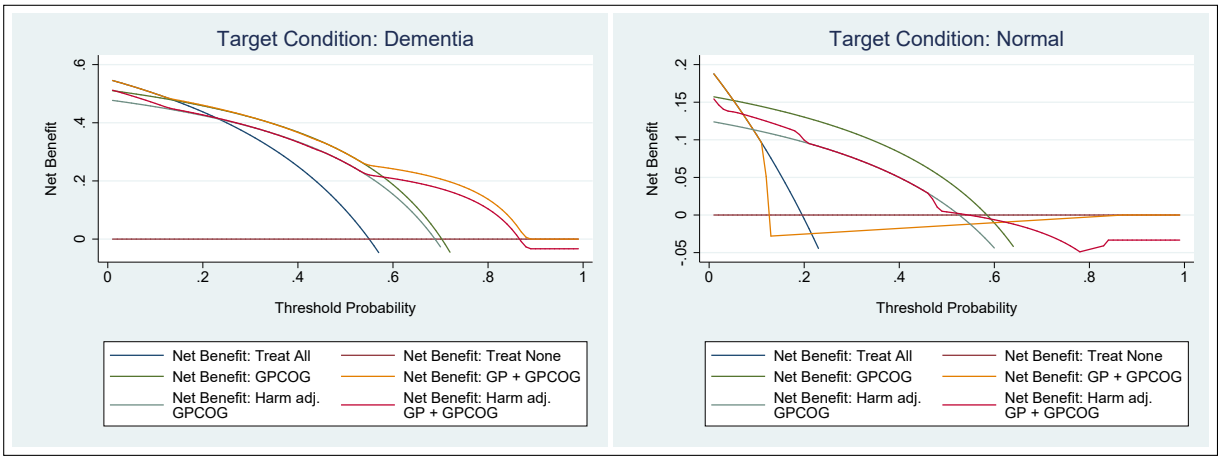
DECISION CURVE ANALYSIS

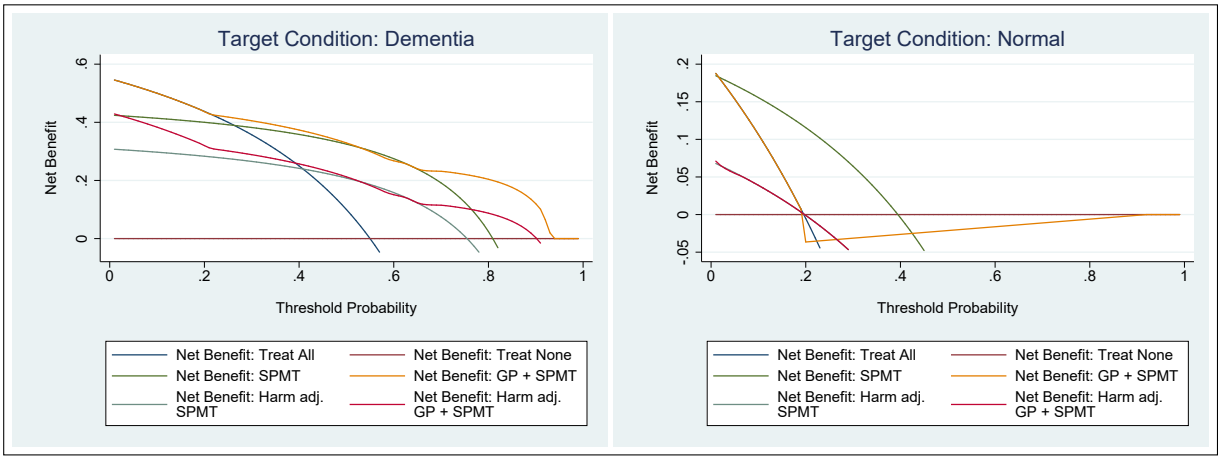
This appendix provides the naïve and harm adjusted decision curves for each of the individual index tests, for both of the target conditions, that are referred to in Section 3.7.4.1 and Section 4.6.4.

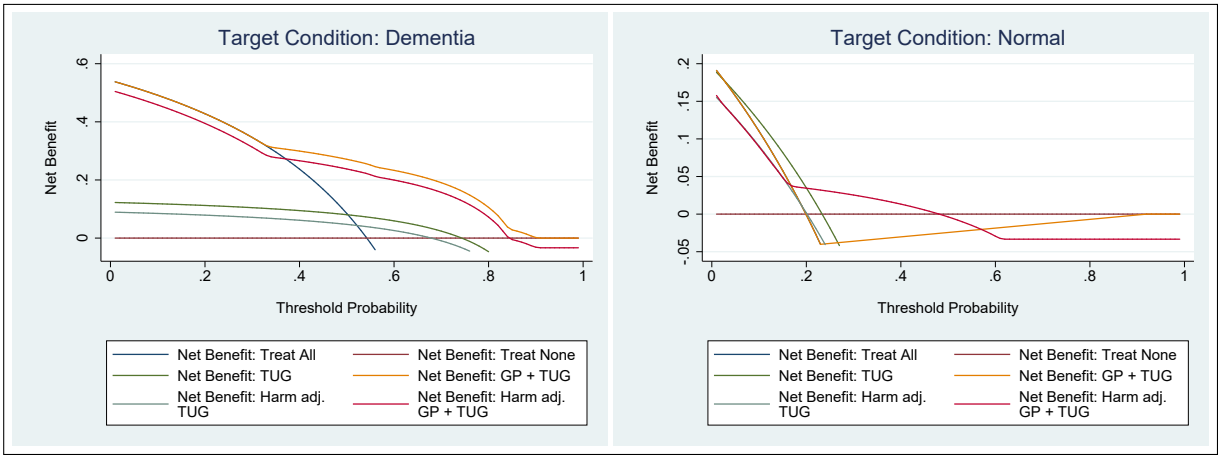
Figure H.1: Naïve and harm adjusted decision curves, with dementia as the target condition on the left, and normal as the target condition on the right



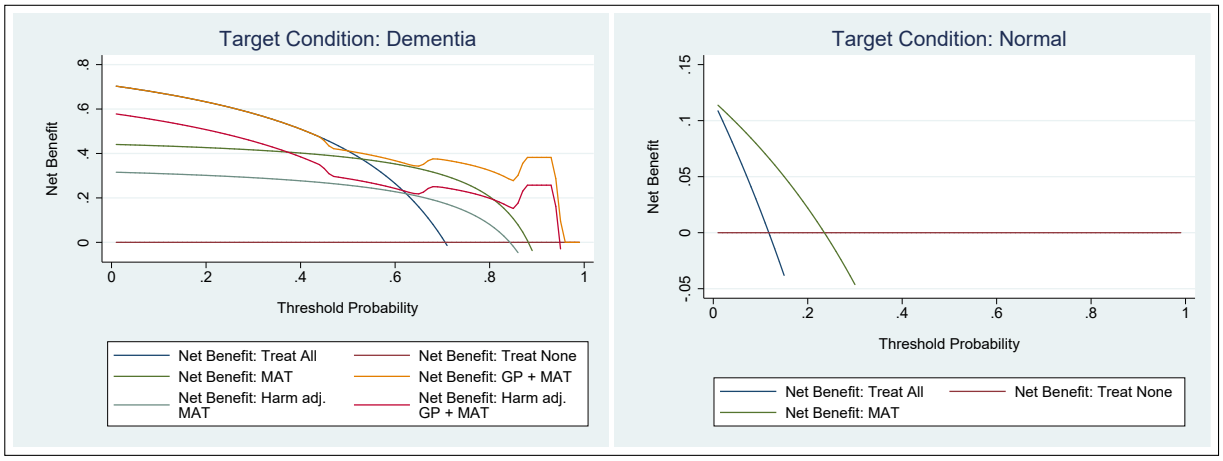
The curves plot the net benefit across all thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]



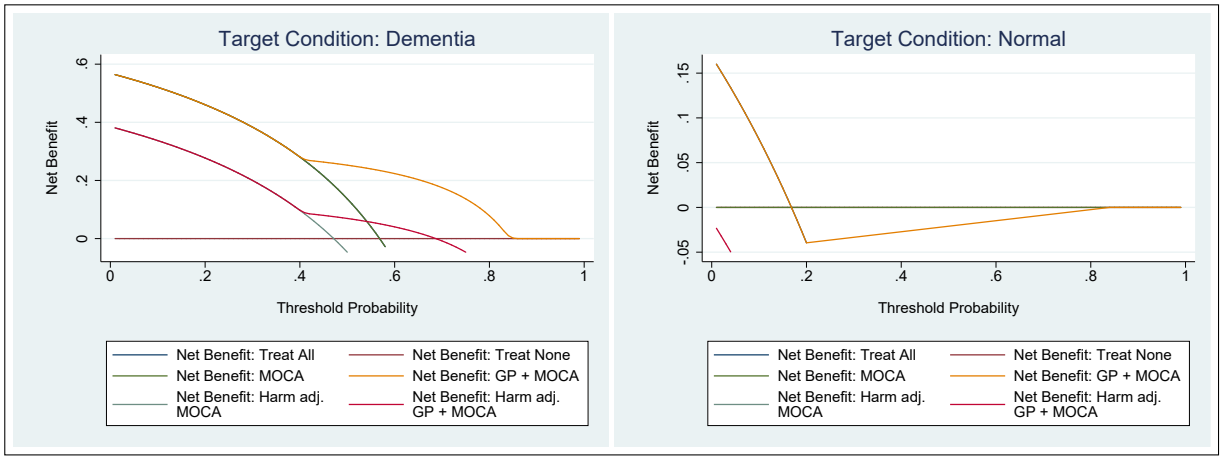


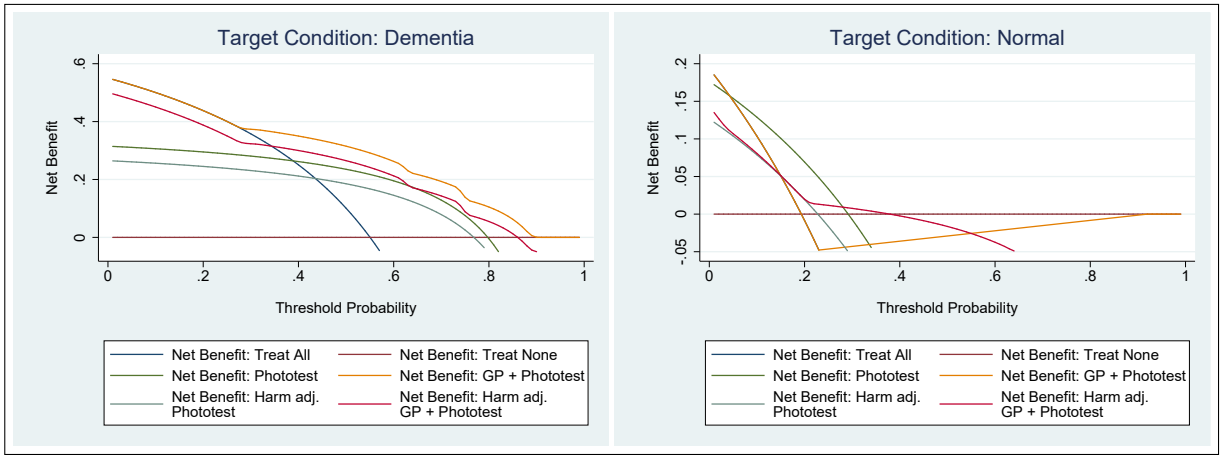


The curves plot the net benefit across all thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]

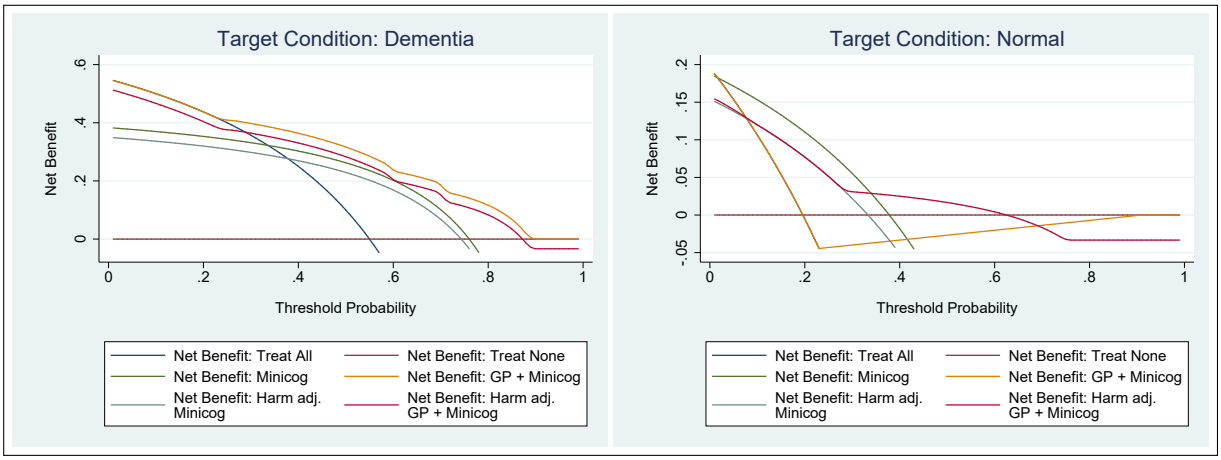


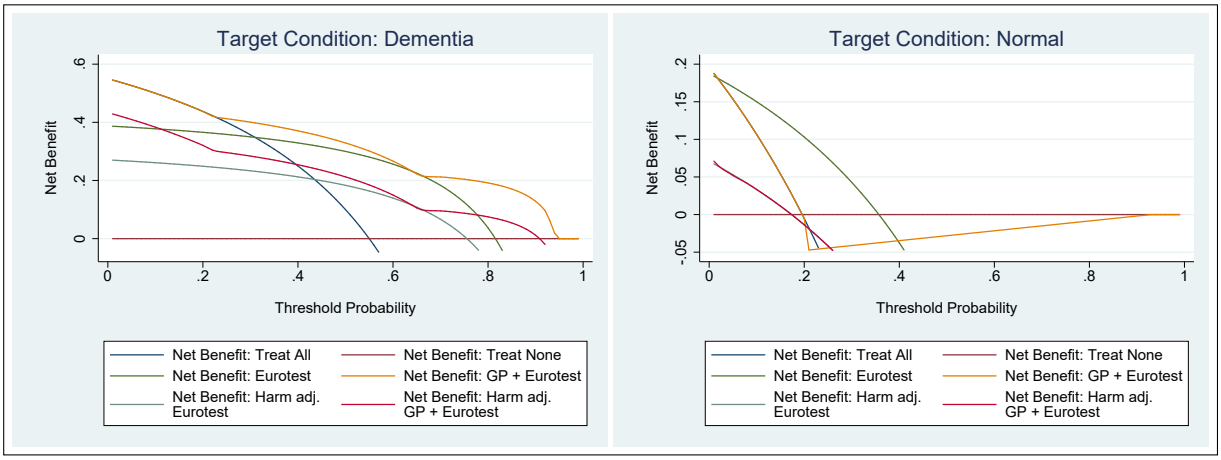
Small numbers make it impossible to plot a curve for GP judgement + MAT for the target condition normal

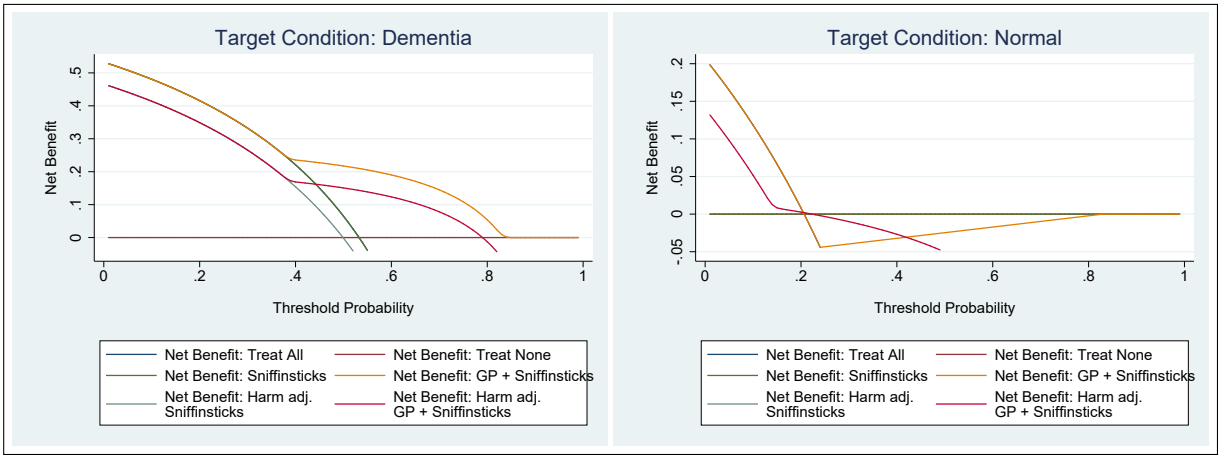




The curves plot the net benefit across all thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]

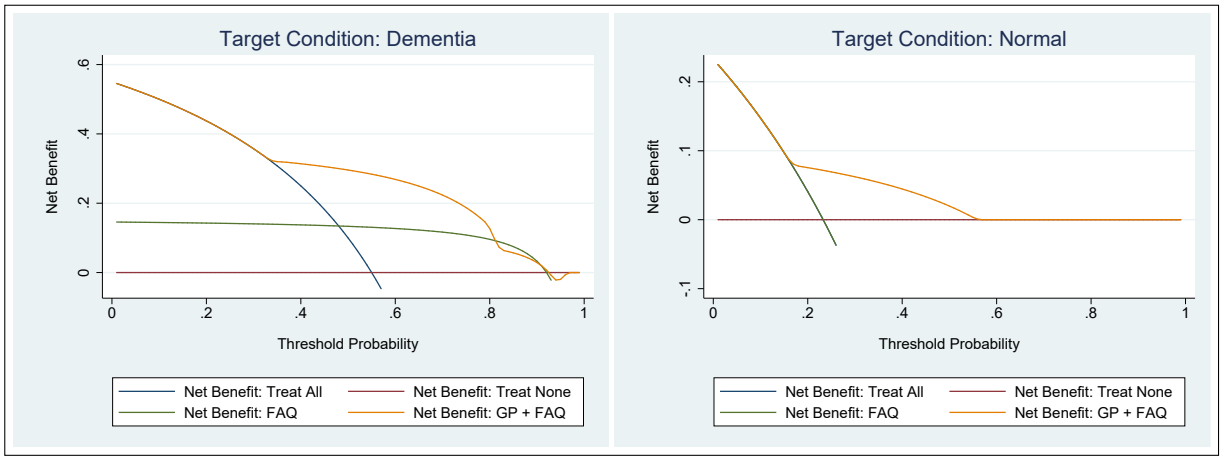


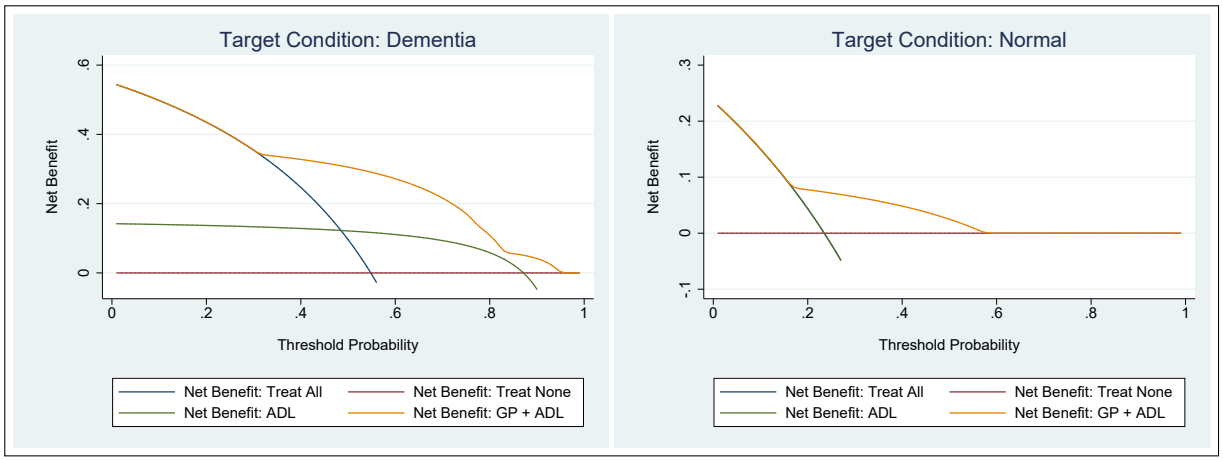


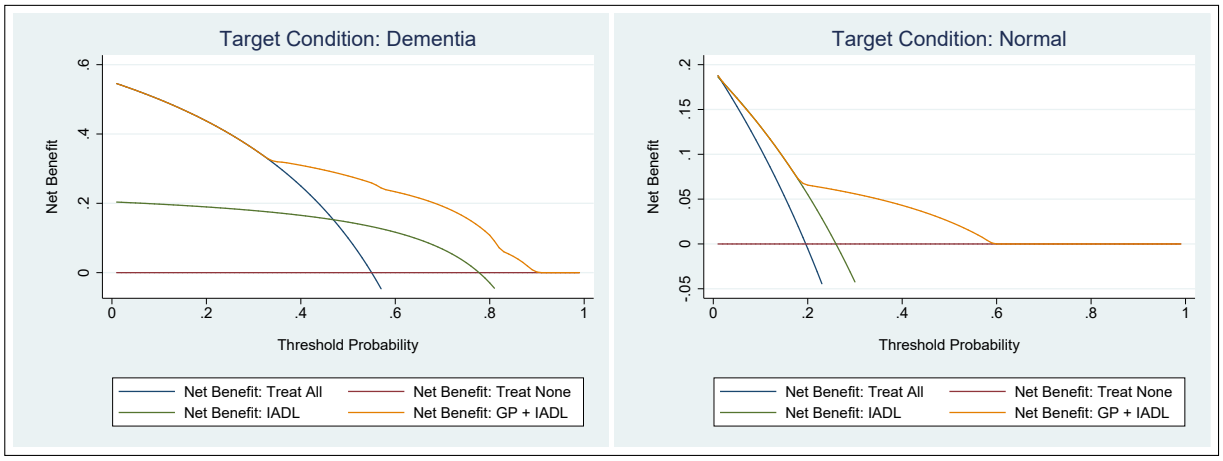


The curves plot the net benefit across all thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]

For informant tests no harm line is drawn

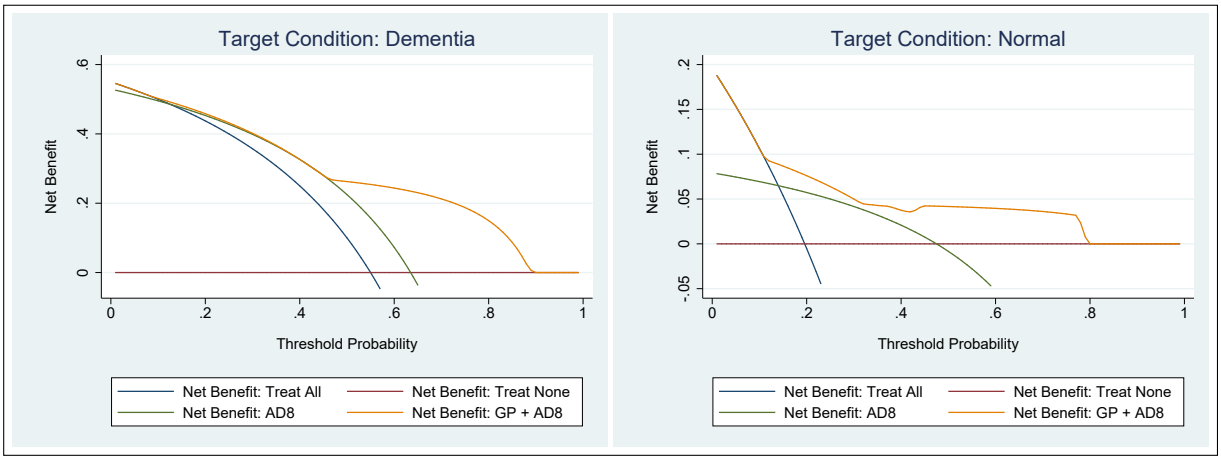


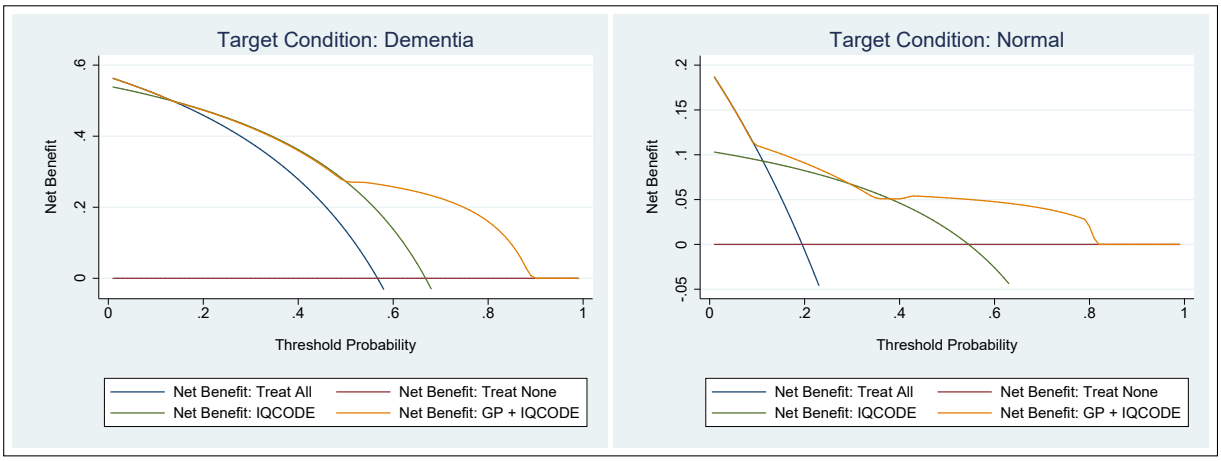


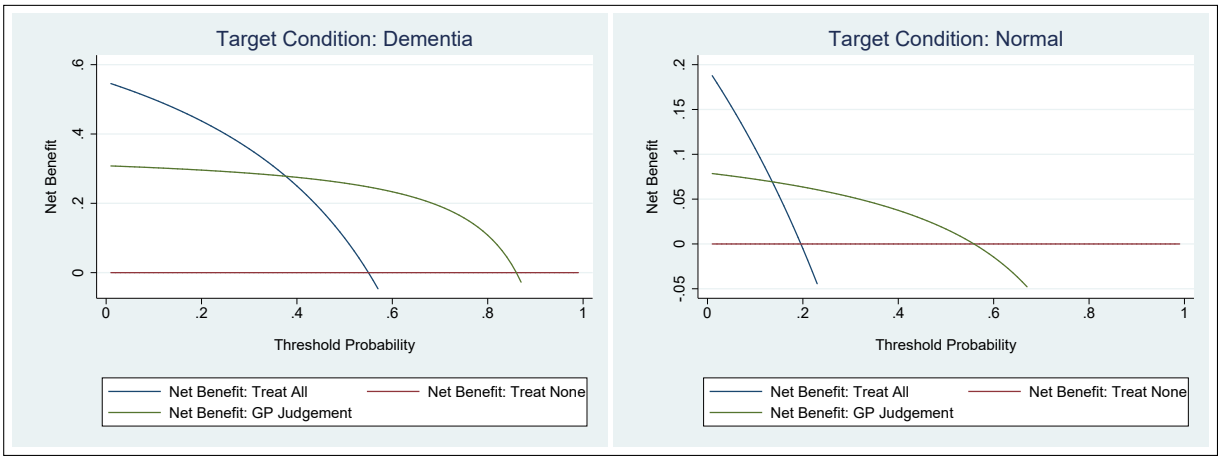


The curves plot the net benefit across all thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]

For informant tests no harm line is drawn







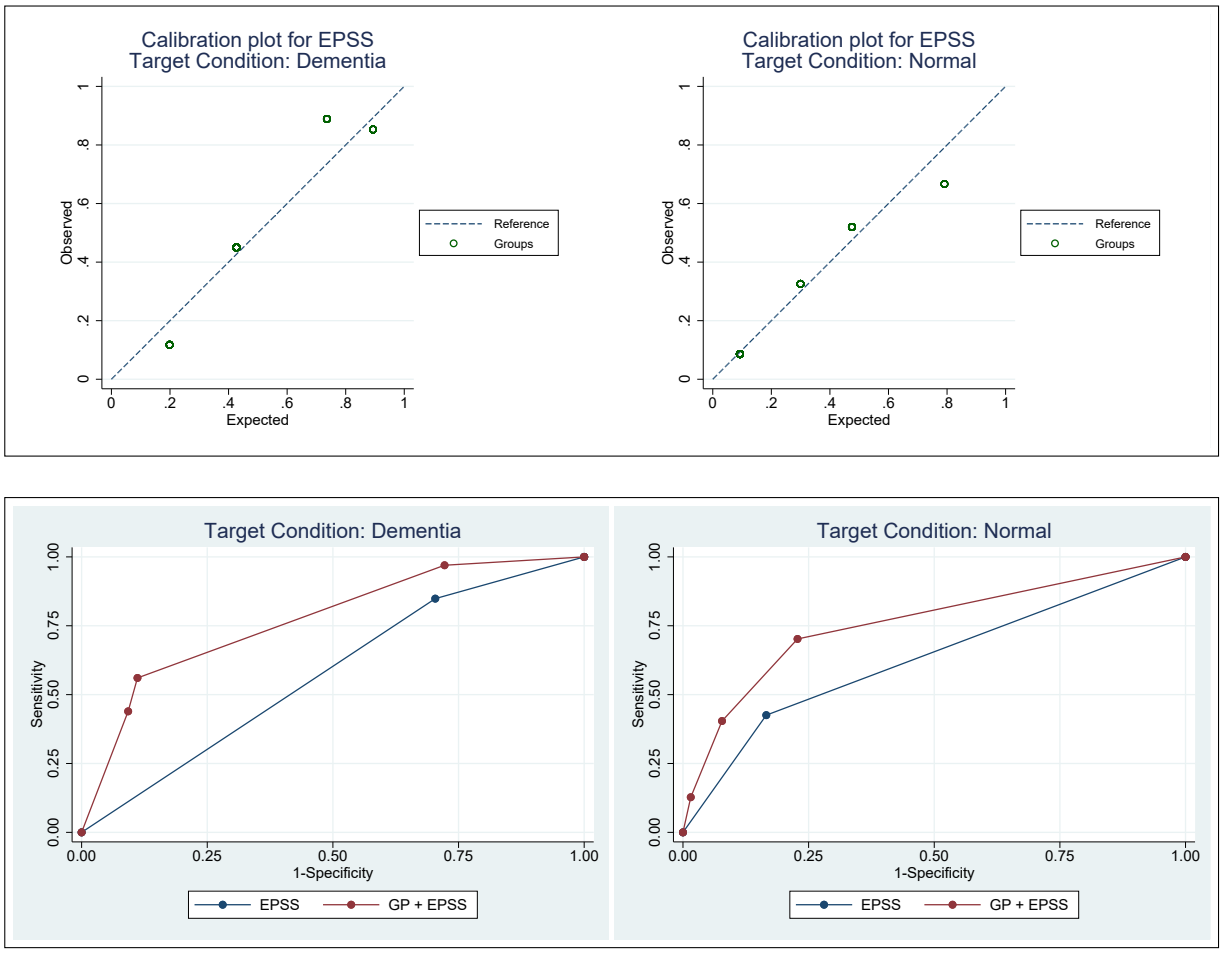
The curves plot the net benefit across all thresholds, along with the treat-all and treat-none lines. Curves are smoothed, as recommended to minimise unstable results [168]

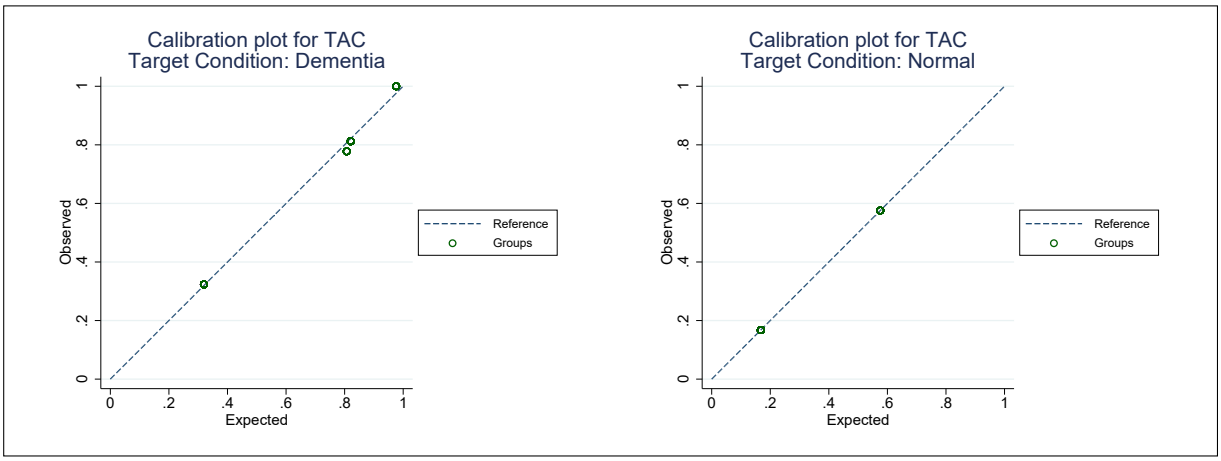
APPENDIX I

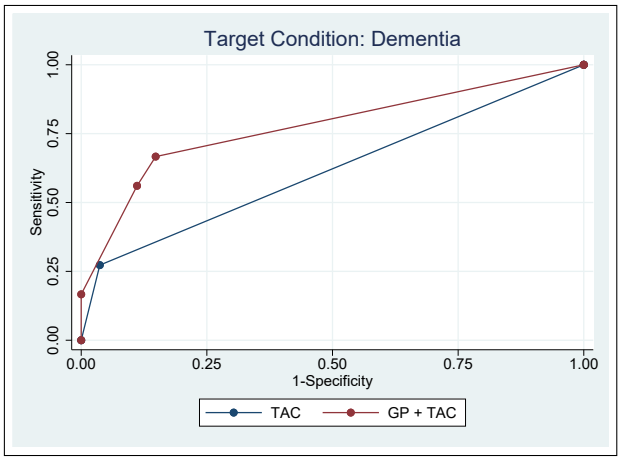
DISCRIMINATION AND CALIBRATION

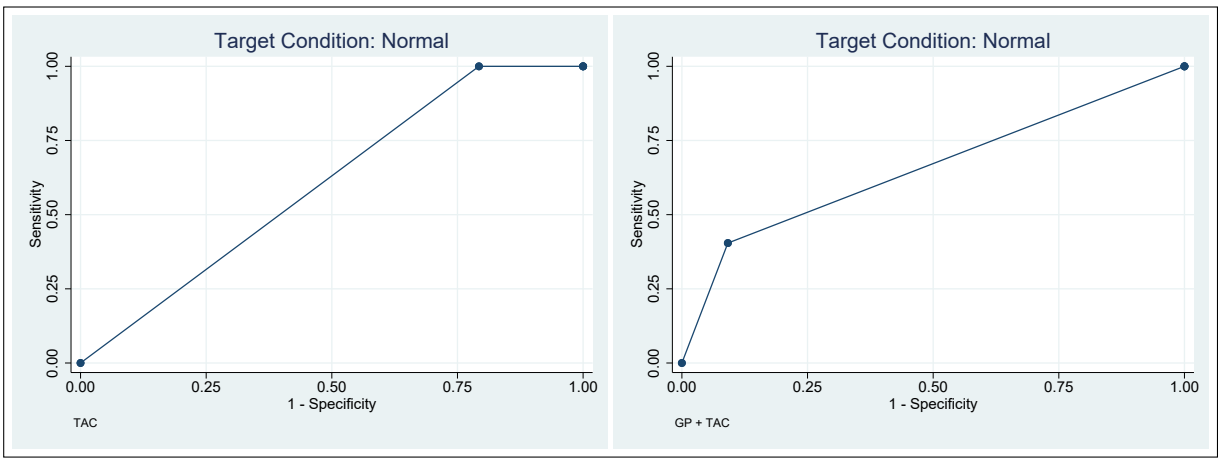
This appendix provides the ROC curves and calibration plots for each of the individual index tests, and GP combined tests, for both of the target conditions, that are referred to in Section 3.7.4.1 and Section 4.6.4. Plots for MAT are not provided because as described in Section 4.6.4 the small numbers who completed this test made the analyses unstable. For dementia as the target condition it was not possible to use `roccomp` to plot both the single test and the GP combined test on the same axis for index tests Sniffin sticks and MOCA, due to perfect sensitivity in the single test. The same applied for normal and TAC, FAQ and ADL. For these tests where it was not possible to plot both single tests and GP combined tests using `roccomp` separate ROC plots were made for the single test and GP combined test using `roctab`.

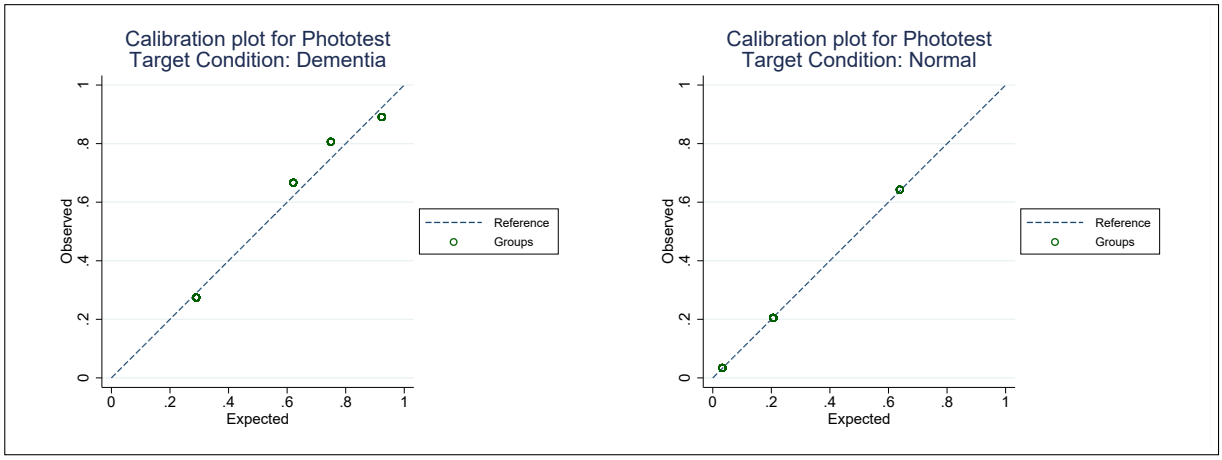
Figure I.1: ROC curves and calibration plots for the individual index tests and GP combined tests, with dementia as the target condition on the left and normal as the target condition on the right.

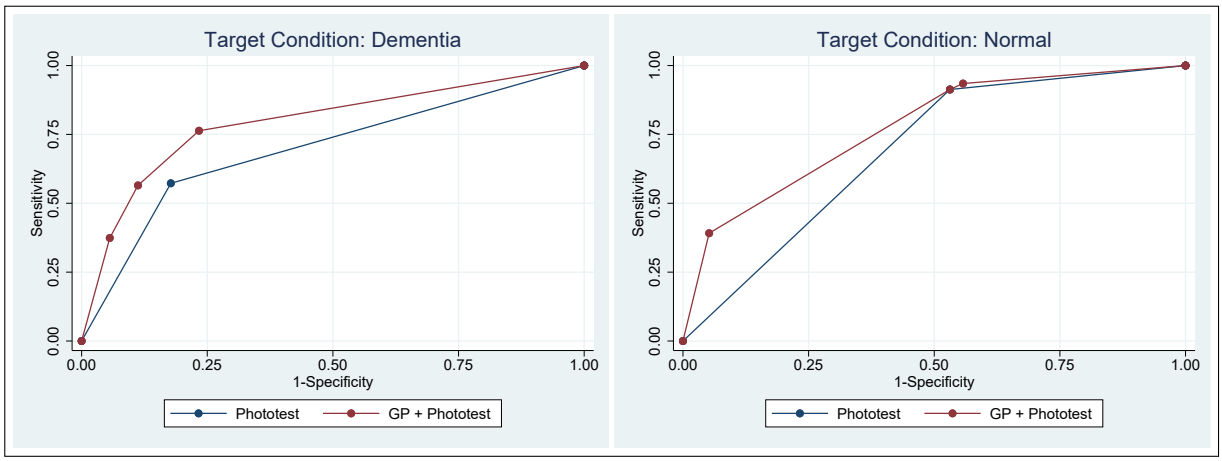


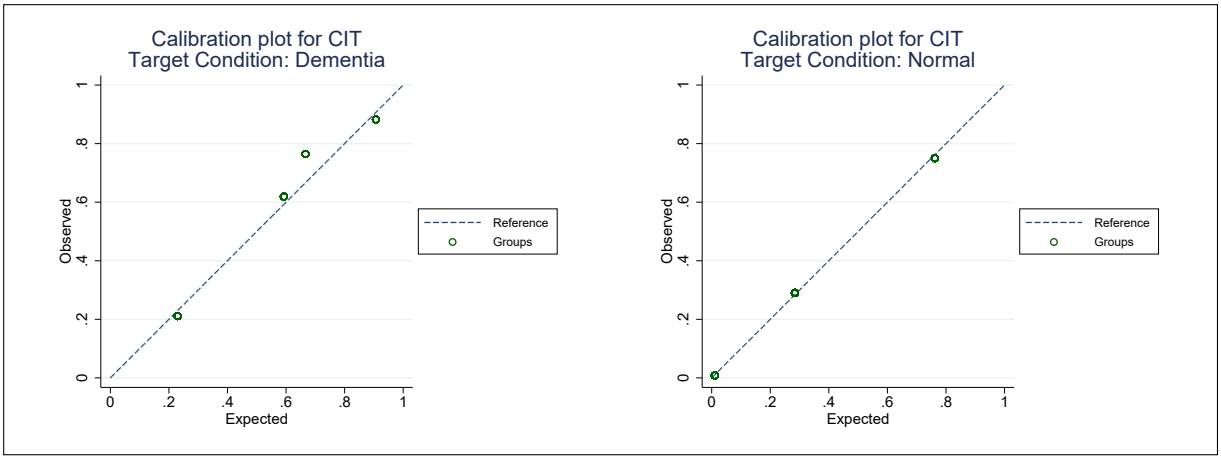

Calibration plot for TAC
Target Condition: Dementia
Observed
Expected
Reference
Groups
Calibration plot for TAC
Target Condition: Normal
Observed
Expected
Reference
Groups

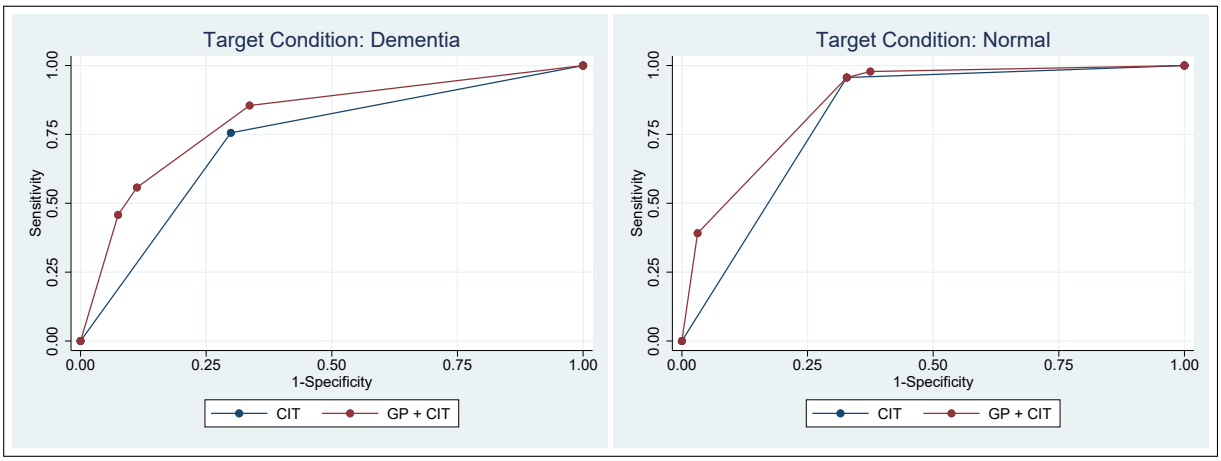

Target Condition: Dementia
Sensitivity
1-Specificity
TAC
GP + TAC

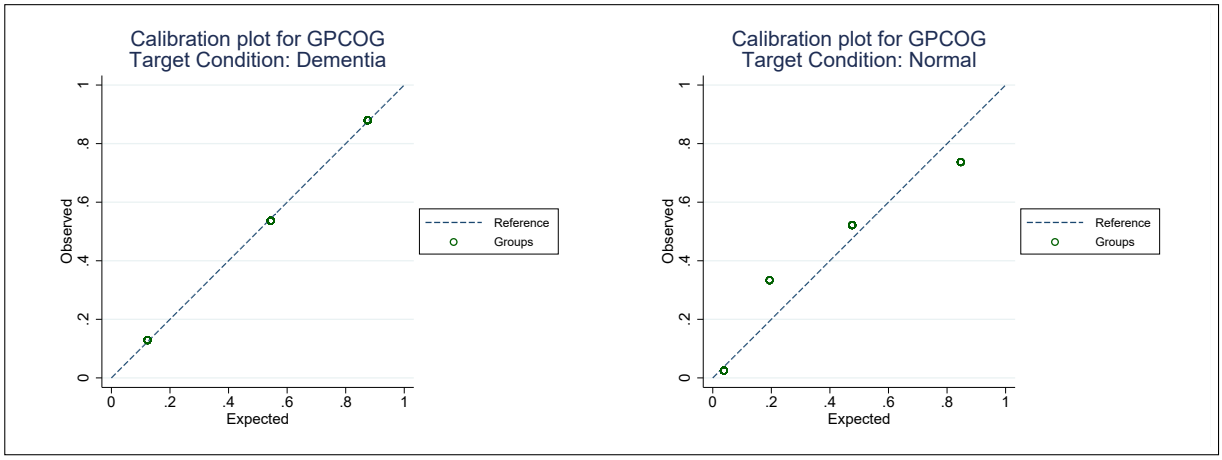

Target Condition: Normal
Sensitivity
1 - Specificity
TAC
Target Condition: Normal
Sensitivity
1 - Specificity
GP + TAC

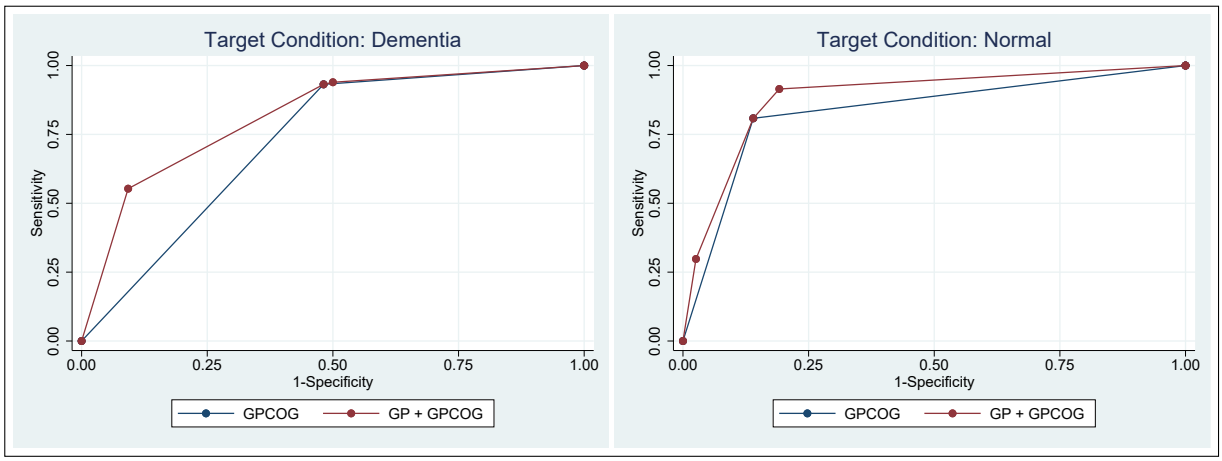

Calibration plot for Phototest
Target Condition: Dementia
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups
Calibration plot for Phototest
Target Condition: Normal
Observed
Expected
Reference
Groups

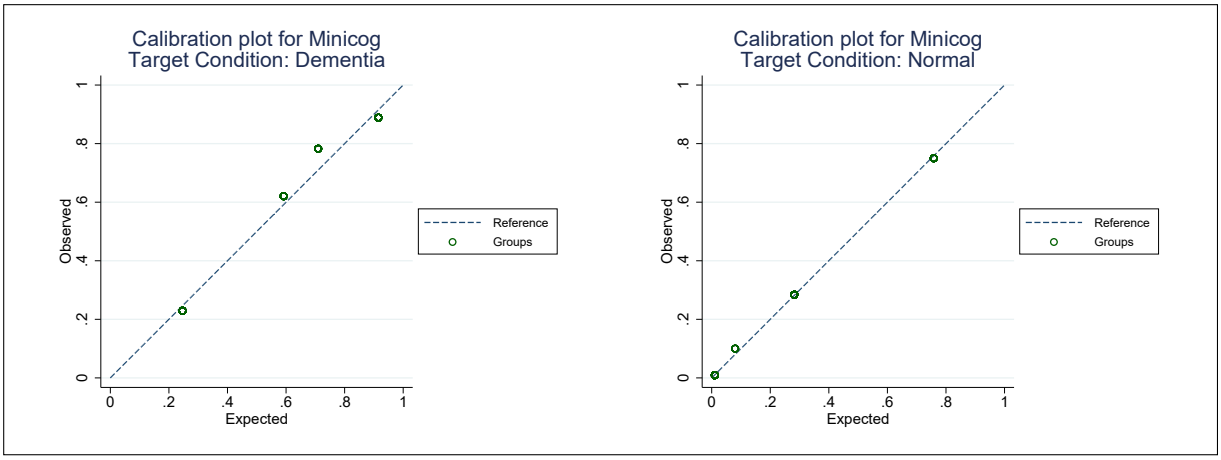

Target Condition: Dementia
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
Phototest
GP + Phototest
Target Condition: Normal
Sensitivity
1-Specificity
Phototest
GP + Phototest

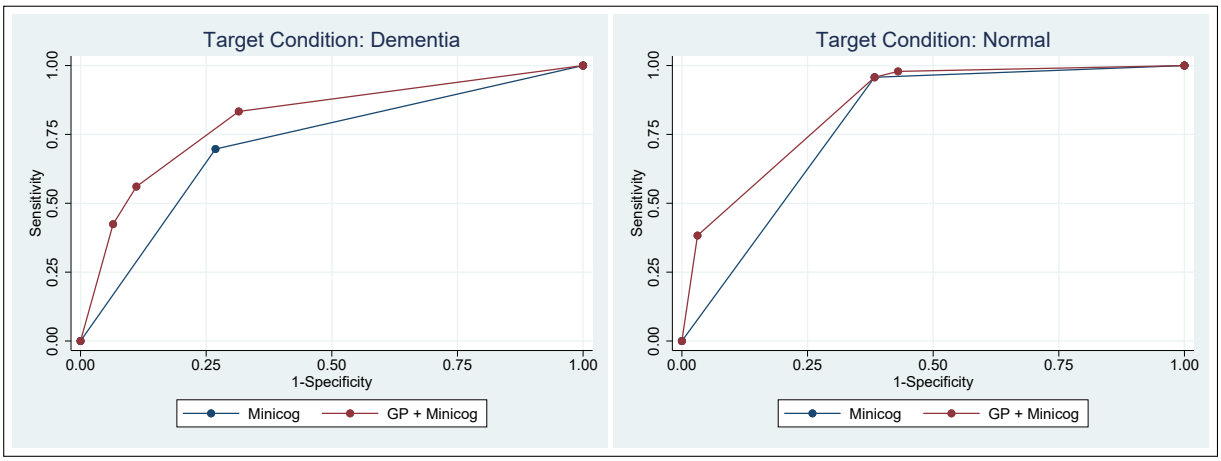

Calibration plot for CIT
Target Condition: Dementia
Observed
Expected
Reference
Groups
Calibration plot for CIT
Target Condition: Normal
Observed
Expected
Reference
Groups

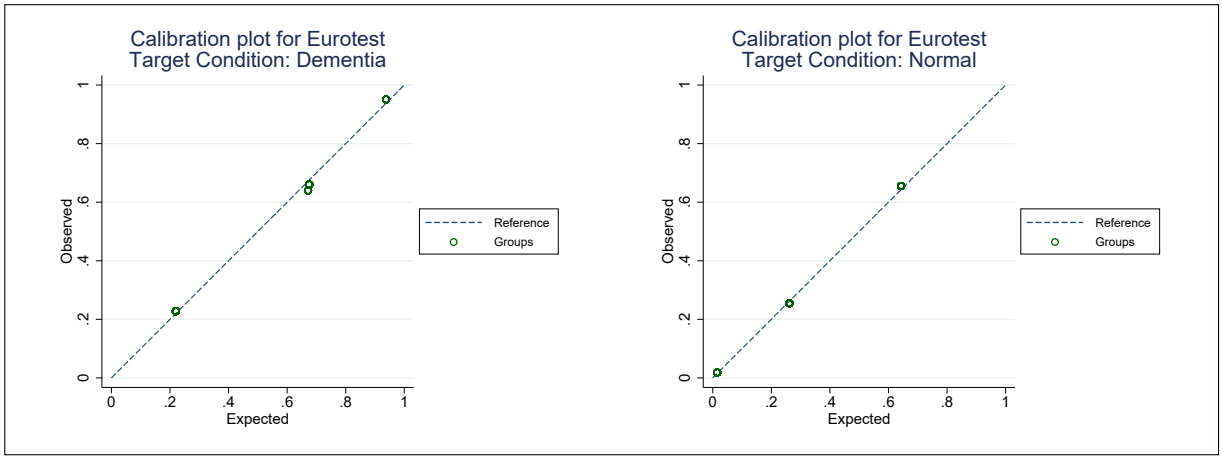

Target Condition: Dementia
Sensitivity
1-Specificity
CIT
GP + CIT
Target Condition: Normal
Sensitivity
1-Specificity
CIT
GP + CIT

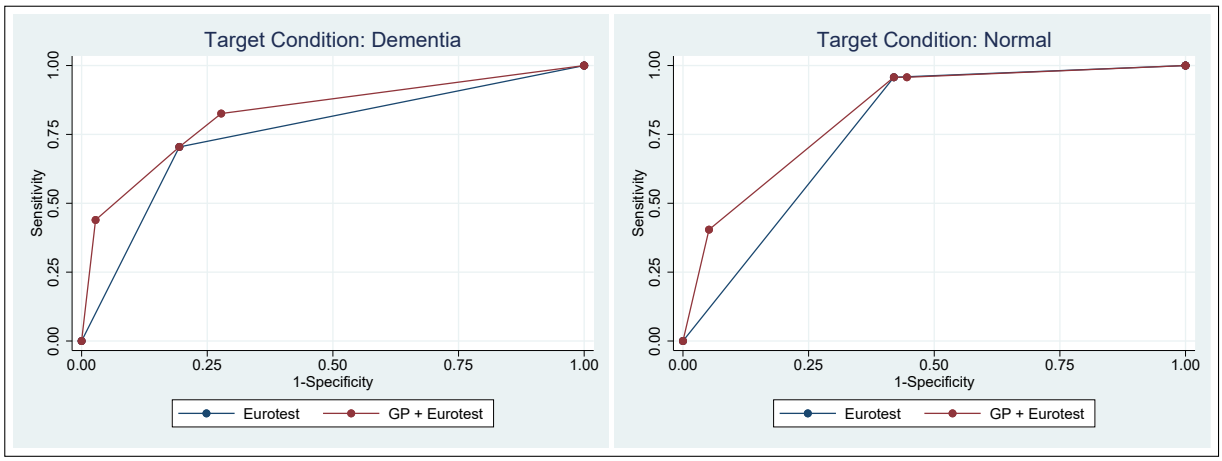

Calibration plot for GPCOG
Target Condition: Dementia
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups
Calibration plot for GPCOG
Target Condition: Normal
Observed
Expected
Reference
Groups

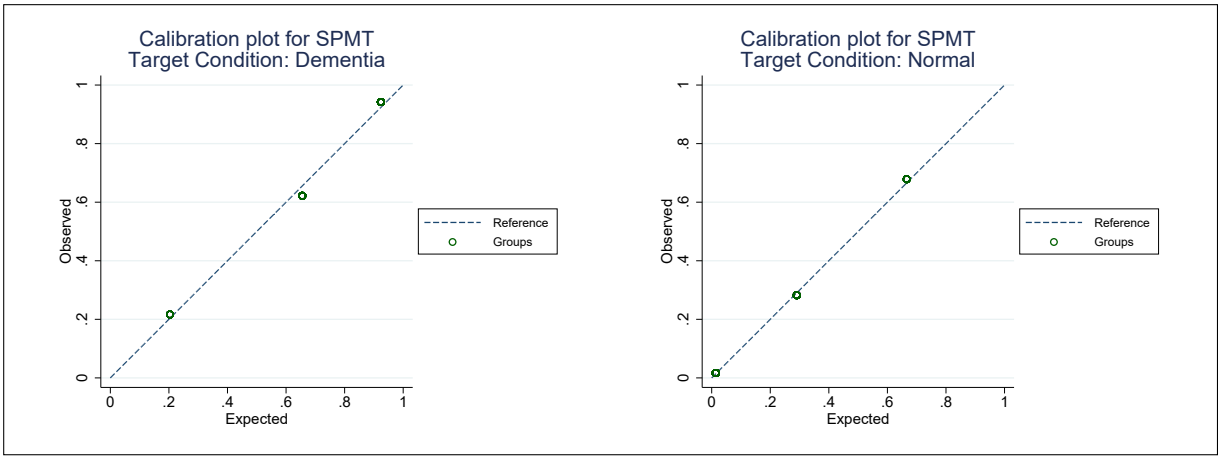

Target Condition: Dementia
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
GPCOG
GP + GPCOG
Target Condition: Normal
Sensitivity
1-Specificity
GPCOG
GP + GPCOG

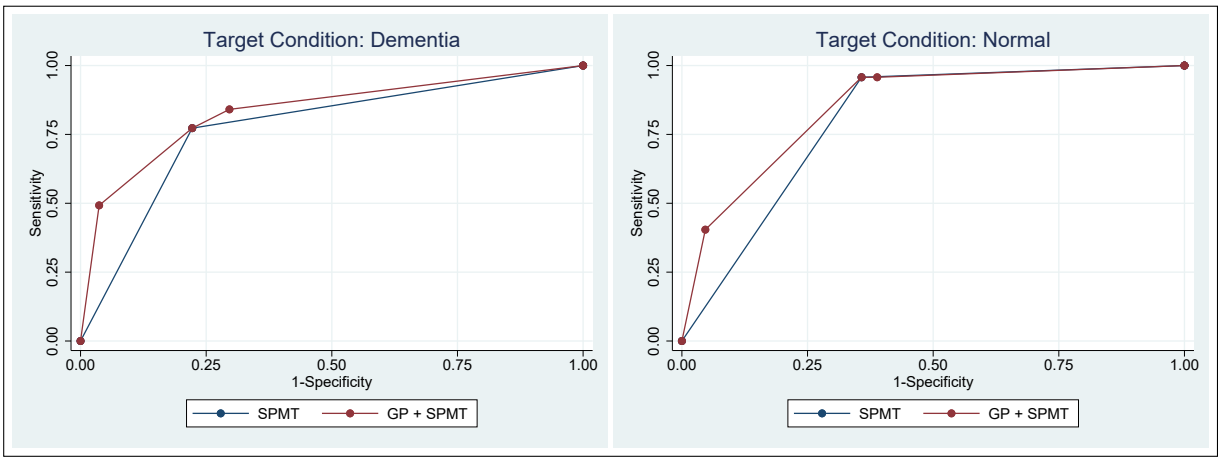

Calibration plot for Minicog
Target Condition: Dementia
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups
Calibration plot for Minicog
Target Condition: Normal
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups

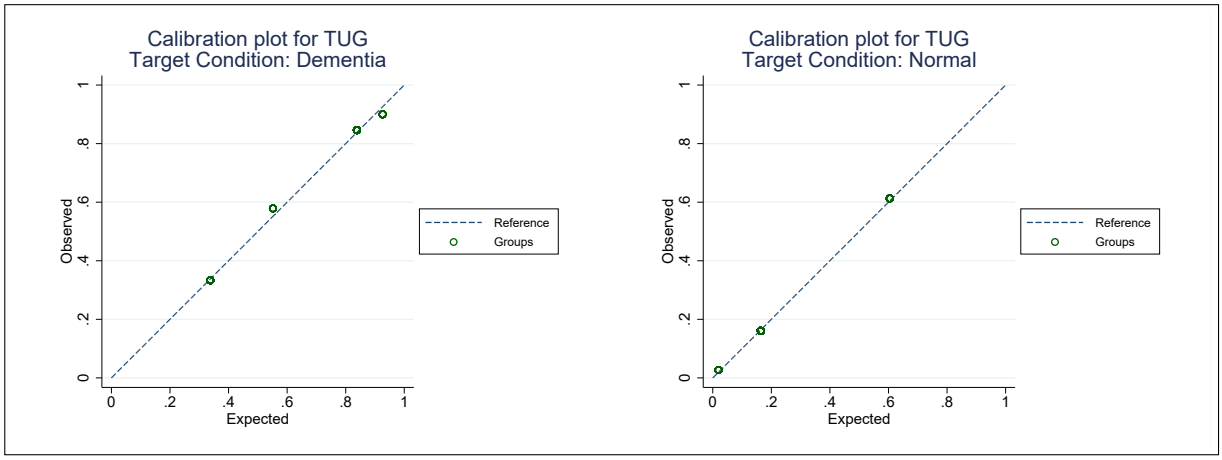

Target Condition: Dementia
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
Minicog
GP + Minicog
Target Condition: Normal
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
Minicog
GP + Minicog

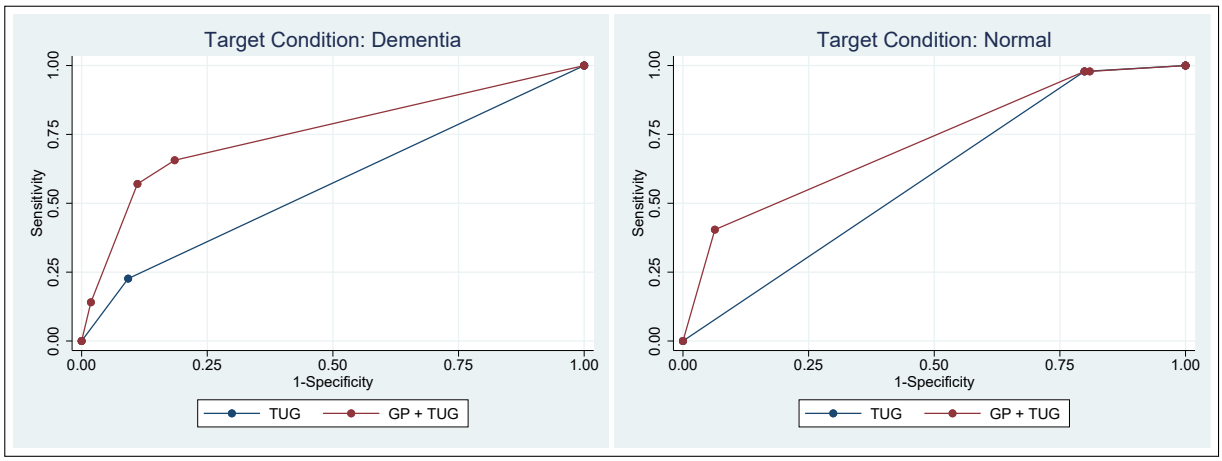

Calibration plot for Eurotest
Target Condition: Dementia
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups
Calibration plot for Eurotest
Target Condition: Normal
Observed
Expected
Reference
Groups

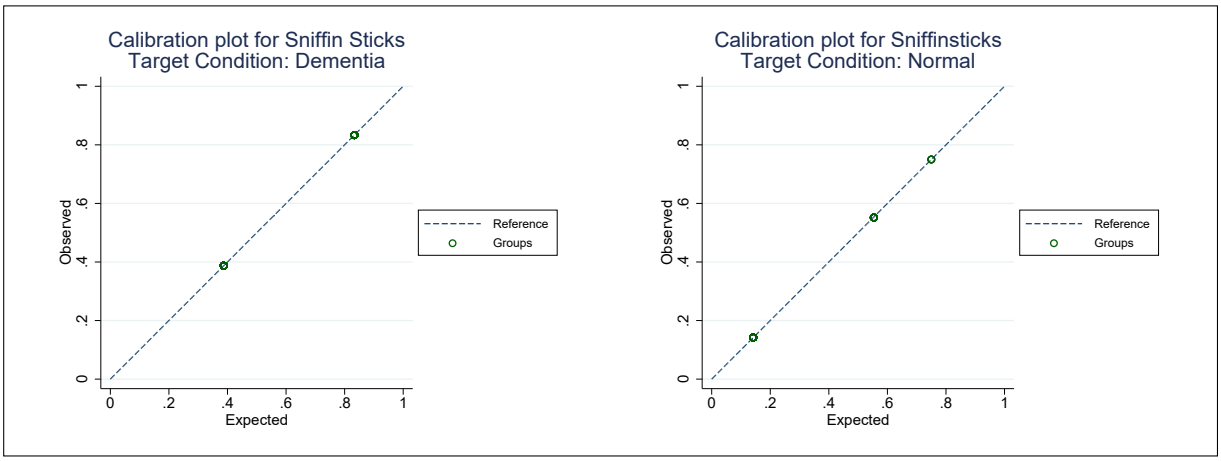

Target Condition: Dementia
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
Eurotest
GP + Eurotest
Target Condition: Normal
Sensitivity
1-Specificity
Eurotest
GP + Eurotest

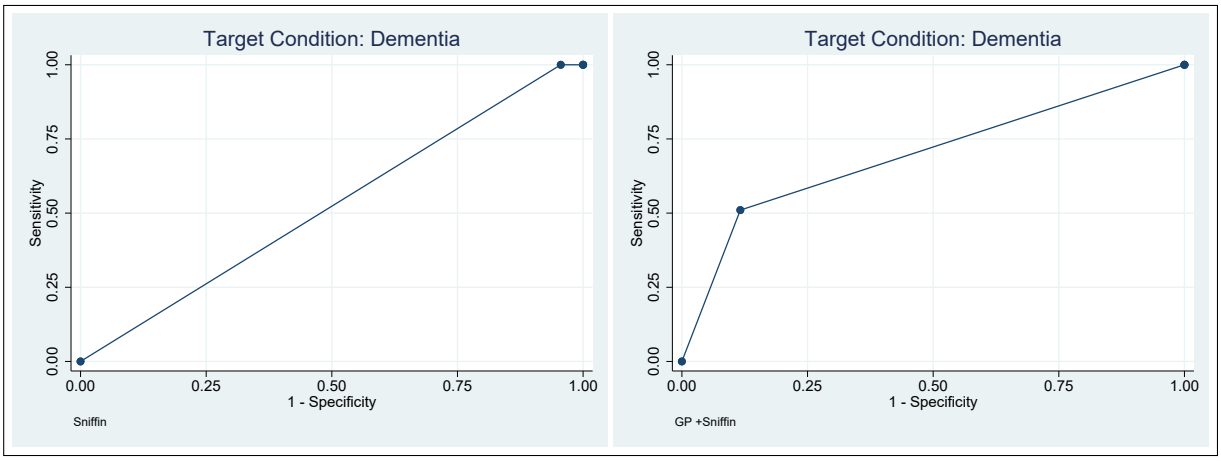

Calibration plot for SPMT
Target Condition: Dementia
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups
Calibration plot for SPMT
Target Condition: Normal
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups

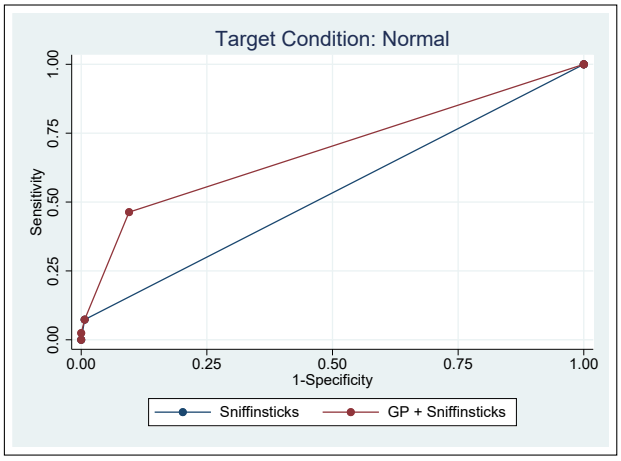

Target Condition: Dementia
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
SPMT
GP + SPMT
Target Condition: Normal
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
SPMT
GP + SPMT

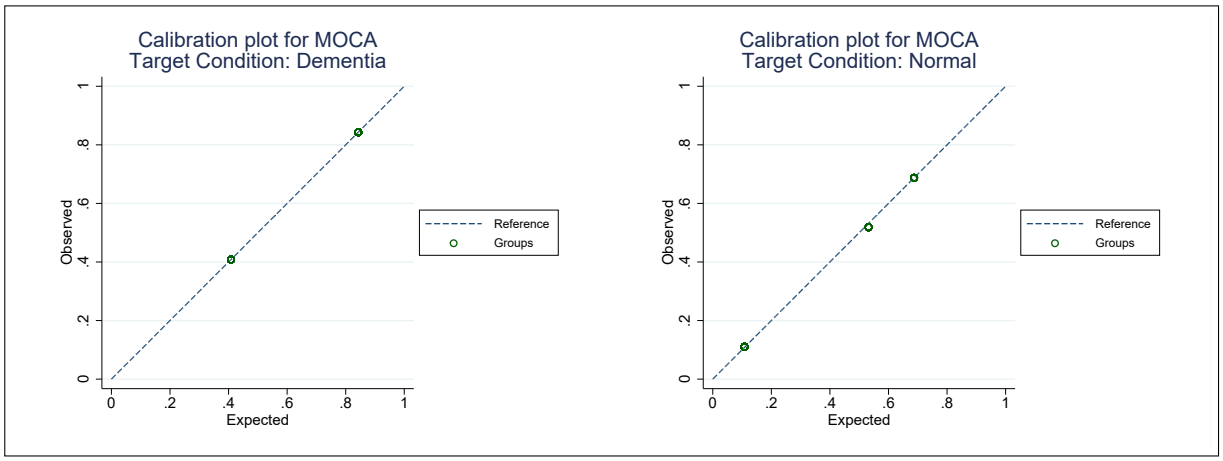

Calibration plot for TUG
Target Condition: Dementia
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups
Calibration plot for TUG
Target Condition: Normal
Observed
Expected
Reference
Groups

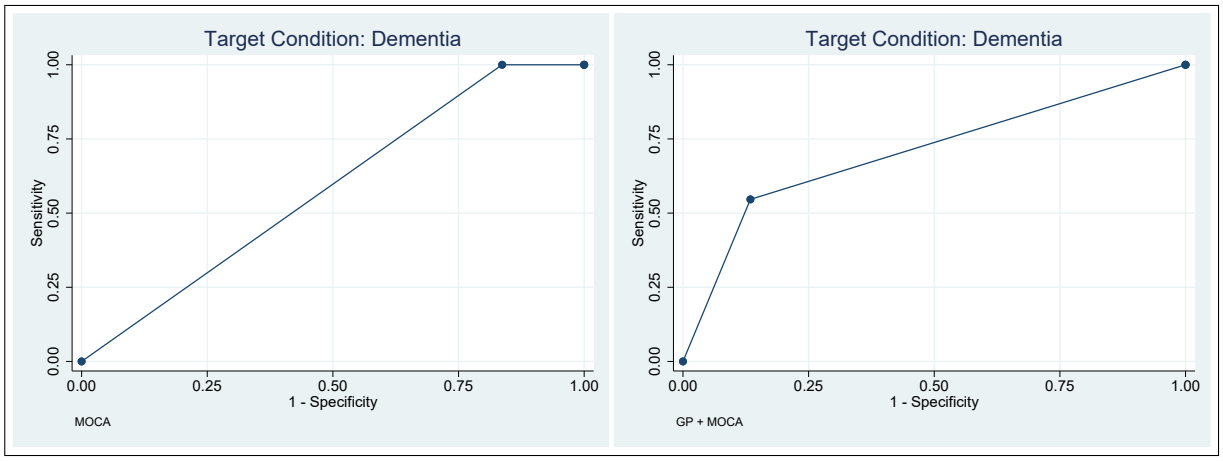

Target Condition: Dementia
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
TUG
GP + TUG
Target Condition: Normal
Sensitivity
1-Specificity
TUG
GP + TUG

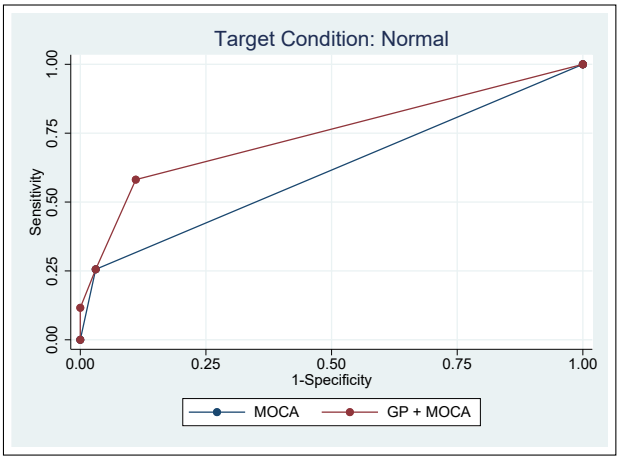

Calibration plot for Sniffin Sticks
Target Condition: Dementia
Observed
Expected
Reference
Groups
Calibration plot for Sniffinsticks
Target Condition: Normal
Observed
Expected
Reference
Groups

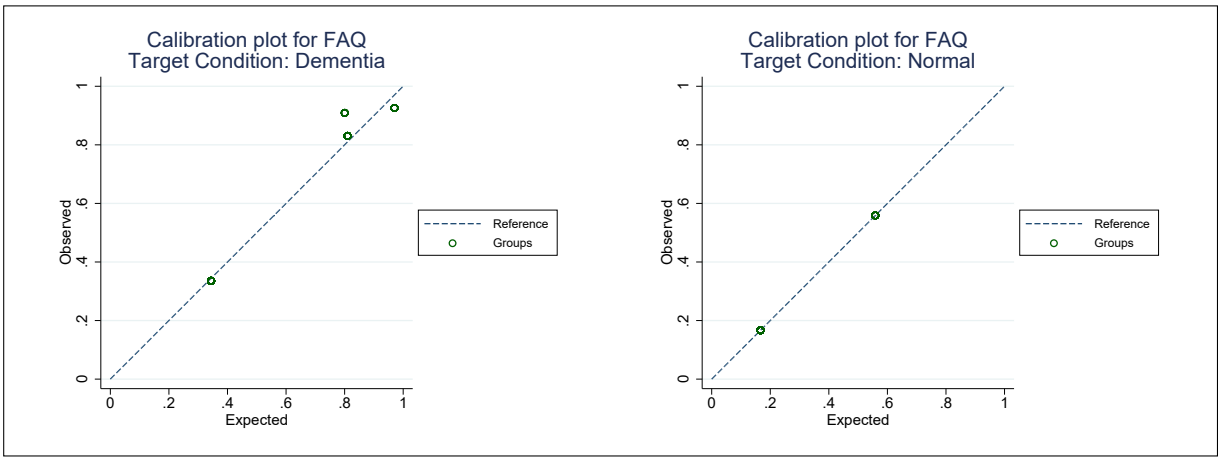

Target Condition: Dementia
Sensitivity
1 - Specificity
Sniffin
Target Condition: Dementia
Sensitivity
1 - Specificity
GP +Sniffin

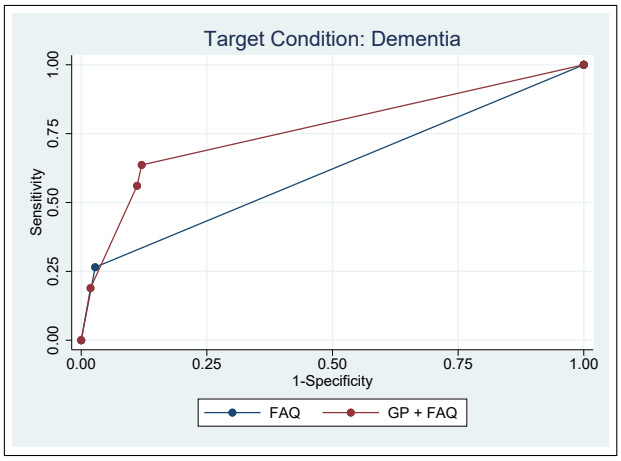

Target Condition: Normal
Sensitivity
1-Specificity
Sniffinsticks
GP + Sniffinsticks

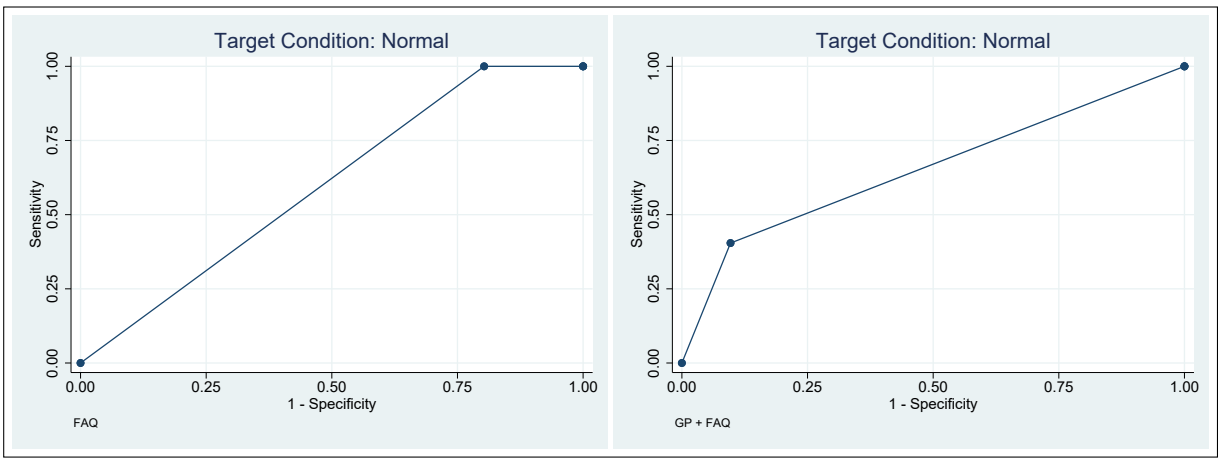

Calibration plot for MOCA
Target Condition: Dementia
Observed
Expected
Reference
Groups
Calibration plot for MOCA
Target Condition: Normal
Observed
Expected
Reference
Groups

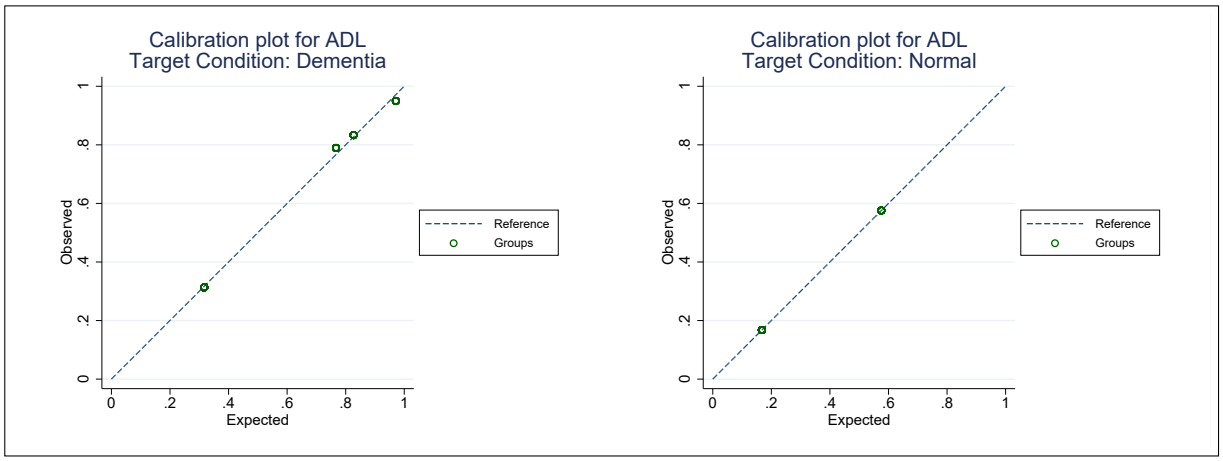

Target Condition: Dementia
Sensitivity
1 - Specificity
MOCA
Target Condition: Dementia
Sensitivity
1 - Specificity
GP + MOCA

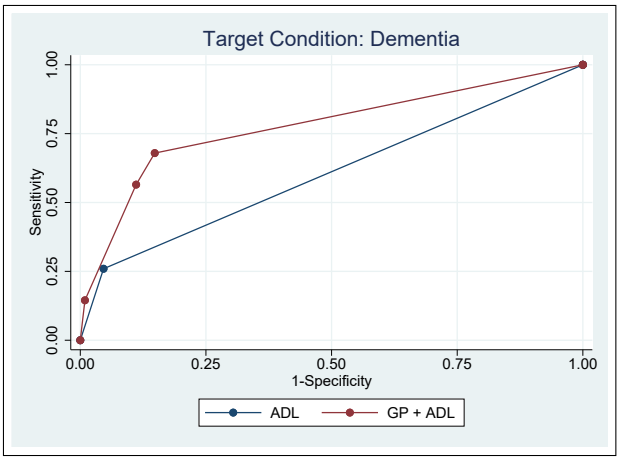

Target Condition: Normal
Sensitivity
1-Specificity
MOCA
GP + MOCA

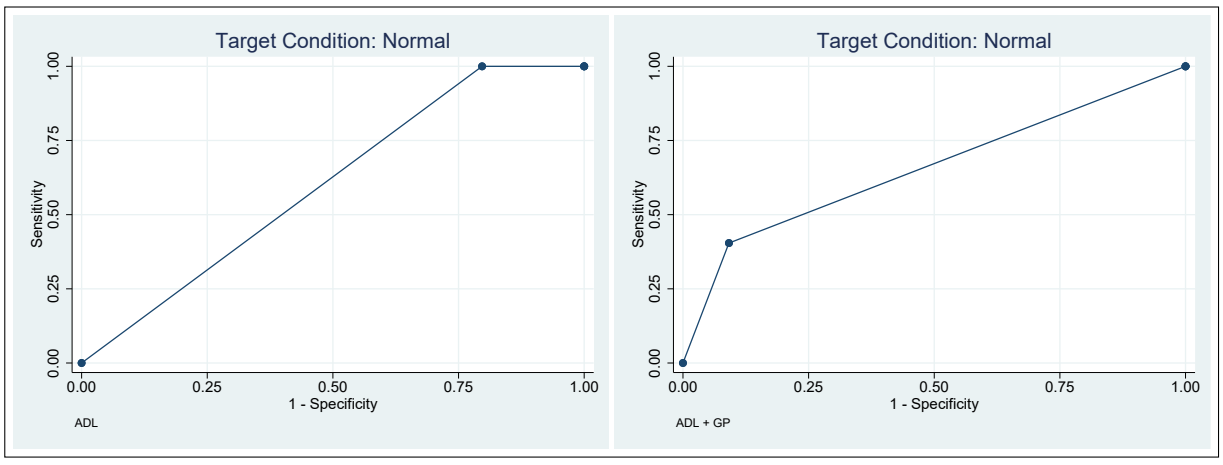

Calibration plot for FAQ
Target Condition: Dementia
Observed
Expected
Reference
Groups
Calibration plot for FAQ
Target Condition: Normal
Observed
Expected
Reference
Groups

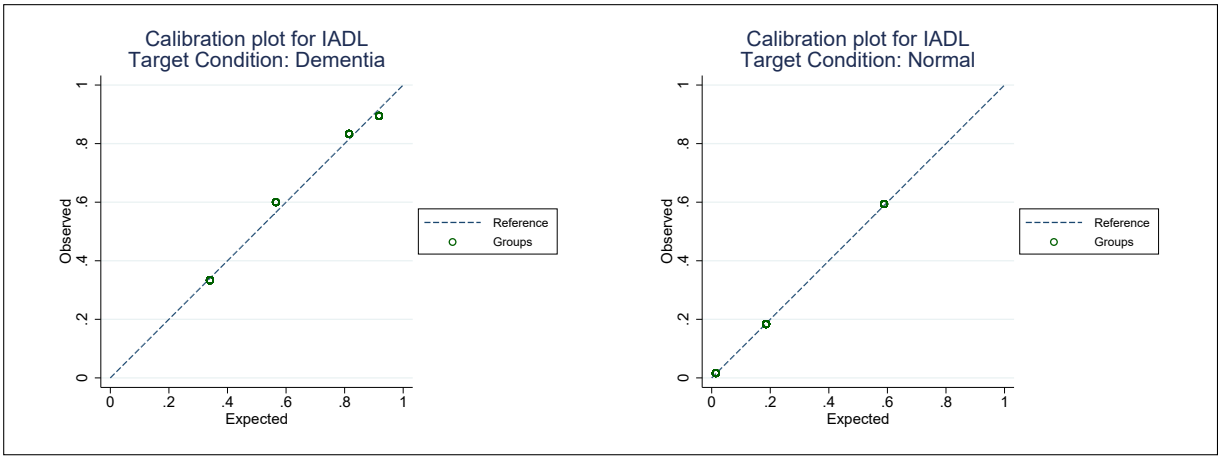

Target Condition: Dementia
Sensitivity
1-Specificity
FAQ
GP + FAQ

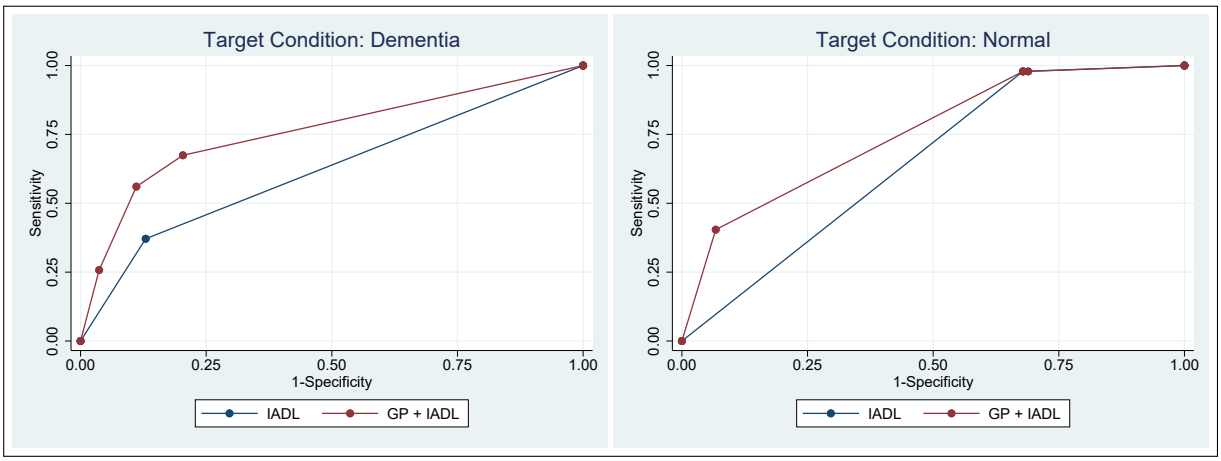

Target Condition: Normal
Sensitivity
1 - Specificity
FAQ
Target Condition: Normal
Sensitivity
1 - Specificity
GP + FAQ

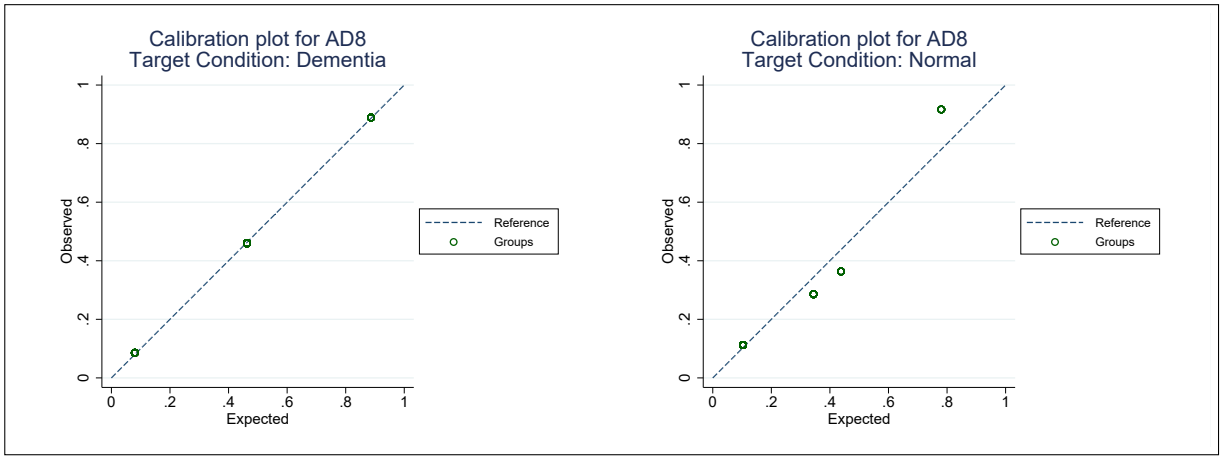

Calibration plot for ADL
Target Condition: Dementia
Observed
Expected
Reference
Groups
Calibration plot for ADL
Target Condition: Normal
Observed
Expected
Reference
Groups

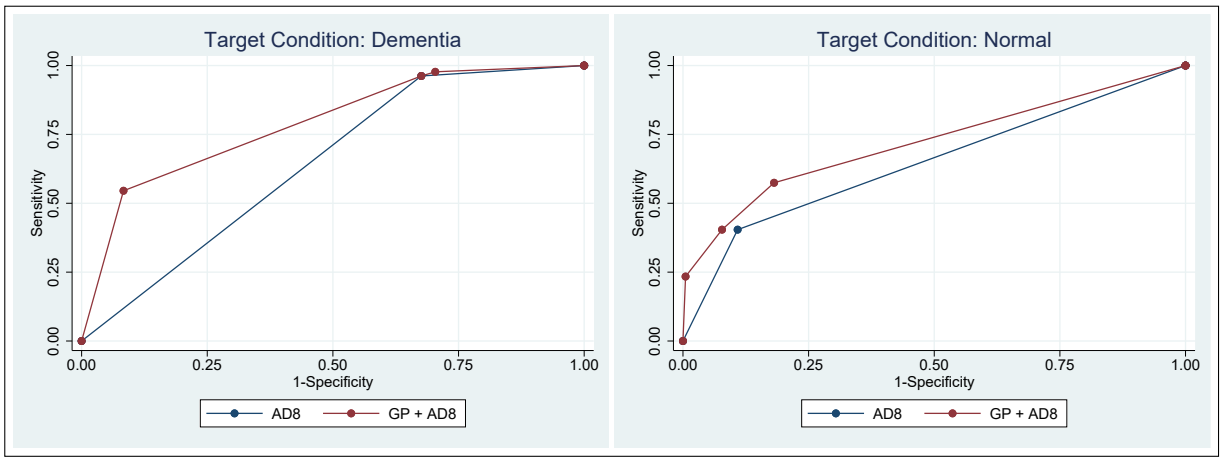

Target Condition: Dementia
Sensitivity
1-Specificity
ADL
GP + ADL

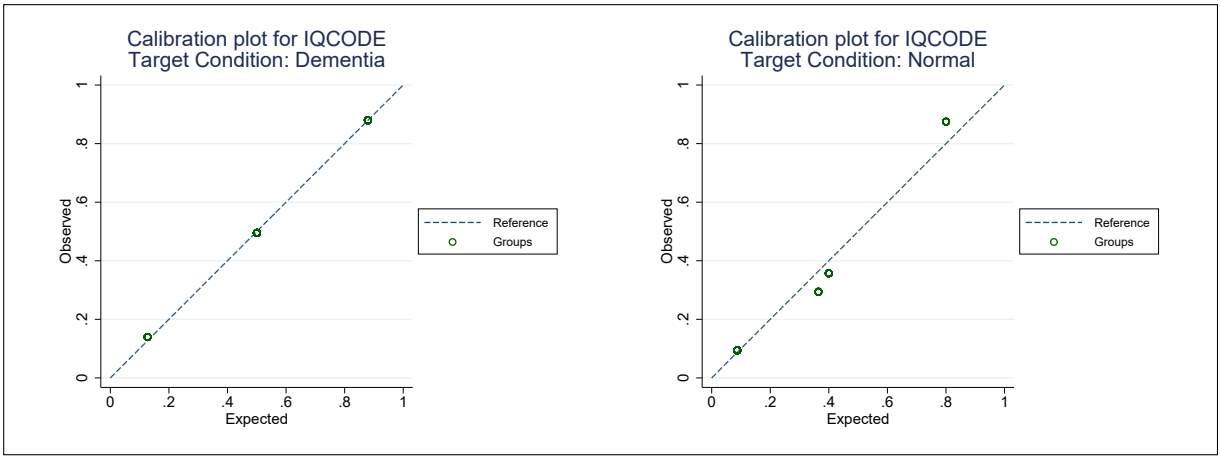

Target Condition: Normal
Sensitivity
1 - Specificity
ADL
Target Condition: Normal
Sensitivity
1 - Specificity
ADL + GP

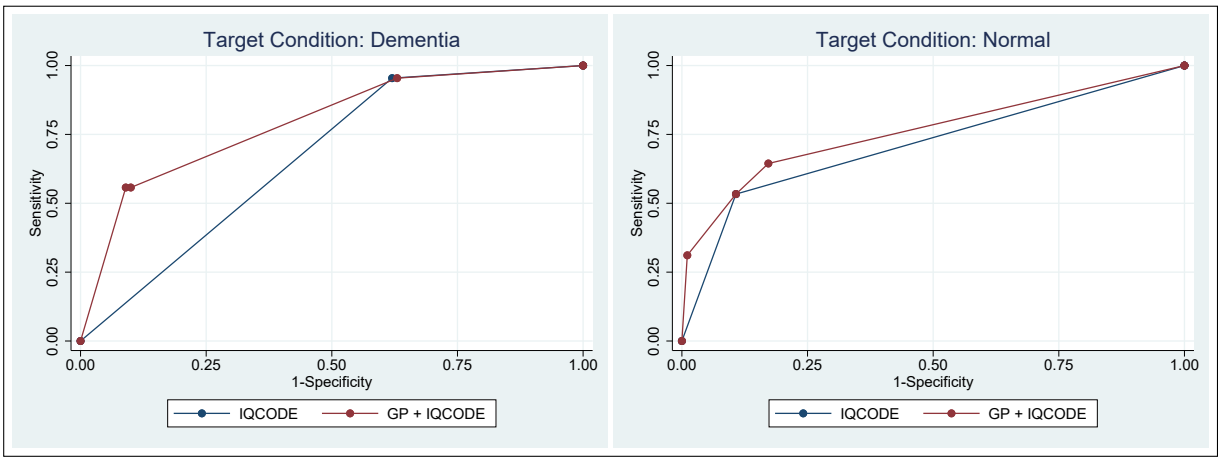

Calibration plot for IADL
Target Condition: Dementia
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups
Calibration plot for IADL
Target Condition: Normal
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups


Target Condition: Dementia
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
IADL
GP + IADL
Target Condition: Normal
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
IADL
GP + IADL

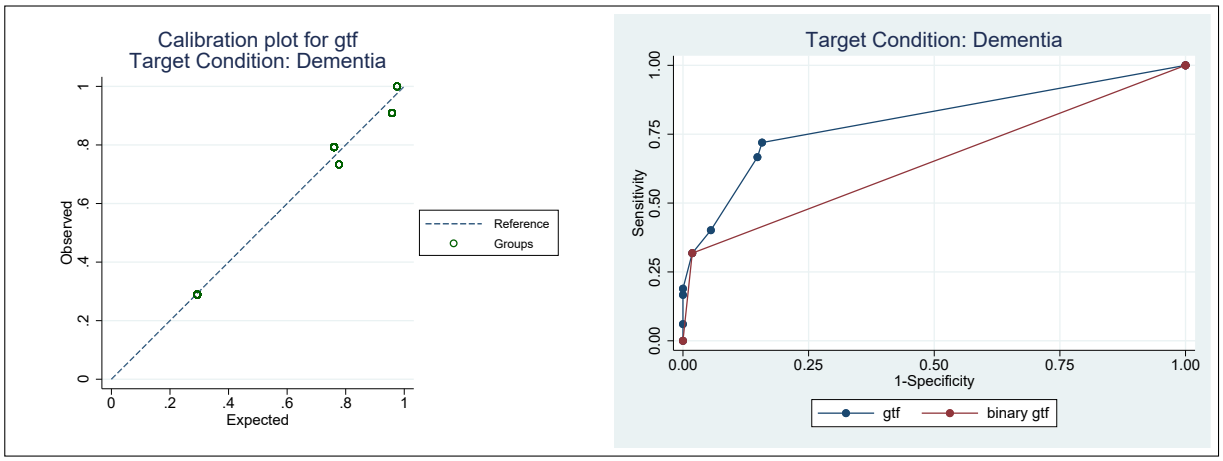

Calibration plot for AD8
Target Condition: Dementia
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups
Calibration plot for AD8
Target Condition: Normal
Observed
Expected
Reference
Groups


Target Condition: Dementia
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
AD8
GP + AD8
Target Condition: Normal
Sensitivity
1-Specificity
AD8
GP + AD8

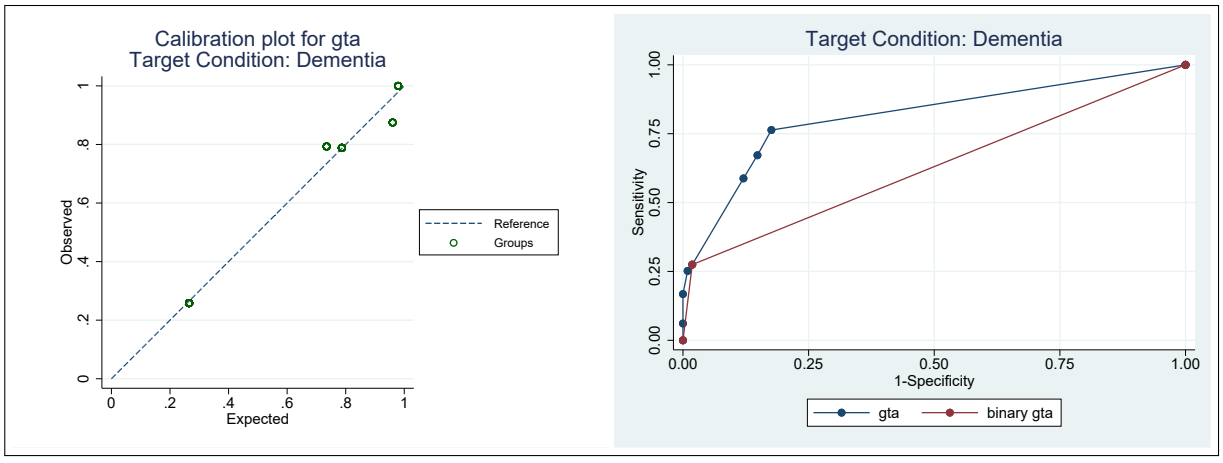

Calibration plot for IQCODE
Target Condition: Dementia
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups
Calibration plot for IQCODE
Target Condition: Normal
Observed
Expected
0
.2
.4
.6
.8
1
Reference
Groups


Target Condition: Dementia
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
IQCODE
GP + IQCODE
Target Condition: Normal
Sensitivity
1-Specificity
0.00
0.25
0.50
0.75
1.00
IQCODE
GP + IQCODE

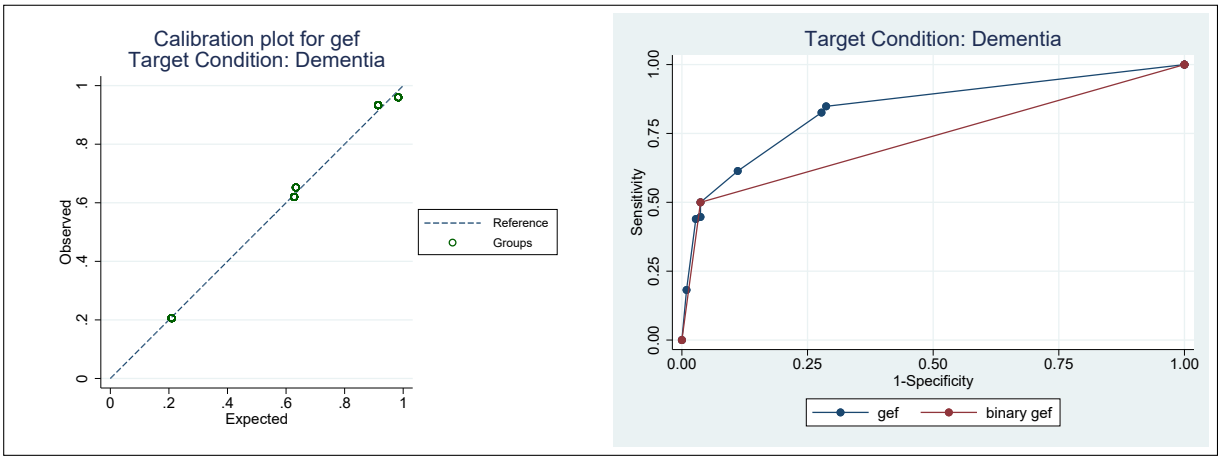
The following calibration plots and ROC plots are for the GP 360 tests which combined GP judgement, a patient completed test, and an informant completed test.



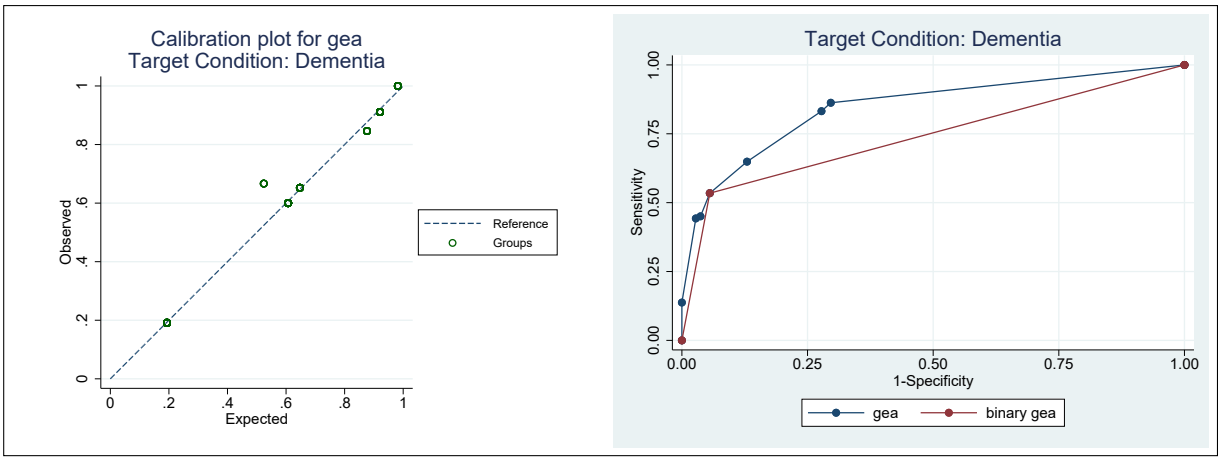
gtf GP + TAC + FAQ



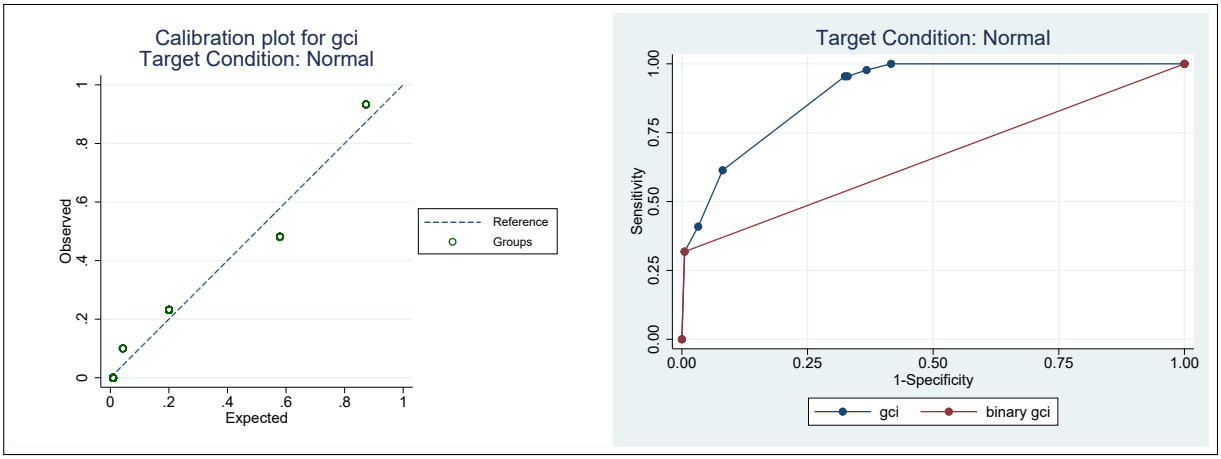
gta GP + TAC + ADL



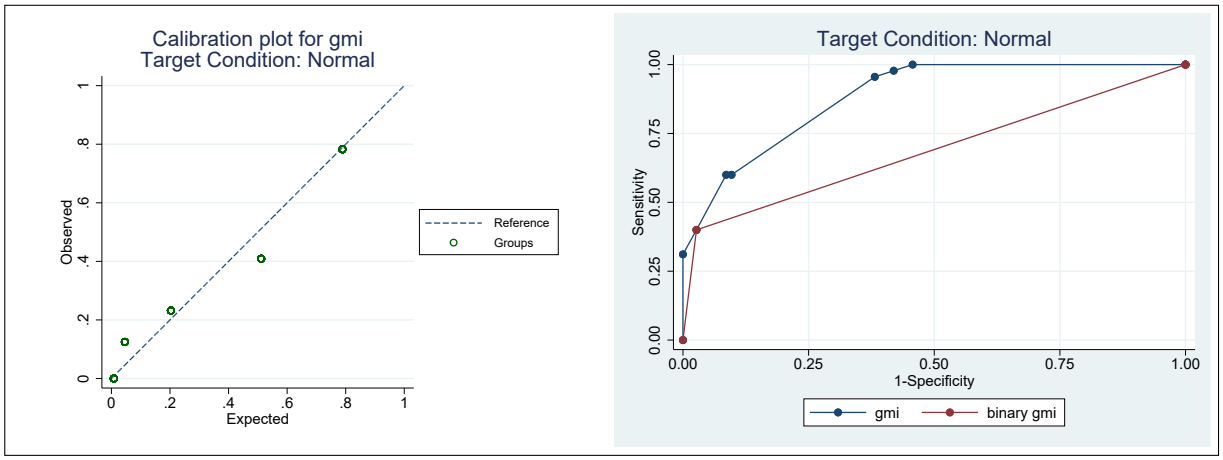
gef GP + Eurotest + FAQ



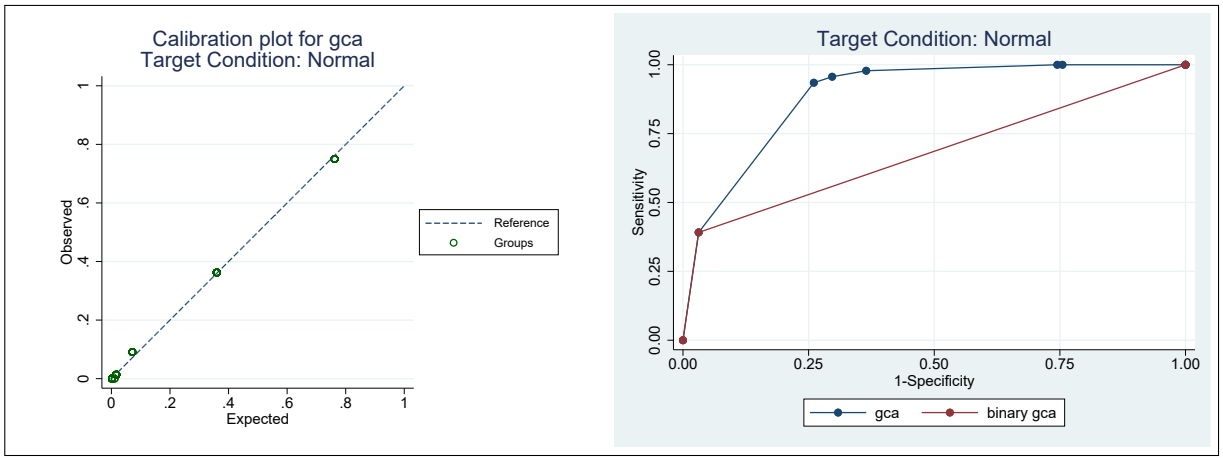
gea GP + Eurotest + ADL



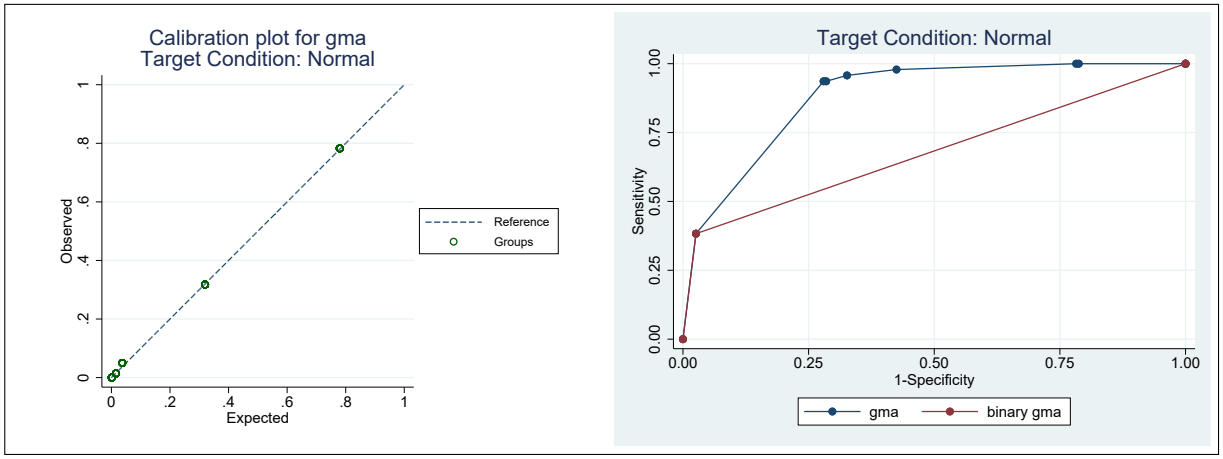
gci GP + 6CIT + IQCODE



gmi GP + Minicog + IQCODE



gca GP + 6CIT + IADL



gmi GP + Minicog + IADL

APPENDIX J

SENSITIVITY ANALYSES TABLES

This appendix gives tables of the sensitivity analyses for the quantitative study.

Table J.1: Sensitivity analysis: continuous tests combined with GP judgement, target condition dementia

Test	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LRP (95% CI)	LRN (95% CI)	DOR (95% CI)	Probability ‡
Tests as continuous scores, dichotomised at a predicted probability of dementia ≥ 80%								
EPSS	39 (31 to 48)	92 (85 to 96)	85 (74 to 93)	55 (48 to 63)	5 (2 to 9)	0.66 (0.57 to 0.77)	7 (3 to 17)	0.99
TAC	67 (58 to 75)	85 (77 to 91)	85 (76 to 91)	68 (59 to 75)	5 (3 to 7)	0.39 (0.3 to 0.5)	12 (6 to 23)	1
Phototest	47 (38 to 55)	93 (86 to 97)	88 (78 to 95)	59 (51 to 66)	6 (3 to 12)	0.58 (0.49 to 0.68)	11 (5 to 28)	0.99
6CIT	47 (39 to 56)	90 (82 to 95)	85 (75 to 92)	58 (50 to 66)	5 (3 to 8)	0.59 (0.49 to 0.7)	8 (4 to 18)	0.99
GPCOG	52 (43 to 61)	94 (87 to 97)	91 (82 to 96)	62 (54 to 69)	8 (4 to 17)	0.51 (0.42 to 0.61)	16 (7 to 43)	0.98
Minicog	42 (34 to 51)	94 (87 to 97)	89 (78 to 95)	57 (49 to 64)	7 (3 to 14)	0.62 (0.53 to 0.72)	11 (4 to 29)	0.96
Eurotest	61 (52 to 69)	97 (92 to 99)	96 (90 to 99)	67 (59 to 74)	22 (7 to 67)	0.41 (0.33 to 0.5)	54 (16 to 275)	1
SPMT	56 (47 to 65)	94 (88 to 98)	93 (84 to 97)	64 (56 to 71)	10 (5 to 22)	0.47 (0.38 to 0.57)	22 (9 to 64)	0.98
TUG	51 (42 to 60)	91 (84 to 95)	87 (77 to 93)	61 (53 to 68)	5 (3 to 10)	0.54 (0.45 to 0.65)	10 (5 to 24)	1
Sniffin sticks	37 (27 to 47)	92 (85 to 97)	84 (69 to 93)	57 (49 to 65)	5 (2 to 10)	0.69 (0.58 to 0.81)	7 (3 to 19)	0.95
MAT	67 (45 to 84)	90 (55 to 100)	94 (71 to 100)	53 (28 to 77)	7 (1 to 44)	0.37 (0.2 to 0.68)	18 (2 to 837)	0.99
MOCA	56 (46 to 65)	92 (85 to 96)	88 (78 to 95)	65 (57 to 73)	7 (3 to 14)	0.48 (0.39 to 0.6)	14 (6 to 36)	1
FAQ	55 (46 to 63)	94 (88 to 98)	92 (84 to 97)	63 (55 to 70)	10 (4 to 22)	0.48 (0.4 to 0.58)	20 (8 to 60)	1
ADL	61 (52 to 69)	89 (81 to 94)	87 (78 to 93)	65 (57 to 73)	5 (3 to 9)	0.44 (0.36 to 0.55)	12 (6 to 27)	1
IADL	48 (40 to 57)	94 (88 to 98)	91 (82 to 97)	60 (52 to 67)	9 (4 to 19)	0.55 (0.46 to 0.65)	16 (6 to 47)	0.96
AD8	49 (40 to 58)	95 (90 to 98)	93 (84 to 98)	61 (53 to 68)	11 (4 to 25)	0.53 (0.45 to 0.63)	20 (7 to 66)	0.96
IQCODE	53 (44 to 61)	95 (89 to 98)	93 (85 to 98)	61 (52 to 68)	11 (4 to 25)	0.5 (0.41 to 0.6)	21 (8 to 70)	0.98
Tests using published thresholds, dichotomised at a predicted probability of dementia ≥ 80%								
EPSS	44 (35 to 53)	91 (84 to 95)	85 (75 to 93)	57 (49 to 64)	5 (3 to 9)	0.62 (0.53 to 0.73)	8 (4 to 18)	0.99
TAC	67 (58 to 75)	85 (77 to 91)	85 (76 to 91)	68 (59 to 75)	5 (3 to 7)	0.39 (0.3 to 0.5)	12 (6 to 23)	1
Phototest	37 (29 to 46)	94 (88 to 98)	89 (78 to 96)	55 (48 to 63)	7 (3 to 15)	0.66 (0.58 to 0.76)	10 (4 to 30)	0.99
6CIT	46 (37 to 55)	93 (86 to 97)	88 (78 to 95)	58 (50 to 66)	6 (3 to 12)	0.59 (0.5 to 0.69)	10 (5 to 27)	0.99
GPCOG	55 (46 to 64)	91 (84 to 95)	88 (79 to 94)	62 (54 to 70)	6 (3 to 11)	0.49 (0.4 to 0.6)	12 (6 to 28)	0.98
Minicog	42 (34 to 51)	94 (87 to 97)	89 (78 to 95)	57 (49 to 64)	7 (3 to 14)	0.62 (0.53 to 0.72)	11 (4 to 29)	0.96
Eurotest	44 (35 to 53)	97 (92 to 99)	95 (86 to 99)	59 (51 to 66)	16 (5 to 49)	0.58 (0.49 to 0.67)	27 (8 to 140)	1
SPMT	49 (40 to 58)	96 (91 to 99)	94 (86 to 98)	61 (53 to 68)	13 (5 to 35)	0.53 (0.44 to 0.63)	25 (9 to 99)	0.98
TUG	57 (48 to 66)	89 (81 to 94)	86 (77 to 92)	64 (55 to 71)	5 (3 to 9)	0.48 (0.39 to 0.6)	11 (5 to 23)	1
Sniffin sticks	51 (41 to 61)	88 (80 to 94)	83 (71 to 92)	61 (52 to 70)	4 (2 to 8)	0.55 (0.45 to 0.69)	8 (4 to 19)	0.95
MAT	63 (41 to 81)	90 (55 to 100)	94 (70 to 100)	50 (26 to 74)	6 (1 to 41)	0.42 (0.24 to 0.73)	15 (2 to 700)	0.99
MOCA	55 (45 to 64)	87 (77 to 93)	84 (74 to 92)	59 (50 to 68)	4 (2 to 7)	0.52 (0.42 to 0.66)	8 (4 to 18)	1
FAQ	64 (55 to 72)	88 (80 to 93)	87 (78 to 93)	66 (58 to 74)	5 (3 to 9)	0.41 (0.33 to 0.52)	13 (6 to 27)	1
ADL	56 (48 to 65)	89 (81 to 94)	86 (77 to 93)	63 (55 to 70)	5 (3 to 9)	0.49 (0.4 to 0.6)	10 (5 to 23)	1
IADL	56 (47 to 65)	89 (81 to 94)	86 (77 to 93)	62 (54 to 70)	5 (3 to 9)	0.49 (0.4 to 0.61)	10 (5 to 22)	0.96
AD8	55 (46 to 63)	92 (85 to 96)	89 (80 to 95)	62 (54 to 70)	7 (3 to 12)	0.5 (0.41 to 0.6)	13 (6 to 32)	0.96
IQCODE	56 (47 to 64)	91 (84 to 96)	89 (80 to 95)	61 (53 to 69)	6 (3 to 12)	0.49 (0.4 to 0.6)	13 (6 to 31)	0.98

‡ predicted probability of dementia in test positive group
see Glossary for abbreviations of tests

Table J.2: Sensitivity analysis: continuous tests combined with GP judgement, target condition normal

Test	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LRP (95% CI)	LRN (95% CI)	DOR (95% CI)	Probability ‡
Tests as continuous scores, dichotomised at a predicted probability of normal ≥ 60%								
EPSS	32 (19 to 47)	97 (93 to 99)	71 (48 to 89)	85 (80 to 90)	10 (4 to 25)	0.7 (0.58 to 0.86)	15 (5 to 49)	0.78
Phototest	35 (21 to 50)	97 (93 to 99)	73 (50 to 89)	86 (81 to 90)	11 (5 to 27)	0.67 (0.54 to 0.83)	17 (6 to 55)	0.95
6CIT	50 (35 to 65)	94 (90 to 97)	68 (49 to 83)	89 (84 to 93)	9 (5 to 17)	0.53 (0.4 to 0.71)	16 (7 to 42)	0.88
GPCOG	30 (17 to 45)	98 (95 to 99)	78 (52 to 94)	85 (80 to 90)	14 (5 to 42)	0.72 (0.59 to 0.86)	20 (6 to 87)	0.89
Minicog	36 (23 to 51)	97 (93 to 99)	74 (52 to 90)	86 (81 to 90)	12 (5 to 28)	0.66 (0.53 to 0.82)	18 (6 to 58)	0.85
Eurotest	38 (25 to 54)	96 (92 to 98)	69 (48 to 86)	86 (81 to 91)	9 (4 to 20)	0.64 (0.51 to 0.81)	14 (5 to 41)	0.96
SPMT	30 (17 to 45)	96 (93 to 99)	67 (43 to 85)	85 (79 to 89)	8 (4 to 19)	0.73 (0.6 to 0.88)	11 (4 to 35)	0.95
TUG	26 (14 to 40)	96 (92 to 98)	60 (36 to 81)	84 (78 to 88)	6 (3 to 14)	0.78 (0.66 to 0.92)	8 (3 to 23)	0.84
Sniffin sticks	32 (18 to 48)	99 (95 to 100)	87 (60 to 98)	84 (77 to 89)	23 (5 to 99)	0.69 (0.56 to 0.85)	34 (7 to 315)	0.86
MAT	25 (1 to 81)	96 (82 to 100)	50 (1 to 99)	90 (73 to 98)	7 (1 to 91)	0.78 (0.44 to 1.38)	9 (0 to 731)	0.69
MOCA	49 (33 to 65)	96 (91 to 98)	75 (55 to 89)	88 (82 to 92)	11 (5 to 25)	0.53 (0.4 to 0.72)	21 (7 to 65)	0.98
FAQ	28 (16 to 43)	98 (95 to 99)	76 (50 to 93)	85 (79 to 89)	13 (5 to 39)	0.74 (0.62 to 0.88)	18 (5 to 79)	0.75
ADL	36 (23 to 51)	94 (90 to 97)	61 (41 to 78)	86 (80 to 90)	6 (3 to 13)	0.68 (0.54 to 0.84)	9 (4 to 24)	0.61
IADL	34 (21 to 49)	97 (93 to 99)	73 (50 to 89)	86 (80 to 90)	11 (5 to 26)	0.68 (0.55 to 0.84)	16 (5 to 53)	0.74
AD8	23 (12 to 38)	98 (96 to 100)	79 (49 to 95)	84 (79 to 89)	15 (4 to 52)	0.78 (0.66 to 0.91)	19 (5 to 111)	0.83
IQCODE	38 (24 to 53)	98 (95 to 99)	81 (58 to 95)	87 (81 to 91)	18 (6 to 50)	0.64 (0.51 to 0.8)	28 (8 to 118)	0.96
Tests using published thresholds, dichotomised at a predicted probability of normal ≥ 60%								
EPSS	32 (19 to 47)	97 (93 to 99)	71 (48 to 89)	85 (80 to 90)	10 (4 to 25)	0.7 (0.58 to 0.86)	15 (5 to 49)	0.78
Phototest	35 (21 to 50)	97 (93 to 99)	73 (50 to 89)	86 (81 to 90)	11 (5 to 27)	0.67 (0.54 to 0.83)	17 (6 to 55)	0.95
6CIT	50 (35 to 65)	94 (90 to 97)	68 (49 to 83)	89 (84 to 93)	9 (5 to 17)	0.53 (0.4 to 0.71)	16 (7 to 42)	0.88
GPCOG	30 (17 to 45)	98 (95 to 99)	78 (52 to 94)	85 (80 to 90)	14 (5 to 42)	0.72 (0.59 to 0.86)	20 (6 to 87)	0.89
Minicog	36 (23 to 51)	97 (93 to 99)	74 (52 to 90)	86 (81 to 90)	12 (5 to 28)	0.66 (0.53 to 0.82)	18 (6 to 58)	0.85
Eurotest	38 (25 to 54)	96 (92 to 98)	69 (48 to 86)	86 (81 to 91)	9 (4 to 20)	0.64 (0.51 to 0.81)	14 (5 to 41)	0.96
SPMT	30 (17 to 45)	96 (93 to 99)	67 (43 to 85)	85 (79 to 89)	8 (4 to 19)	0.73 (0.6 to 0.88)	11 (4 to 35)	0.95
TUG	26 (14 to 40)	96 (92 to 98)	60 (36 to 81)	84 (78 to 88)	6 (3 to 14)	0.78 (0.66 to 0.92)	8 (3 to 23)	0.84
Sniffin sticks	32 (18 to 48)	99 (95 to 100)	87 (60 to 98)	84 (77 to 89)	23 (5 to 99)	0.69 (0.56 to 0.85)	34 (7 to 315)	0.86
MAT	25 (1 to 81)	96 (82 to 100)	50 (1 to 99)	90 (73 to 98)	7 (1 to 91)	0.78 (0.44 to 1.38)	9 (0 to 731)	0.69
MOCA	49 (33 to 65)	96 (91 to 98)	75 (55 to 89)	88 (82 to 92)	11 (5 to 25)	0.53 (0.4 to 0.72)	21 (7 to 65)	0.98
FAQ	28 (16 to 43)	98 (95 to 99)	76 (50 to 93)	85 (79 to 89)	13 (5 to 39)	0.74 (0.62 to 0.88)	18 (5 to 79)	0.75
ADL	36 (23 to 51)	94 (90 to 97)	61 (41 to 78)	86 (80 to 90)	6 (3 to 13)	0.68 (0.54 to 0.84)	9 (4 to 24)	0.61
IADL	34 (21 to 49)	97 (93 to 99)	73 (50 to 89)	86 (80 to 90)	11 (5 to 26)	0.68 (0.55 to 0.84)	16 (5 to 53)	0.74
AD8	23 (12 to 38)	98 (96 to 100)	79 (49 to 95)	84 (79 to 89)	15 (4 to 52)	0.78 (0.66 to 0.91)	19 (5 to 111)	0.83
IQCODE	38 (24 to 53)	98 (95 to 99)	81 (58 to 95)	87 (81 to 91)	18 (6 to 50)	0.64 (0.51 to 0.8)	28 (8 to 118)	0.96

‡ predicted probability of normal in test positive group
see Glossary for abbreviations of tests

Table J.3: Sensitivity analysis: age, tests combined with GP judgement, target condition dementia

Test	Sensitivity	Sensitivity lb	Sensitivity ub	Specificity	Specificity lb	Specificity ub ‡
EPSS all	44	35	53	91	84	95
age80 =0	32	19	47	94	86	98
age80 =1	51	40	62	84	69	94
TAC all	67	58	75	85	77	91
age80 =0	15	6	28	100	95	100
age80 =1	68	57	78	79	63	90
Phototest all	37	29	46	94	88	98
age80 =0	38	25	54	96	88	99
age80 =1	57	46	68	84	68	94
6CIT all	46	37	55	93	86	97
age80 =0	43	28	58	96	88	99
age80 =1	48	37	59	87	72	96
GPCOG all	55	46	64	91	84	95
age80 =0	53	38	68	94	86	98
age80 =1	56	45	67	84	69	94
Minicog all	42	34	51	94	87	97
age80 =0	38	25	54	96	88	99
age80 =1	45	34	56	89	75	97
Eurotest all	44	35	53	97	92	99
age80 =0	34	21	49	97	90	100
age80 =1	49	38	60	97	86	100
SPMT all	49	40	58	96	91	99
age80 =0	47	32	62	97	90	100
age80 =1	80	70	88	76	60	89
TUG all	57	48	66	89	81	94
age80 =0	50	35	65	94	86	98
age80 =1	57	46	68	84	69	94
Sniffin Sticks all	51	41	61	88	80	94
age80 =0	55	38	71	91	81	97
age80 =1	48	35	62	82	63	94
MAT all	63	41	81	90	55	100
age80 =0	0	0	71	100	48	100
age80 =1	74	49	91	60	15	95
MOCA all	55	45	64	87	77	93
age80 =0	0	0	8	100	93	100
age80 =1	53	40	65	84	66	95
FAQ all	64	55	72	88	80	93
age80 =0	13	5	26	99	92	100
age80 =1	66	55	76	82	66	92
ADL all	56	48	65	89	81	94
age80 =0	4	1	15	99	92	100
age80 =1	69	58	79	76	60	89
IADL all	56	47	65	89	81	94
age80 =0	55	40	70	91	82	97
age80 =1	56	45	67	84	69	94
AD8 all	55	46	63	92	85	96
age80 =0	55	40	70	94	86	98
age80 =1	54	43	65	87	72	96
IQCODE all	56	47	64	91	84	96
age80 =0	55	40	70	94	85	98
age80 =1	56	45	67	85	69	95

see Glossary for abbreviations of tests

Table J.4: Sensitivity analysis: sex, tests combined with GP judgement, target condition dementia

Test	Sensitivity	Sensitivity lb	Sensitivity ub	Specificity	Specificity lb	Specificity ub ‡
EPSS all	44	35	53	91	84	95
Female	63	50	74	94	83	99
Male	43	31	55	86	75	94
TAC all	67	58	75	85	77	91
Female	72	59	82	92	81	98
Male	18	9	29	100	94	100
Phototest all	37	29	46	94	88	98
Female	63	50	75	94	83	99
Male	37	25	49	93	83	98
6CIT all	46	37	55	93	86	97
Female	63	50	74	94	83	99
Male	39	27	51	89	78	96
GPCOG all	55	46	64	91	84	95
Female	63	50	74	96	86	100
Male	49	36	61	86	75	94
Minicod all	42	34	51	94	87	97
Female	53	40	66	98	89	100
Male	32	22	45	90	79	96
Eurotest all	44	35	53	97	92	99
Female	52	39	64	98	89	100
Male	37	25	49	97	88	100
SPMT all	49	40	58	96	91	99
Female	63	50	74	94	83	99
Male	44	32	57	97	88	100
TUG all	57	48	66	89	81	94
Female	64	51	76	94	83	99
Male	10	4	20	98	91	100
Sniffin Sticks all	51	41	61	88	80	94
Female	57	41	72	95	83	99
Male	0	0	7	100	92	100
MAT all	63	41	81	90	55	100
Female	0	0	71	100	40	100
Male	50	19	81	83	36	100
MOCA all	55	45	64	87	77	93
Female	60	45	74	92	79	98
Male	0	0	6	100	92	100
FAQ all	64	55	72	88	80	93
Female	67	54	78	92	81	98
Male	13	6	24	97	88	100
ADL all	56	48	65	89	81	94
Female	73	60	83	92	81	98
Male	9	3	18	98	91	100
IADL all	56	47	65	89	81	94
Female	63	50	74	94	83	99
Male	24	14	35	97	88	100
IQCODE all	56	47	64	91	84	96
Female	62	49	74	100	92	100
Male	0	0	5	100	94	100

see Glossary for abbreviations of tests

Table J.5: Sensitivity analysis: age, tests combined with GP judgement, target condition normal

Test	Sensitivity	Sensitivity lb	Sensitivity ub	Specificity	Specificity lb	Specificity ub ‡
EPSS all	13	5	26	98	96	100
age80 =0	12	3	27	98	92	100
age80 =1	15	2	45	99	95	100
TAC all	0	0	8	100	98	100
age80 =0	41	25	59	87	77	94
age80 =1	0	0	25	100	96	100
Phototest all	39	25	55	95	91	97
age80 =0	41	25	59	93	85	97
age80 =1	0	0	26	100	97	100
6CIT all	39	25	55	97	93	99
age80 =0	42	25	61	96	90	99
age80 =1	31	9	61	97	92	99
GPCOG all	30	17	45	97	94	99
age80 =0	35	20	54	96	90	99
age80 =1	15	2	45	98	94	100
Minicog all	38	25	54	97	93	99
age80 =0	41	25	59	96	90	99
age80 =1	31	9	61	97	92	99
Eurotest all	40	26	56	95	91	97
age80 =0	41	25	59	86	73	95
age80 =1	0	0	25	100	97	100
SPMT all	40	26	56	95	91	98
age80 =0	41	25	59	93	85	97
age80 =1	38	14	68	97	92	99
TUG all	40	26	56	94	89	97
age80 =0	41	25	59	89	80	95
age80 =1	38	14	68	96	89	99
Sniffin sticks all	7	2	20	99	96	100
age80 =0	44	26	62	88	78	95
age80 =1	0	0	37	100	95	100
MAT all	0	0	60	100	72	100
age80 =0	0	0	84	100	40	100
age80 =1	0	0	84	100	59	100
MOCA all	12	4	25	100	98	100
age80 =0	34	19	53	96	89	99
age80 =1	0	0	28	100	96	100
FAQ all	0	0	8	100	98	100
age80 =0	41	25	59	88	78	94
age80 =1	0	0	25	100	96	100
ADL all	0	0	8	100	98	100
age80 =0	41	25	59	88	78	94
age80 =1	0	0	25	100	95	100
IADL all	0	0	8	100	98	100
age80 =0	41	25	59	88	78	95
age80 =1	0	0	25	100	97	100
AD8 all	23	12	38	99	97	100
age80 =0	29	15	47	99	93	100
age80 =1	8	0	36	100	97	100
IQCODE all	31	18	47	99	96	100
age80 =0	33	18	52	98	91	100
age80 =1	25	5	57	100	97	100

see Glossary for abbreviations of tests

Table J.6: Sensitivity analysis: sex, tests combined with GP judgement, target condition normal

Test	Sensitivity	Sensitivity lb	Sensitivity ub	Specificity	Specificity lb	Specificity ub ‡
EPSS all	13	5	26	98	96	100
Female	17	5	37	98	92	100
Male	9	1	28	99	95	100
TAC all	0	0	8	100	98	100
Female	0	0	14	100	95	100
Male	43	23	66	93	84	97
Phototest all	39	25	55	95	91	97
Female	0	0	15	100	96	100
Male	39	20	61	97	92	99
6CIT all	39	25	55	97	93	99
Female	33	16	55	96	89	99
Male	45	24	68	98	93	100
GPCOG all	30	17	45	97	94	99
Female	83	63	95	90	82	95
Male	30	13	53	97	92	99
Minicog all	38	25	54	97	93	99
Female	33	16	55	97	91	99
Male	43	23	66	97	92	99
Eurotest all	40	26	56	95	91	97
Female	38	19	59	94	88	98
Male	43	23	66	90	79	97
SPMT all	40	26	56	95	91	98
Female	0	0	14	100	96	100
Male	43	23	66	98	93	100
TUG all	40	26	56	94	89	97
Female	0	0	14	100	96	100
Male	43	23	66	93	85	97
Sniffin sticks all	7	2	20	99	96	100
Female	0	0	17	100	94	100
Male	0	0	18	100	96	100
MAT all	0	0	60	100	72	100
Female	0	0	98	100	48	100
Male	0	0	71	100	54	100
MOCA all	12	4	25	100	98	100
Female	30	13	53	99	93	100
Male	10	1	32	100	96	100
FAQ all	0	0	8	100	98	100
Female	0	0	14	100	95	100
Male	43	23	66	93	85	97
ADL all	0	0	8	100	98	100
Female	0	0	14	100	95	100
Male	43	23	66	93	85	97
IADL all	0	0	8	100	98	100
Female	0	0	14	100	94	100
Male	43	23	66	95	89	98
AD8 all	23	12	38	99	97	100
Female	25	10	47	100	96	100
Male	22	7	44	99	95	100
IQCODE all	31	18	47	99	96	100
Female	35	16	57	98	92	100
Male	27	11	50	100	96	100

see Glossary for abbreviations of tests

Table J.7: Diagnostic accuracy of all GP 360 tests for target condition dementia

Test	Sensitivity	Sensitivity lb	Sensitivity ub	Specificity	Specificity lb	Specificity ub ‡
mat+faq	0	0	31	100	69	100
mat+adl	0	0	34	100	69	100
mat+iadl	0	0	34	100	69	100
tac+faq	32	24	40	98	93	100
tac+adl	27	20	36	98	93	100
euro+ad8	42	34	51	98	93	100
euro+iqcode	44	35	52	98	93	100
spmt+ad8	48	39	57	98	93	100
spmt+iqcode	49	40	58	98	93	100
sniffin+faq	11	6	19	98	92	100
euro+iadl	44	35	53	97	92	99
sniffin+iadl	21	14	31	97	90	99
photo+ad8	37	28	46	96	91	99
photo+iqcode	38	29	47	96	90	99
euro+faq	50	41	59	96	91	99
spmt+faq	54	45	62	96	91	99
spmt+adl	59	50	67	96	91	99
spmt+iadl	49	40	58	96	91	99
tug+faq	28	21	37	96	91	99
moca+iadl	23	16	32	95	88	99
tac+iadl	37	29	46	94	88	98
photo+faq	45	36	54	94	88	98
cit+ad8	44	36	53	94	88	98
cit+iqcode	45	37	54	94	87	98
minicog+faq	51	42	60	94	87	97
minicog+ad8	41	32	50	94	88	98
minicog+iq e	42	33	51	94	87	98
euro+adl	53	45	62	94	88	98
epss+ad8	42	34	51	93	86	97
photo+adl	44	35	53	93	87	97
photo+iadl	44	35	52	93	87	97
cit+iadl	46	37	55	93	86	97
gpcog+ad8	54	45	62	93	86	97
minicog+adl	55	46	64	93	86	97
epss+iqcode	44	35	52	92	85	96
cit+faq	56	47	64	92	85	96
cit+adl	58	49	66	92	85	96
gpcog+iqcode	55	46	64	92	85	96
minicog+iadl	48	39	57	92	85	96
tug+ad8	55	46	64	92	85	96
epss+adl	44	36	53	91	84	95
epss+iadl	45	37	54	91	84	95
gpcog+iadl	55	46	64	91	84	95
tug+iqcode	57	48	65	91	84	96
sniffin+ad8	49	39	59	91	82	96
epss+faq	54	45	62	90	83	95
tac+ad8	67	58	75	90	83	95
gpcog+faq	62	53	70	90	83	95
sniffin+iq e	51	40	61	90	81	96
mat+ad8	63	41	81	90	55	100
moca+ad8	53	43	62	90	82	96
tac+iqcode	66	57	74	89	81	94
gpcog+adl	66	58	74	89	81	94
tug+iadl	57	48	66	89	81	94
moca+iqcode	54	44	64	89	80	95
tug+adl	62	53	70	88	80	93
sniffin+adl	52	41	62	88	80	94
moca+faq	55	45	64	87	77	93
moca+adl	55	45	65	87	77	93
mat+iqcode	71	49	87	71	29	96

see Glossary for abbreviations of tests

Table J.8: Diagnostic accuracy of all GP 360 tests for target condition normal

Test	Sensitivity	Sensitivity lb	Sensitivity ub	Specificity	Specificity lb	Specificity ub ‡
tac+faq	0	0	8	100	97	100
tac+adl	0	0	8	100	97	100
tac+iadl	0	0	8	100	98	100
mat+faq	0	0	60	100	69	100
mat+adl	0	0	60	100	66	100
mat+iadl	0	0	60	100	66	100
mat+ad8	0	0	60	100	66	100
mat+iqcode	0	0	60	100	69	100
moca+faq	0	0	9	100	97	100
tac+ad8	23	12	38	99	96	100
tac+iqcode	31	18	47	99	95	100
photo+ad8	24	13	39	99	97	100
photo+iqcode	32	19	48	99	96	100
cit+iqcode	32	19	48	99	97	100
euro+ad8	23	12	38	99	97	100
euro+iqcode	31	18	47	99	96	100
spmt+ad8	23	12	38	99	97	100
spmt+iqcode	31	18	47	99	96	100
tug+ad8	23	12	38	99	97	100
tug+iqcode	31	18	47	99	96	100
sniffin+faq	7	2	20	99	96	100
sniffin+adl	7	2	20	99	95	100
sniffin+iadl	5	1	17	99	96	100
sniffin+ad8	29	16	46	99	95	100
sniffin+iq e	31	17	48	99	96	100
moca+ad8	40	25	56	99	96	100
moca+iqcode	44	28	60	99	95	100
epss+faq	13	5	26	98	94	100
epss+adl	13	5	26	98	94	100
epss+iadl	13	5	26	98	96	100
epss+ad8	30	17	45	98	95	99
gpcog+iadl	30	17	45	98	95	99
gpcog+iqcode	31	18	47	98	95	99
moca+iadl	26	14	41	98	95	100
epss+iqcode	40	26	56	97	94	99
cit+adl	39	25	55	97	92	99
cit+iadl	39	25	55	97	93	99
cit+ad8	39	25	55	97	93	99
gpcog+faq	30	17	45	97	93	99
gpcog+adl	30	17	45	97	93	99
gpcog+ad8	30	17	45	97	94	99
minicog+adl	38	25	54	97	93	99
minicog+iadl	38	25	54	97	94	99
minicog+ad8	38	25	54	97	93	99
minicog+iq e	40	26	56	97	94	99
moca+adl	26	14	41	97	92	99
photo+iadl	39	25	55	96	92	98
cit+faq	39	25	55	96	92	99
minicog+faq	38	25	54	96	92	99
spmt+iadl	40	26	56	96	92	98
euro+iadl	40	26	56	95	91	98
spmt+adl	40	26	56	95	90	98
photo+faq	39	25	55	94	88	97
photo+adl	39	25	55	94	89	97
euro+faq	40	26	56	94	88	97
euro+adl	40	26	56	94	89	97
spmt+faq	40	26	56	94	89	97
tug+iadl	40	26	56	94	90	97
tug+faq	40	26	56	92	87	96
tug+adl	40	26	56	92	87	96

see Glossary for abbreviations of tests

APPENDIX K

INDEX TEST POSITION IN THE BATTERY

This appendix provides tabulated figures of the position of each index test in the index test battery.

Table K.1: Test order in index test battery

	Number (%) of tests in each test order position									
Test order position	**MAT or MOCA**	**Eurotest**	**TAC**	**Phototest**	**SPMT**	**6CIT**	**GPCOG**	**EPSS**	**TUG**	**Sniffin Sticks**
First	36 (15)	16 (7)	33 (14)	22 (9)	25 (10)	30 (13)	30 (13)	30 (13)	18 (8)	()
Second	24 (10)	21 (9)	26 (11)	24 (10)	31 (13)	25 (10)	35 (15)	36 (15)	18 (8)	0 (0)
Third	21 (9)	29 (12)	30 (13)	37 (16)	26 (11)	19 (8)	26 (11)	26 (11)	26 (11)	0 (0)
Fourth	30 (13)	27 (11)	21 (9)	34 (14)	23 (10)	32 (13)	17 (7)	21 (9)	34 (14)	0 (0)
Fifth	24 (10)	27 (11)	31 (13)	22 (9)	29 (12)	32 (13)	27 (11)	24 (10)	23 (10)	0 (0)
Sixth	26 (11)	26 (11)	24 (10)	28 (12)	32 (13)	23 (10)	28 (12)	22 (9)	29 (12)	0 (0)
Seventh	30 (13)	30 (13)	27 (11)	19 (8)	21 (9)	29 (12)	25 (10)	31 (13)	27 (11)	0 (0)
Eighth	24 (10)	27 (11)	33 (14)	30 (13)	30 (13)	22 (9)	18 (8)	20 (8)	35 (15)	0 (0)
Ninth	24 (10)	36 (15)	15 (6)	22 (9)	22 (9)	28 (12)	33 (14)	29 (12)	29 (12)	0 (0)
Tenth	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)	189 (100)

see Glossary for abbreviations of tests

APPENDIX L

PRISMA-DTA CHECKLIST

This appendix includes the Prisma-DTA checklist for reporting systematic reviews available at http://prisma-statement.org.

PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	49
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	i
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	49-50
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	49-50
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	50
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	49
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	50-52
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	53
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	Appendix A
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	53-54
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	54
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	53-54
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	54
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	54-55
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	54-55

PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	54-55
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	56
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	57
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	59-63
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	63-65
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	66-70
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	66,67,69
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	71-80
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	81-83
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	81,83-85
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	86-92
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	iii

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

For more information, visit: **www.prisma-statement.org**.

BIBLIOGRAPHY

1. Burns, A and Iliffe, S.
   Dementia.
   BMJ (Clinical research ed.) 2009;338:b75.

2. American Psychiatric Association.
   Diagnostic and Statistical Manual of Mental Disorders.
   DSM-V.
   Washington, DC: Author, 2013.

3. Gauthier, S, Reisberg, B, Zaudig, M, et al.
   Mild cognitive impairment.
   The Lancet 2006;367:1262–70.

4. American Psychiatric Association.
   Diagnostic and Statistical Manual of Mental Disorders.
   DSM-IV-TR.
   Washington, DC: Author, 2000.

5. World Health Organization.
   The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research.
   Geneva: World Health Organisation, 1993.

6. American psychiatric Association.
   Diagnostic and statistical manual of mental disorders.
   DSM-III-R.
   Washington, DC: Author, 1987.

7. World Health Organization.
   The ICD-11 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.
   Geneva: World Health Organisation, 2018.

8. Henderson, AS, Jorm, AF, Mackinnon, A, et al.
   A survey of dementia in the Canberra population: experience with ICD-10 and DSM-III-R criteria.
   Psychological Medicine 1994;24:473.

9. Fichter, MM, Meller, I, Schröppel, H, and Steinkirchner, R.
Dementia and cognitive impairment in the oldest old in the community. Prevalence and comorbidity.
British Journal of Psychiatry 1995;166:621–9.

10. Erkinjuntti, T, Østbye, T, Steenhuis, R, and Hachinski, V.
The Effect of Different Diagnostic Criteria on the Prevalence of Dementia.
New England Journal of Medicine 1997;337:1667–1674.

11. Wancata, J, Börjesson-Hanson, A, Östling, S, Sjögren, K, and Skoog, I.
Diagnostic Criteria Influence Dementia Prevalence.
The American Journal of Geriatric Psychiatry 2007;15:1034–1045.

12. Chagas, MHN, Pessoa, RMP, and Almeida, OP.
Comparison of DSM-IV and DSM-5 dementia criteria among older people living in a community sample.
International Journal of Geriatric Psychiatry 2018;33:801–802.

13. Riedel-Heller, SG, Busse, A, Aurich, C, Matschinger, H, and Angermeyer, MC.
Prevalence of dementia according to DSM–III–R and ICD–10.
British Journal of Psychiatry 2001;179:250–254.

14. Prince, M, Wimo, A, Guerchet, M, Gemma-Claire, A, Wu, YT, and Prina, M.
World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost And trends.
Tech. rep.
London: Alzheimer's Disease International, 2015:
84.
URL: http://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf.

15. Prince, M, Ali, GC, Guerchet, M, Prina, AM, Albanese, E, and Wu, YT.
Recent global trends in the prevalence and incidence of dementia, and survival with dementia.
Alzheimer's Research & Therapy 2016;8:23.

16. Satizabal, CL, Beiser, AS, Chouraki, V, Chêne, G, Dufouil, C, and Seshadri, S.
Incidence of Dementia over Three Decades in the Framingham Heart Study.
New England Journal of Medicine 2016;374:523–532.

17. Matthews, FE, Stephan, BCM, Robinson, L, et al.
A two decade dementia incidence comparison from the Cognitive Function and Ageing Studies I and II.
Nature Communications 2016;7:11398.

18. Gao, S, Ogunniyi, A, Hall, KS, et al.
Dementia incidence declined in African-Americans but not in Yoruba.

Alzheimer's & Dementia 2016;12:244–251.

19. Chan, KY, Wang, W, Wu, JJ, et al.
Epidemiology of Alzheimer's disease and other forms of dementia in China, 1990-2010: a systematic review and analysis.
Lancet (London, England) 2013;381:2016–23.

20. Dodge, HH, Buracchio, TJ, Fisher, GG, et al.
Trends in the Prevalence of Dementia in Japan.
International Journal of Alzheimer's Disease 2012;2012:1–11.

21. Wu, YT, Brayne, C, and Matthews, FE.
Prevalence of dementia in East Asia: a synthetic review of time trends.
International Journal of Geriatric Psychiatry 2015;30:793–801.

22. Bussel, EF van, Richard, E, Arts, DL, et al.
Dementia incidence trend over 1992-2014 in the Netherlands: Analysis of primary care data.
PLOS Medicine 2017;14. Ed. by Miller, BL:e1002235.

23. Primary Care Domain NHS Digital.
Recorded Dementia Diagnoses - October 2018.
Tech. rep.
2018.
URL: https://digital.nhs.uk/data-and-information/publications/statistical/recorded-dementia-diagnoses/october-2018.

24. Pérès, K, Brayne, C, Matharan, F, et al.
Trends in Prevalence of Dementia in French Farmers from Two Epidemiological Cohorts.
Journal of the American Geriatrics Society 2017;65:415–420.

25. Patterson, C.
World Alzheimer Report 2018 - The state of the art of dementia research: New frontiers.
Tech. rep.
London: Alzheimer's Disease International (ADI), 2018:
46.
URL: https://www.alz.co.uk/research/WorldAlzheimerReport2018.pdf.

26. Ahmadi-Abhari, S, Guzman-Castillo, M, Bandosz, P, et al.
Temporal trend in dementia incidence since 2002 and projections for prevalence in England and Wales to 2040: modelling study.
BMJ (Clinical research ed.) 2017;358:j2856.

27. Brayne, C and Davis, D.
Making Alzheimer's and dementia research fit for populations.
The Lancet 2012;380:1441–3.

28. Haroutunian, V, Schnaider-Beeri, M, Schmeidler, J, et al.
Role of the neuropathology of Alzheimer disease in dementia in the oldest-old.
Archives of neurology 2008;65:1211–7.

29. Brayne, C, Richardson, K, Matthews, FE, et al.
Neuropathological correlates of dementia in over-80-year-old brain donors from the population-based Cambridge city over-75s cohort (CC75C) study.
Journal of Alzheimer's disease : JAD 2009;18:645–58.

30. Matthews, FE, Brayne, C, Lowe, J, McKeith, I, Wharton, SB, and Ince, P.
Epidemiological pathology of dementia: attributable-risks at death in the Medical Research Council Cognitive Function and Ageing Study.
PLoS medicine 2009;6:e1000180.

31. Nelson, PT, Jicha, GA, Kryscio, RJ, et al.
Low sensitivity in clinical diagnoses of dementia with Lewy bodies.
Journal of Neurology 2010;257:359–366.

32. Beach, TG, Monsell, SE, Phillips, LE, and Kukull, W.
Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010.
Journal of neuropathology and experimental neurology 2012;71:266–73.

33. Savva, GM, Wharton, SB, Ince, PG, Forster, G, Matthews, FE, and Brayne, C.
Age, neuropathology, and dementia.
The New England journal of medicine 2009;360:2302–9.

34. Scheltens, P and Rockwood, K.
How golden is the gold standard of neuropathology in dementia?
Alzheimer's and Dementia 2011;7:486–489.

35. Walker, Z, Possin, KL, Boeve, BF, and Aarsland, D.
Lewy body dementias.
The Lancet 2015;386:1683–1697.

36. Bang, J, Spina, S, and Miller, BL.
Frontotemporal dementia.
The Lancet 2015;386:1672–1682.

37. Tola-Arribas, MA, Yugueros, MI, Garea, MJ, et al.
Prevalence of Dementia and Subtypes in Valladolid, Northwestern Spain: The DEMINVALL Study.
PLoS ONE 2013;8. Ed. by Ikram, MA:e77688.

38. Scheltens, P, Blennow, K, Breteler, MMB, et al.
Alzheimer's disease.

The Lancet 2016;1:15056.

39. O'Brien, JT and Thomas, A.
Vascular dementia.
The Lancet 2015;386:1698–1706.

40. Stevens, T, Livingston, G, Kitchen, G, Manela, M, Walker, Z, and Katona, C.
Islington study of dementia subtypes in the community.
British Journal of Psychiatry 2002;180:270–276.

41. Lim, A, Tsuang, D, Kukull, W, et al.
Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series.
Journal of the American Geriatrics Society 1999;47:564–9.

42. Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS).
Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales.
The Lancet 2001;357:169–175.

43. White, L, Small, BJ, Petrovitch, H, et al.
Recent clinical-pathologic research on the causes of dementia in late life: Update from the Honolulu-Asia Aging Study.
Journal of Geriatric Psychiatry and Neurology 2005;18:224–227.

44. Schneider, Ja, Arvanitakis, Z, Bang, W, and Bennett, Da.
Mixed brain pathologies account for most dementia cases in community-dwelling older persons.
Neurology 2007;69:2197–204.

45. Kawas, CH, Kim, RC, Sonnen, Ja, Bullain, SS, and Trieu, T.
Multiple pathologies are common and related to dementia in the oldest-old.
2015:535–542.

46. Ritchie, K, Carriere, I, Ritchie, CW, Berr, C, Artero, S, and Ancelin, ML.
Designing prevention programmes to reduce incidence of dementia: prospective cohort study of modifiable risk factors.
BMJ 2010;341:c3885–c3885.

47. Gatz, M, Fiske, A, Reynolds, CA, Wetherell, JL, Johansson, B, and Pedersen, NL.
Sex differences in genetic risk for dementia.
Behavior genetics 2003;33:95–105.

48. Lautenschlager, NT, Cupples, LA, Rao, VS, et al.

Risk of dementia among relatives of Alzheimer's disease patients in the MIRAGE study: What is in store for the oldest old?
Neurology 1996;46:641–50.

49. Neu, SC, Pa, J, Kukull, W, et al.
Apolipoprotein E Genotype and Sex Risk Factors for Alzheimer Disease.
JAMA Neurology 2017;74:1178.

50. Sleegers, K, Roks, G, Theuns, J, et al.
Familial clustering and genetic risk for dementia in a genetically isolated Dutch population.
Brain 2004;127:1641–1649.

51. Livingston, G, Sommerlad, A, Orgeta, V, et al.
Dementia prevention, intervention, and care.
The Lancet 2017;390:2673–2734.

52. Norton, S, Matthews, FE, Barnes, DE, Yaffe, K, and Brayne, C.
Potential for primary prevention of Alzheimer's disease: an analysis of population-based data.
The Lancet. Neurology 2014;13:788–94.

53. Nguyen, TT, Tchetgen, EJT, Kawachi, I, et al.
Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk.
Annals of Epidemiology 2016;26:71–76.

54. Seblova, D, Fischer, M, Fors, S, et al.
Is There a Direct Causal Effect of Education on Dementia? A Swedish Natural Experiment on 1.3 Million Individuals.
URL: https://ssrn.com/abstract=3320156.

55. Kuźma, E, Hannon, E, Zhou, A, et al.
Which Risk Factors Causally Influence Dementia? A Systematic Review of Mendelian Randomization Studies.
Journal of Alzheimer's Disease 2018;64:181–193.

56. Anderson, EL, Howe, LD, Wade, KH, et al.
Education, intelligence and Alzheimer's disease: Evidence from a multivariable two-sample Mendelian randomization study.
bioRxiv 2018:401042.

57. Stern, Y.
What is cognitive reserve? Theory and research application of the reserve concept.
2002.

58. Clare, L, Wu, YT, Teale, JC, et al.

Potentially modifiable lifestyle factors, cognitive reserve, and cognitive function in later life: A cross-sectional study.
PLoS medicine 2017;14:e1002259.

59. Loy, CT, Schofield, PR, Turner, AM, and Kwok, JBJ.
Genetics of dementia.
Lancet 2014;383:828–40.

60. Barnes, J, Dickerson, BC, Frost, C, Jiskoot, LC, Wolk, D, and Van Der Flier, WM.
Alzheimer's disease first symptoms are age dependent: Evidence from the NACC dataset.
Alzheimer's and Dementia 2015;11:1349–1357.

61. Crutch, SJ, Lehmann, M, Schott, JM, Rabinovici, GD, Rossor, MN, and Fox, NC.
Posterior cortical atrophy.
The Lancet Neurology 2012;11:170–178.

62. McKhann, G, Drachman, D, Folstein, M, Katzman, R, Price, D, and Stadlan, EM.
Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease.
Neurology 1984;34:939–44.

63. 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.
Alzheimer's & dementia : the journal of the Alzheimer's Association 2011;7:263–9.

64. Winikates, J and Jankovic, J.
Clinical Correlates of Vascular Parkinsonism.
Archives of Neurology 1999;56:98.

65. Román, GC, Tatemichi, TK, Erkinjuntti, T, et al.
Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop.
Neurology 1993;43:250–60.

66. Neary, D, Snowden, JS, Gustafson, L, et al.
Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria.
Neurology 1998;51:1546–54.

67. Brun, A, Englund, B, Gustafson, L, et al.
Clinical and neuropathological criteria for frontotemporal dementia. The Lund and Manchester Groups.
Journal of neurology, neurosurgery, and psychiatry 1994;57:416–8.

68. Boxer, AL and Miller, BL.

Clinical features of frontotemporal dementia.
Alzheimer disease and associated disorders 2005;19 Suppl 1:3–6.

69. Chui, HC, Victoroff, JI, Margolin, D, Jagust, W, Shankle, R, and Katzman, R.
Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's Disease Diagnostic and Treatment Centers.
Neurology 1992;42:473–80.

70. McKeith, IG, Galasko, D, Kosaka, K, et al.
Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop.
Neurology 1996;47:1113–24.

71. McKeith, IG, Dickson, DW, Lowe, J, et al.
Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium.
Neurology 2005;65:1863–72.

72. Larner, aJ.
Neurological signs of possible diagnostic value in the cognitive disorders clinic.
Practical neurology 2014:1–4.

73. Hughes, CP, Berg, L, Danziger, WL, Coben, LA, and Martin, RL.
A new clinical scale for the staging of dementia.
The British journal of psychiatry : the journal of mental science 1982;140:566–72.

74. Manning, CA and Ducharme, JK.
Dementia Syndromes in the Older Adult.
Handbook of Assessment in Clinical Gerontology 2010:155–178.

75. Morris, JC.
The Clinical Dementia Rating (CDR): current version and scoring rules.
Neurology 1993;43:2412–4.

76. Lynch, C, Walsh, C, Blanco, A, et al.
The Clinical Dementia Rating Sum of Box Score in Mild Dementia.
Dementia and Geriatric Cognitive Disorders 2006;21:40–43.

77. O'Bryant, SE, Waring, SC, Cullum, CM, et al.
Staging Dementia Using Clinical Dementia Rating Scale Sum of Boxes Scores.
Archives of Neurology 2008;65:1091.

78. O'Bryant, SE, Lacritz, LH, Hall, J, et al.
Validation of the New Interpretive Guidelines for the Clinical Dementia Rating Scale Sum of Boxes Score in the National Alzheimer's Coordinating Center Database.
Archives of Neurology 2010;67:746–749.

79. Reisberg, B, Ferris, SH, Leon, MJ de, and Crook, T.
The Global Deterioration Scale for assessment of primary degenerative dementia.
American Journal of Psychiatry 1982;139:1136–1139.

80. Jack, CR, Knopman, DS, Jagust, WJ, et al.
Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade.
The Lancet. Neurology 2010;9:119–28.

81. Sperling, RA, Aisen, PS, Beckett, LA, et al.
Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease.
Alzheimer's & dementia : the journal of the Alzheimer's Association 2011;7:280–92.

82. Cooper, C, Bebbington, P, Lindesay, J, et al.
The meaning of reporting forgetfulness: a cross-sectional study of adults in the English 2007 Adult Psychiatric Morbidity Survey.
Age and Ageing 2011;40:711–717.

83. Mitchell, AJ, Beaumont, H, Ferguson, D, Yadegarfar, M, and Stubbs, B.
Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: meta-analysis.
Acta Psychiatrica Scandinavica 2014;130:439–451.

84. Mitchell, AJ.
The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: a meta-analysis.
International journal of geriatric psychiatry 2008;23:1191–202.

85. Ford, E, Greenslade, N, Paudyal, P, et al.
Predicting dementia from primary care records: A systematic review and meta-analysis.
PLOS ONE 2018;13:e0194735.

86. Tang, EYH, Harrison, SL, Errington, L, et al.
Current developments in dementia risk prediction modelling: An updated systematic review.
PLoS ONE 2015;10:1–31.

87. Fisher, S, Hsu, A, Mojaverian, N, et al.
Dementia Population Risk Tool (DemPoRT): study protocol for a predictive algorithm assessing dementia risk in the community.
BMJ open 2017;7:e018018.

88. Li, J, Ogrodnik, M, Devine, S, Auerbach, S, Wolf, PA, and Au, R.
Practical risk score for 5-, 10-, and 20-year prediction of dementia in elderly persons: Framingham Heart Study.

Alzheimer's & Dementia 2018;14:35–42.

89. Jessen, F, Wiese, B, Bickel, H, et al.
Prediction of dementia in primary care patients.
PloS one 2011;6:e16852.

90. 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 Medicine 2016;14:6.

91. Licher, S, Leening, MJ, Yilmaz, P, et al.
Development and Validation of a Dementia Risk Prediction Model in the General Population: An Analysis of Three Longitudinal Studies.
American Journal of Psychiatry 2018:appi.ajp.2018.1.

92. Bhome, R, Berry, AJ, Huntley, JD, and Howard, RJ.
Interventions for subjective cognitive decline: systematic review and meta-analysis.
BMJ open 2018;8:e021610.

93. Petersen, RC.
Mild cognitive impairment as a diagnostic entity.
Journal of internal medicine 2004;256:183–94.

94. Winblad, B, Palmer, K, Kivipelto, M, et al.
Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment.
Journal of internal medicine 2004;256:240–246.

95. Petersen, RC, Caracciolo, B, Brayne, C, Gauthier, S, Jelic, V, and Fratiglioni, L.
Mild cognitive impairment: A concept in evolution.
Journal of Internal Medicine 2014;275:214–228.

96. Petersen, RC, Lopez, O, Armstrong, MJ, et al.
Practice guideline update summary: Mild cognitive impairment.
Neurology 2018;90:126–135.

97. Mitchell, aJ and Shiri-Feshki, M.
Rate of progression of mild cognitive impairment to dementia - Meta-analysis of 41 robust inception cohort studies.
Acta Psychiatrica Scandinavica 2009;119:252–265.

98. Kaduszkiewicz, H, Eisele, M, Wiese, B, et al.
Prognosis of mild cognitive impairment in general practice: results of the German AgeCoDe study.
Annals of family medicine 2014;12:158–65.

99. Aerts, L, Heffernan, M, Kochan, NA, et al.
Effects of MCI subtype and reversion on progression to dementia in a community sample.
Neurology 2017:10.1212/WNL.0000000000004015.

100. Mitchell, SL.
Advanced Dementia.
New England Journal of Medicine 2015;372:2533–2540.

101. Mitchell, SL, Shaffer, ML, Loeb, MB, et al.
Infection management and multidrug-resistant organisms in nursing home residents with advanced dementia.
JAMA Internal Medicine 2014;174:1660–1667.

102. Mitchell, SL, Teno, JM, Kiely, DK, et al.
The clinical course of advanced dementia.
The New England journal of medicine 2009;361:1529–38.

103. Hendriks, SA, Smalbrugge, M, Hertogh, CM, and Steen, JT van der.
Dying With Dementia: Symptoms, Treatment, and Quality of Life in the Last Week of Life.
Journal of Pain and Symptom Management 2014;47:710–720.

104. Fleming, J, Calloway, R, Perrels, A, Farquhar, M, Barclay, S, and Brayne, C.
Dying comfortably in very old age with or without dementia in different care settings – a representative “older old” population study.
BMC Geriatrics 2017;17:222.

105. Wolfson, C, Wolfson, DB, Asgharian, M, et al.
A Reevaluation of the Duration of Survival after the Onset of Dementia.
New England Journal of Medicine 2001;344:1111–1116.

106. Xie, J, Brayne, C, Matthews, FE, Function, MRCC, Ageing Study collaborators, tMRCCF, and Study, A.
Survival times in people with dementia: analysis from population based cohort study with 14 year follow-up.
BMJ (Clinical research ed.) 2008;336:258–62.

107. Rait, G, Walters, K, Bottomley, C, Petersen, I, Iliffe, S, and Nazareth, I.
Survival of people with clinical diagnosis of dementia in primary care: cohort study.
Bmj 2010;341:c3584.

108. 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;3:e004023.

109. Pimouguet, C, Delva, F, Le Goff, M, et al.

Survival and early recourse to care for dementia: A population based study.
Alzheimer's & Dementia 2015;11:385–393.

110. Fong, TG, Davis, D, Growdon, ME, Albuquerque, A, and Inouye, SK.
The interface between delirium and dementia in elderly adults.
The Lancet. Neurology 2015;14:823–832.

111. Marcantonio, ER, Ngo, LH, O'Connor, M, et al.
3D-CAM: Derivation and Validation of a 3-Minute Diagnostic Interview for CAM-Defined Delirium: A Cross-sectional Diagnostic Test Study.
Annals of internal medicine 2014;161:554–561.

112. Bellelli, G, Morandi, A, Davis, DHJ, et al.
Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people.
Age and Ageing 2014;43:496–502.

113. Richardson, SJ, Davis, DHJ, Bellelli, G, et al.
Detecting delirium superimposed on dementia: diagnostic accuracy of a simple combined arousal and attention testing procedure.
International Psychogeriatrics 2017;29:1585–1593.

114. Davis, DH, Skelly, DT, Murray, C, et al.
Worsening cognitive impairment and neurodegenerative pathology progressively increase risk for delirium.
American Journal of Geriatric Psychiatry 2015;23:403–415.

115. Davis, DHJ, Barnes, LE, Stephan, BCM, et al.
The descriptive epidemiology of delirium symptoms in a large population-based cohort study: results from the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS).
BMC geriatrics 2014;14:87.

116. Richardson, SJ, Davis, DHJ, Stephan, B, et al.
Protocol for the Delirium and Cognitive Impact in Dementia (DECIDE) study: A nested prospective longitudinal cohort study.
BMC geriatrics 2017;17:98.

117. Davis, D, Richardson, S, Hornby, J, et al.
The delirium and population health informatics cohort study protocol: ascertaining the determinants and outcomes from delirium in a whole population.
BMC geriatrics 2018;18:45.

118. Davis, DH, Muniz Terrera, G, Keage, H, et al.
Delirium is a strong risk factor for dementia in the oldest-old: A population-based cohort study.

Brain 2012;135:2809–2816.

119. Davis, DHJ, Muniz-Terrera, G, Keage, HAD, et al.
Association of Delirium With Cognitive Decline in Late Life.
JAMA Psychiatry 2017;74:244.

120. Morandi, A, Davis, D, Bellelli, G, et al.
The Diagnosis of Delirium Superimposed on Dementia: An Emerging Challenge.
Journal of the American Medical Directors Association 2017;18:12–18.

121. Gual, N, Richardson, SJ, Davis, DHJ, et al.
Impairments in balance and mobility identify delirium in patients with comorbid dementia.
International Psychogeriatrics 2018:1–5.

122. European Delirium Association and American Delirium Society.
The DSM-5 criteria, level of arousal and delirium diagnosis: inclusiveness is safer.
BMC medicine 2014;12:141.

123. Department of Health.
G8 Dementia summit declaration.
Tech. rep.
London, 2013.
URL: https://www.gov.uk/government/publications/g8-dementia-summit-agreements/g8-dementia-summit-declaration.

124. Wu, YT, Fratiglioni, L, Matthews, FE, et al.
Dementia in western Europe: epidemiological evidence and implications for policy making.
The Lancet Neurology 2016;15:116–124.

125. Steen, JT van der, Soest-Poortvliet, MC van, Wouden, JC van der, Bruinsma, MS, Scholten, RJ, and Vink, AC.
Music-based therapeutic interventions for people with dementia.
The Cochrane database of systematic reviews 2017;5:CD003477.

126. Woods, B, O'Philbin, L, Farrell, EM, Spector, AE, and Orrell, M.
Reminiscence therapy for dementia.
The Cochrane database of systematic reviews 2018;3:CD001120.

127. Forbes, D, Forbes, SC, Blake, CM, Thiessen, EJ, and Forbes, S.
Exercise programs for people with dementia.
The Cochrane database of systematic reviews 2015:CD006489.

128. Woods, B, Aguirre, E, Spector, AE, and Orrell, M.
Cognitive stimulation to improve cognitive functioning in people with dementia.
The Cochrane database of systematic reviews 2012:CD005562.

129. Liu, Z, Sun, YY, and Zhong, Bl.

Mindfulness-based stress reduction for family carers of people with dementia.
The Cochrane database of systematic reviews 2018;8:CD012791.

130. Lins, S, Hayder-Beichel, D, Rücker, G, et al.
Efficacy and experiences of telephone counselling for informal carers of people with dementia.
The Cochrane database of systematic reviews 2014:CD009126.

131. Maayan, N, Soares-Weiser, K, and Lee, H.
Respite care for people with dementia and their carers.
The Cochrane database of systematic reviews 2014:CD004396.

132. Reilly, S, Miranda-Castillo, C, Malouf, R, et al.
Case management approaches to home support for people with dementia.
The Cochrane database of systematic reviews 2015;1:CD008345.

133. Richter, T, Meyer, G, Möhler, R, and Köpke, S.
Psychosocial interventions for reducing antipsychotic medication in care home residents.
The Cochrane database of systematic reviews 2012;12:CD008634.

134. Flynn, E, Smith, CH, Walsh, CD, and Walshe, M.
Modifying the consistency of food and fluids for swallowing difficulties in dementia.
The Cochrane database of systematic reviews 2018;9:CD011077.

135. Francis, PT, Palmer, AM, Snape, M, and Wilcock, GK.
The cholinergic hypothesis of Alzheimer's disease: a review of progress.
Journal of Neurology, Neurosurgery & Psychiatry 1999;66:137–147.

136. Birks, JS and Harvey, RJ.
Donepezil for dementia due to Alzheimer's disease.
The Cochrane database of systematic reviews 2018;6:CD001190.

137. Loy, C and Schneider, L.
Galantamine for Alzheimer's disease and mild cognitive impairment.
The Cochrane database of systematic reviews 2006:CD001747.

138. Birks, JS and Grimley Evans, J.
Rivastigmine for Alzheimer's disease.
The Cochrane database of systematic reviews 2015:CD001191.

139. Birks, J.
Cholinesterase inhibitors for Alzheimer's disease.
The Cochrane database of systematic reviews 2006:CD005593.

140. Croft, P, Dinant, GJ, Coventry, P, and Barraclough, K.
Looking to the future: Should 'prognosis' be heard as often as 'diagnosis' in medical education?
Education for Primary Care 2015;26:367–371.

141. Porta, M, ed.
A Dictionary of Epidemiology.
Sixth edit.
Oxford: Oxford University Press, 2008:
376.

142. Heneghan, C, Glasziou, P, Thompson, M, et al.
Diagnostic strategies used in primary care.
BMJ (Clinical research ed.) 2009;338:b946.

143. Pauker, SG and Kassirer, JP.
The Threshold Approach to Clinical Decision Making.
New England Journal of Medicine 1980;302:1109–1117.

144. Kahneman, D.
A perspective on judgment and choice: mapping bounded rationality - London School of Economics and Political Science.
1996:1–35.

145. Balla, JI, Heneghan, C, Glasziou, P, Thompson, M, and Balla, ME.
A model for reflection for good clinical practice.
Journal of Evaluation in Clinical Practice 2009;15:964–969.

146. Knottnerus, JA, Weel, CV, and Muris, JWM.
Evaluation of diagnostic procedures.
Bmj 2002;324:1391.

147. Irving, G and Holden, J.
The time-efficiency principle: time as the key diagnostic strategy in primary care.
Family practice 2013;30:386–9.

148. Knottnerus, JA and Buntinx, F, eds.
The Evidence Base of Clinical Diagnosis.
Second edi.
Oxford, UK: Wiley-Blackwell, 2008:
316.
DOI: 10.1002/9781444300574.
URL: http://core.ac.uk/download/pdf/11357358.pdf.

149. Youden, W.
Index for rating diagnostic tests.
Cancer 1950;3:32–5.

150. Macaskill, P, Gatsonis, C, Deeks, J, Harbord, R, and Takwoingi, Y.
Analysing and Presenting Results.

In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*.
Ed. by Deeks, J, Bossuyt, P, and Gatsonis, C.
First edit.
The Cochrane Collaboration, 2010.
Chap. 10.
URL: `http://srdta.cochrane.org`.

151. Metz, CE.
Basic principles of ROC analysis.
Seminars in Nuclear Medicine 1978;8:283–298.

152. Lalkhen, AG and McCluskey, A.
Clinical tests: sensitivity and specificity.
Continuing Education in Anaesthesia Critical Care & Pain 2008;8:221–223.

153. Begg, CB.
Biases in the assessment of diagnostic tests.
Statistics in medicine 1987;6:411–23.

154. Leeflang, MMG, Rutjes, AWS, Reitsma, JB, Hooft, L, and Bossuyt, PM.
Variation of a test's sensitivity and specificity with disease prevalence.
CMAJ : Canadian Medical Association journal journal de l'Association medicale canadienne 2013;185:537–44.

155. Elder, A, Japp, A, and Verghese, A.
How valuable is physical examination of the cardiovascular system?
Bmj 2016:i3309.

156. Folstein, MF, Folstein, SE, and McHugh, PR.
"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.
Journal of psychiatric research 1975;12:189–98.

157. 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.
The Cochrane database of systematic reviews 2016;1:CD011145.

158. Steyerberg, EW, Vickers, AJ, Cook, NR, et al.
Assessing the Performance of Prediction Models.
Epidemiology 2010;21:128–138.

159. Harrell, FE, Califf, RM, Pryor, DB, Lee, KL, and Rosati, RA.
Evaluating the yield of medical tests.
JAMA 1982;247:2543–6.

160. Uno, H, Cai, T, Pencina, MJ, D'Agostino, RB, and Wei, LJ.
On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data.
Statistics in medicine 2011;30:1105–17.

161. Royston, P and Sauerbrei, W.
A new measure of prognostic separation in survival data.
Statistics in medicine 2004;23:723–48.

162. Ensor, J, Snell, KI, and Martin, E.
PMCALPLOT Stata module to produce calibration plot of prediction model performance.
Boston College Department of Economics, 2018.

163. Stevens, RJ and Poppe, KK.
Validation of clinical prediction models: what does the "calibration slope" really measure?
Journal of Clinical Epidemiology 2019.

164. Hanley, JA and McNeil, BJ.
The meaning and use of the area under a receiver operating characteristic (ROC) curve.
Radiology 1982;143:29–36.

165. Mallett, S, Halligan, S, Thompson, M, Collins, GS, and Altman, DG.
Interpreting diagnostic accuracy studies for patient care.
Bmj 2012;345:e3999–e3999.

166. Hand, DJ.
Evaluating diagnostic tests: The area under the ROC curve and the balance of errors.
Statistics in medicine 2010;29:1502–10.

167. Vickers, AJ and Elkin, EB.
Decision curve analysis: a novel method for evaluating prediction models.
Medical decision making : an international journal of the Society for Medical Decision Making 2006;26:565–74.

168. Vickers, AJ, Cronin, AM, Elkin, EB, and Gonen, M.
Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers.
BMC Medical Informatics and Decision Making 2008;8:53.

169. Moons, KG, Stijnen, T, Michel, BC, et al.
Application of treatment thresholds to diagnostic-test evaluation: an alternative to the comparison of areas under receiver operating characteristic curves.
Medical decision making : an international journal of the Society for Medical Decision Making 1997;17:447–54.

170. Chan, QL, Xu, X, Shaik, MA, et al.

Clinical utility of the informant AD8 as a dementia case finding instrument in primary healthcare.
Journal of Alzheimer's disease : JAD 2016;49:121–7.

171. Tokuhara, KG, Valcour, VG, Masaki, KH, and Blanchette, PL.
Utility of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for dementia in a Japanese-American population.
Hawaii medical journal 2006;65:72–5.

172. Moons, KG, Es, GA van, Deckers, JW, Habbema, JD, and Grobbee, DE.
Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example.
Epidemiology (Cambridge, Mass.) 1997;8:12–7.

173. Chan, SF, Deeks, JJ, Macaskill, P, and Irwig, L.
Three methods to construct predictive models using logistic regression and likelihood ratios to facilitate adjustment for pretest probability give similar results.
Journal of Clinical Epidemiology 2008;61:52–63.

174. Helman, CG.
Disease versus illness in general practice.
The Journal of the Royal College of General Practitioners 1981;31:548–52.

175. Croft, P, Altman, DG, Deeks, JJ, 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 2015;13:20.

176. Van Straaten, EC, Scheltens, P, Knol, DL, et al.
Operational definitions for the NINDS-AIREN criteria for vascular dementia an interobserver study.
Stroke 2003;34:1907–1912.

177. Vet, HC de, Mokkink, LB, Terwee, CB, Hoekstra, OS, and Knol, DL.
Clinicians are right not to like Cohen's κ.
BMJ (Clinical research ed.) 2013;346:1–7.

178. Mandrekar, JN.
Measures of interrater agreement.
Journal of Thoracic Oncology 2011;6:6–7.

179. Simel, DL, Rennie, D, and Bossuyt, PMM.
The STARD statement for reporting diagnostic accuracy studies: application to the history and physical examination.
Journal of general internal medicine 2008;23:768–74.

180. Brush, JE, Sherbino, J, and Norman, GR.
How Expert Clinicians Intuitively Recognize a Medical Diagnosis.
The American Journal of Medicine 2017;130:629–634.

181. Kahneman, D.
Thinking, fast and slow.
First edit.
New York, NY: Farrar, Straus and Giroux, 2011:
499.

182. Lehman, R.
Siddharta Mukherjee's three laws of medicine.
BMJ 2015;6708:h6708.

183. Lambe, KA, O'Reilly, G, Kelly, BD, and Curristan, S.
Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review.
BMJ quality & safety 2016;25:808–20.

184. Pauker, SG and Kassirer, JP.
Therapeutic Decision Making: A Cost-Benefit Analysis.
New England Journal of Medicine 1975;293:229–234.

185. Charlin, B, Tardif, J, and Boshuizen, HP.
Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research.
Academic medicine : journal of the Association of American Medical Colleges 2000;75:182–190.

186. Greenhalgh, T.
Narrative based medicine: narrative based medicine in an evidence based world.
BMJ (Clinical research ed.) 1999;318:323–5.

187. Foot, C, Naylor, C, and Imison, C.
The quality of GP diagnosis and referral.
Tech. rep.
London, UK: The King's Fund, 2010:
79.
URL: https://www.kingsfund.org.uk/sites/default/files/Diagnosis%20and%20referral.pdf.

188. NHS England.
Numbers of Patients Registered at a GP Practice - April 2018.
Tech. rep.
NHS England, 2018.

URL: https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice/patients-registered-at-a-gp-practice-april-2018-special-topic---registered-patients-compared-to-the-projected-resident-population-in-england.

189. Stolper, E, Bokhoven, M van, Houben, P, et al.
The diagnostic role of gut feelings in general practice A focus group study of the concept and its determinants.
BMC Family Practice 2009;10:17.

190. Woolley, A and Kostopoulou, O.
Clinical intuition in family medicine: More than first impressions.
Annals of Family Medicine 2013;11:60–66.

191. Borson, S, Scanlan, JM, Watanabe, J, Tu, SP, and Lessig, M.
Improving identification of cognitive impairment in primary care.
International journal of geriatric psychiatry 2006;21:349–55.

192. Löppönen, M, Räihä, I, Isoaho, R, et al.
Diagnosing cognitive impairment and dementia in primary health care - a more active approach is needed.
Age and Ageing 2003;32:606–612.

193. Buntinx, F, Mant, D, Van den Bruel, A, Donner-Banzhof, N, and Dinant, GJ.
Dealing with low-incidence serious diseases in general practice.
The British journal of general practice : the journal of the Royal College of General Practitioners 2011;61:43–6.

194. Barraclough, K.
Medical intuition.
BMJ 2006;332:497.2.

195. Hjertholm, P, Moth, G, Ingeman, ML, and Vedsted, P.
Predictive values of GPs' suspicion of serious disease: a population-based follow-up study.
The British journal of general practice : the journal of the Royal College of General Practitioners 2014;64:346–53.

196. Hamilton, W.
Five misconceptions in cancer diagnosis.
The British journal of general practice : the journal of the Royal College of General Practitioners 2009;59:441–5.

197. 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 of Systematic Reviews 2017;2:CD012558.

198. Moher, D, Shamseer, L, Clarke, M, et al.
Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement.
Systematic reviews 2015;4:1.

199. Whiting, PF, Rutjes, AWS, Westwood, ME, et al.
QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.
Annals of internal medicine 2011;155:529–36.

200. O'Connor, DW, Fertig, A, Grande, MJ, et al.
Dementia in general practice: the practical consequences of a more positive approach to diagnosis.
The British journal of general practice : the journal of the Royal College of General Practitioners 1993;43:185–8.

201. Pentzek, M, Fuchs, A, Wiese, B, et al.
General practitioners' judgment of their elderly patients' cognitive status.
Journal of general internal medicine 2009;24:1314–7.

202. Blaeuer, SR, Bally, K, Tschudi, P, Martina, B, and Zeller, A.
Acute cough illness in general practice - predictive value of clinical judgement and accuracy of requesting chest x-rays.
Praxis 2013;102:1287–92.

203. Di Somma, S, Magrini, L, De Berardinis, B, et al.
Additive value of blood neutrophil gelatinase-associated lipocalin to clinical judgement in acute kidney injury diagnosis and mortality prediction in patients hospitalized from the emergency department.
Critical care (London, England) 2013;17:R29.

204. Body, R, Cook, G, Burrows, G, Carley, S, and Lewis, PS.
Can emergency physicians 'rule in' and 'rule out' acute myocardial infarction with clinical judgement?
Emergency Medicine Journal 2014;31:872–876.

205. Chen, L, Reed, C, Happich, M, Nyhuis, A, and Lenox-Smith, A.
Health care resource utilisation in primary care prior to and after a diagnosis of Alzheimer's disease: a retrospective, matched case-control study in the United Kingdom.
BMC geriatrics 2014;14:76.

206. Ramakers, IHGB, Visser, PJ, Aalten, P, et al.
Symptoms of preclinical dementia in general practice up to five years before dementia diagnosis.

Dementia and geriatric cognitive disorders 2007;24:300–6.

207. Ydstebø, AE, Bergh, S, Selbæk, G, Benth, JŠ, Lurås, H, and Vossius, C.
The impact of dementia on the use of general practitioners among the elderly in Norway.
Scandinavian journal of primary health care 2015;33:199–205.

208. Dewey, ME and Copeland, JR.
Computerized psychiatric diagnosis in the elderly: AGECAT.
Journal of Microcomputer Applications 1986;9:135–140.

209. Roth, M, Tym, E, Mountjoy, CQ, et al.
CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia.
The British journal of psychiatry : the journal of mental science 1986;149:698–709.

210. Zaudig, M, Mittelhammer, J, Hiller, W, et al.
SIDAM–A structured interview for the diagnosis of dementia of the Alzheimer type, multi-infarct dementia and dementias of other aetiology according to ICD-10 and DSM-III-R.
Psychological medicine 1991;21:225–36.

211. Petersen, RC, Smith, GE, Waring, SC, Ivnik, RJ, Tangalos, EG, and Kokmen, E.
Mild cognitive impairment: clinical characterization and outcome.
Archives of neurology 1999;56:303–8.

212. Whiting, P, Westwood, M, Beynon, R, Burke, M, Sterne, JA, and Glanville, J.
Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies.
Journal of clinical epidemiology 2011;64:602–7.

213. Veritas Health Innovation.
Covidence.
Melbourne, 2019.

214. The Cochrane Collaboration.
Review Manager (RevMan) [Computer program].
Copenhagen: The Nordic Cochrane Centre, 2014.

215. StataCorp.
Stata Statistical Software: Release 13.
College Station, Tx, 2013.

216. Freeman, SC, Kerby, CR, Patel, A, Cooper, NJ, Quinn, T, and Sutton, AJ.
Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA.
BMC Medical Research Methodology 2019;19:81.

217. Reitsma, JB, Glas, AS, Rutjes, AWS, Scholten, RJPM, Bossuyt, PM, and Zwinderman, AH.

Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews.
Journal of clinical epidemiology 2005;58:982–90.

218. Chu, H and Cole, SR.
Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach.
Journal of clinical epidemiology 2006;59:1331–2.

219. Harbord, RM, Deeks, JJ, Egger, M, Whiting, P, and Sterne, J.
A unification of models for meta-analysis of diagnostic accuracy studies.
Biostatistics 2007;8:239–251.

220. Bossuyt, P, Davenport, C, Deeks, J, Hyde, C, Leeflang, M, and Scholten, R.
Interpreting results and drawing conclusions.
In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*.
Ed. by Deeks, J, Bossuyt, P, and Gatsonis, C.
version0.9.
The Cochrane Collaboration, 2013.
Chap. 11.
URL: `http://srdta.cochrane.org/`.

221. Valcour, VG, Masaki, KH, Curb, JD, and Blanchette, PL.
The detection of dementia in the primary care setting.
Archives of internal medicine 2000;160:2964–8.

222. Chong, SA, Abdin, E, Vaingankar, J, Ng, LL, and Subramaniam, M.
Diagnosis of dementia by medical practitioners: a national study among older adults in Singapore.
Aging & Mental Health 2016;20:1271–1276.

223. Engedal, K, Gilje, K, and Lilleaas, F.
Diagnostic Evaluation of the Mentally Impaired Elderly Living at Home.
Scandinavian Journal of Primary Health Care 1989;7:5–11.

224. Pittman, J, Andrews, H, Tatemichi, T, et al.
Diagnosis of dementia in a heterogeneous population. A comparison of paradigm-based diagnosis and physician's diagnosis.
Archives of neurology 1992;49:461–7.

225. Belmin, J, Min, L, Roth, C, Reuben, D, and Wenger, N.
Assessment and management of patients with cognitive impairment and dementia in primary care.
The journal of nutrition, health & aging 2012;16:462–467.

226. Jacinto, AF, Nitrini, R, Brucki, SMD, and Porto, CS.
Detection of cognitive impairment in the elderly by general practitioners in Brazil.
In: *Alzheimer's & Dementia*.
Vol. 5.
4.
Vienna Austria: Alzheimer's Association International Conference on Alzheimer's Disease, 2009:
189–190.
DOI: 10.1016/j.jalz.2009.04.066.

227. Wilkins, CH, Wilkins, KL, Meisel, M, Depke, M, Williams, J, and Edwards, DF.
Dementia Undiagnosed in Poor Older Adults with Functional Impairment.
Journal of the American Geriatrics Society 2007;55:1771–1776.

228. Dilts, SL, Mann, N, and Dilts, JG.
Accuracy of referring psychiatric diagnosis on a consultation-liaison service.
Psychosomatics;44:407–11.

229. Jansen, APD, Hout, HPJ van, Nijpels, G, Marwijk, HWJ van, Vet, HCW de, and Stalman, WAB.
Yield of a new method to detect cognitive impairment in general practice.
International Journal of Geriatric Psychiatry 2007;22:590–597.

230. Tierney, MC, Naglie, G, Upshur, R, et al.
Factors associated with primary care physicians' recognition of cognitive impairment in their older patients.
Alzheimer disease and associated disorders;28:320–5.

231. Waldorff, FB, Rishøj, S, and Waldemar, G.
Identification and diagnostic evaluation of possible dementia in general practice.
Scandinavian Journal of Primary Health Care 2005;23:221–226.

232. Mant, A, Eyland, EA, Pond, DC, Saunders, NA, and Chancellor, AH.
Recognition of dementia in general practice: comparison of general practitioners' opinions with assessments using the mini-mental state examination and the Blessed dementia rating scale.
Family practice 1988;5:184–8.

233. Blessed, G, Tomlinson, BE, and Roth, M.
The Association Between Quantitative Measures of Dementia and of Senile Change in the Cerebral Grey Matter of Elderly Subjects.
British Journal of Psychiatry 1968;114:797–811.

234. Hara, J, Macias, D, Russell, J, et al.
Prevalence of cognitive impairment based on the annual wellness visit.
Alzheimer's & Dementia 2013;9:P119.

235. Bushnell, J and MaGPIe Research Group.
Frequency of consultations and general practitioner recognition of psychological symptoms.
The British journal of general practice : the journal of the Royal College of General Practitioners 2004;54:838–43.

236. Hopman-Rock, M, Tak, ECPM, and Staats, PGM.
Development and validation of the Observation List for early signs of Dementia (OLD).
International Journal of Geriatric Psychiatry 2001;16:406–414.

237. Leung, WC.
GPs' diagnosis of dementia.
The British journal of general practice : the journal of the Royal College of General Practitioners 2000;50:666.

238. Schaub, RT, Linden, M, and Copeland, JRM.
A comparison of GMS-A/AGECAT, DSM-III-R for dementia and depression, including subthreshold depression (SD)–results from the Berlin Aging Study (BASE).
International journal of geriatric psychiatry 2003;18:109–17.

239. Livingston, G, Sax, K, Willison, J, Blizard, B, and Mann, A.
The Gospel Oak Study stage II: the diagnosis of dementia in the community.
Psychological medicine 1990;20:881–91.

240. Aldus, C, Arthur, A, Fox, C, et al.
Cognitive function and ageing study II dementia diagnosis study (CADDY): the prevalence, causes and consequences of dementia undetected or undiagnosed in primary care in England.
Alzheimer's & Dementia 2018;14:P573–P574.

241. Camicioli, R, Willert, P, Lear, J, Grossmann, S, Kaye, J, and Butterfield, P.
Dementia in Rural Primary Care Practices in Lake County, Oregon.
Journal of Geriatric Psychiatry and Neurology 2000;13:87–92.

242. Mok, W, Chow, TW, Zheng, L, Mack, WJ, and Miller, C.
Clinicopathological concordance of dementia diagnoses by community versus tertiary care clinicians.
American Journal of Alzheimer's Disease & Other Dementiasr 2004;19:161–165.

243. Ólafsdóttir, M, Skoog, I, Marcusson, J, Olafsdóttir, M, Skoog, I, and Marcusson, J.
Detection of Dementia in Primary Care: The Linköping Study.
Dementia and Geriatric Cognitive Disorders 2000;11:223–229.

244. Hout, HPJ van, Vernooij-Dassen, MJFM, Hoefnagels, WHL, et al.
Dementia: predictors of diagnostic accuracy and the contribution of diagnostic recommendations.
The Journal of family practice 2002;51:693–9.

245. Hout, HP van, Vernooij-Dassen, MJ, Jansen, DA, and Stalman, WA.
Do general practitioners disclose correct information to their patients suspected of dementia and their caregivers? A prospective observational study.
Aging & Mental Health 2006;10:151–155.

246. Hout, H van, Vernooij-Dassen, M, Hoefnagels, W, and Grol, R.
Use of mini-mental state examination by GPs to diagnose dementia may be unnecessary.
BMJ (Clinical research ed.) 1999;319:190.

247. Hout, H van, Vernooij-Dassen, M, Poels, P, Hoefnagels, WHL, and Grol, RPTM.
Are general practitioners able to accurately diagnose dementia and identify Alzheimer's disease? A comparison with an outpatient memory clinic.
The British journal of general practice : the journal of the Royal College of General Practitioners 2000;50:311–2.

248. Van Hout, H.
Applicability of diagnostic recommendations on dementia in family practice.
International Journal for Quality in Health Care 2001;13:127–133.

249. Hout, HP van, Vernooij-Dassen, MJ, and Stalman, WA.
Diagnosing dementia with confidence by GPs.
Family Practice 2007;24:616–621.

250. Hout, H van, Vernooij-Dassen, M, Jansen, D, and Stalman, W.
Geven huisartsen aan dementerende patiënten en hun verzorgers de juiste informatie?
Huisarts en Wetenschap 2007;50:424–430.

251. Hout, H van, Vernooij-Dassen, M, Hoefnagels, W, et al.
De diagnostische waarde van de aanbevelingen uit de NHG-Standaard Dementie.
Huisarts en Wetenschap 2003;46:71–78.

252. Dinesen, O, Frijs-Madsen, B, Almbjerg, F, Fromholt, P, and Torpdahl, P.
[Dementia diagnosis in general practice].
Ugeskrift for laeger 1997;159:5795–9.

253. Lionis, C, Tzagournissakis, M, Iatraki, E, Kozyraki, M, Antonakis, N, and Plaitakis, A.
Are primary care physicians able to assess dementia? An estimation of their capacity after a short-term training program in rural Crete.
The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry 2001;9:315.

254. Wang, SZ, Tan, GC, Wang, XF, De Roza, JG, Lim, LL, and Kandiah, N.
Experience with a Community-based Multidisciplinary Memory Clinic: A Primary Care Perspective.
Annals of the Academy of Medicine, Singapore 2017;46:321–323.

255. De Lepeleire, J, Heyrman, J, Baro, F, and Buntinx, F.
A combination of tests for the diagnosis of dementia had a significant diagnostic value.
Journal of Clinical Epidemiology 2005;58:217–225.

256. Heyrman, J, Dessers, L, Munter, MB de, Haepers, K, and Craenen, J.
Functional Status Assessment in the Elderly.
In: *Functional Status Measurement in Primary Care*.
Ed. by WONCA Classification Committee.
First.
New York, NY: Springer, 1990.
Chap. Functional:213–221.
DOI: doi.org/10.1007/978-1-4613-8977-4{_}14.

257. Hessler, J, Brönner, M, Etgen, T, et al.
Suitability of the 6CIT as a screening test for dementia in primary care patients.
Aging & Mental Health 2014;18:515–520.

258. Brooke, P and Bullock, R.
Validation of a 6 item cognitive impairment test with a view to primary care usage.
International Journal of Geriatric Psychiatry 1999;14:936–940.

259. Juva, K, Sulkava, R, Erkinjuntti, T, Ylikoski, R, Valvanne, J, and Tilvis, R.
Staging the severity of dementia: comparison of clinical (CDR, DSM-III-R), functional (ADL, IADL) and cognitive (MMSE) scales.
Acta neurologica Scandinavica 1994;90:293–8.

260. Kurz, X, Broers, M, Scuvée-Moreau, J, et al.
Methodological issues in a cost-of-dementia study in Belgium: the NAtional Dementia Economic Study (NADES).
Acta neurologica Belgica 1999;99:167–75.

261. Noda, H, Yamagishi, K, Ikeda, A, Asada, T, and Iso, H.
Identification of dementia using standard clinical assessments by primary care physicians in Japan.
Geriatrics & Gerontology International 2018;18:738–744.

262. Mitchell, AJ, Meader, N, and Pentzek, M.
Clinical recognition of dementia and cognitive impairment in primary care: a meta-analysis of physician accuracy.
Acta psychiatrica Scandinavica 2011;124:165–83.

263. Dungen, P van den, Marwijk, HWM van, Horst, HE van der, et al.
The accuracy of family physicians' dementia diagnoses at different stages of dementia: a systematic review.
International journal of geriatric psychiatry 2012;27:342–54.

264. Cooper, B, Bickel, H, and Schäufele, M.
The ability of general practitioners to detect dementia and cognitive impairment in their elderly patients: A study in Mannheim.
International Journal of Geriatric Psychiatry 1992;7:591–598.

265. Eefsting, Ja, Boersma, F, Van den Brink, W, and Van Tilburg, W.
Differences in prevalence of dementia based on community survey and general practitioner recognition.
Psychological medicine 1996;26:1223–30.

266. Pond, CD, Mant, A, Kehoe, L, Hewitt, H, and Brodaty, H.
General practitioner diagnosis of depression and dementia in the elderly: can academic detailing make a difference?
Family practice 1994;11:141–7.

267. Brayne, C and Calloway, P.
The case identification of dementia in the community: A comparison of methods.
International Journal of Geriatric Psychiatry 1990;5:309–316.

268. De Lepeleire, J, Aertgeerts, B, Umbach, I, et al.
The diagnostic value of IADL evaluation in the detection of dementia in general practice.
Aging & mental health 2004;8:52–7.

269. Kaduszkiewicz, H, Zimmermann, T, Van den Bussche, H, et al.
Do general practitioners recognize mild cognitive impairment in their patients?
The journal of nutrition, health & aging 2010;14:697–702.

270. O'Connor, DW, Pollitt, PA, Hyde, JB, Brook, CP, Reiss, BB, and Roth, M.
Do general practitioners miss dementia in elderly patients?
BMJ (Clinical research ed.) 1988;297:1107–10.

271. Pentzek, M, Wollny, A, Wiese, B, 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 2009;17:965–75.

272. 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;7:433–450.

273. Wind, AW, Van Staveren, G, Schellevis, FG, Jonker, C, and Eijk, JTM van.
The validity of the judgement of general practitioners on dementia.
International Journal of Geriatric Psychiatry 1994;9:543–549.

274. Boise, L, Neal, MB, Kaye, J, et al.
Dementia Assessment in Primary Care: Results From a Study in Three Managed Care Systems.
The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 2004;59:M621–M626.

275. Boustani, M, Callahan, CM, Unverzagt, FW, et al.
Implementing a screening and diagnosis program for dementia in primary care.
Journal of general internal medicine 2005;20:572–7.

276. Bowers, J, Jorm, AF, Henderson, S, and Harris, P.
General practitioners' detection of depression and dementia in elderly patients.
The Medical journal of Australia 1990;153:192–6.

277. Callahan, CM, Hendrie, HC, and Tierney, WM.
Documentation and evaluation of cognitive impairment in elderly primary care patients.
Annals of internal medicine 1995;122:422–9.

278. Chodosh, J, Petitti, DB, Elliott, M, et al.
Physician recognition of cognitive impairment: evaluating the need for improvement.
Journal of the American Geriatrics Society 2004;52:1051–9.

279. Ganguli, M, Rodriguez, E, Mulsant, B, et al.
Detection and management of cognitive impairment in primary care: The steel valley seniors survey.
Journal of the American Geriatrics Society 2004;52:1668–1675.

280. Iliffe, S, Booroff, A, Gallivan, S, Goldenberg, E, Morgan, P, and Haines, A.
Screening for cognitive impairment in the elderly using the mini-mental state examination.
The British journal of general practice : the journal of the Royal College of General Practitioners 1990;40:277–9.

281. Cooper, B, Bickel, H, and Schäufele, M.
[Dementia diseases and minor cognitive impairments in elderly patients in general practice. Results of a cross-sectional study].
Der Nervenarzt 1992;63:551–60.

282. Morris, JC, Heyman, A, Mohs, RC, et al.
The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease.
Neurology 1989;39:1159–65.

283. Pfeiffer, E.
A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients.
Journal of the American Geriatrics Society 1975;23:433–41.

284. Kawas, C, Segal, J, Stewart, WF, Corrada, M, and Thal, LJ.
A validation study of the Dementia Questionnaire.
Archives of neurology 1994;51:901–6.

285. Barberger-Gateau, P, Fabrigoule, C, Helmer, C, Rouch, I, and Dartigues, JF.
Functional impairment in instrumental activities of daily living: an early clinical sign of dementia?
Journal of the American Geriatrics Society 1999;47:456–62.

286. Takwoingi, Y, Riley, RD, and Deeks, JJ.
Meta-analysis of diagnostic accuracy studies in mental health.
Evidence Based Mental Health 2015;18:103–109.

287. Matthews, FE, Arthur, A, Barnes, LE, et al.
A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II.
The Lancet 2013;382:1405–1412.

288. Sibbett, RA, Russ, TC, Deary, IJ, and Starr, JM.
Dementia ascertainment using existing data in UK longitudinal and cohort studies: a systematic review of methodology.
BMC Psychiatry 2017;17:239.

289. Ensor, J, Deeks, JJ, Martin, EC, and Riley, RD.
Meta-analysis of test accuracy studies using imputation for partial reporting of multiple thresholds.
Research synthesis methods 2018;9:100–115.

290. Larson, EB, McCurry, SM, Graves, AB, et al.
Standardization of the clinical diagnosis of the dementia syndrome and its subtypes in a cross-national study: the Ni-Hon-Sea experience.
The journals of gerontology. Series A, Biological sciences and medical sciences 1998;53:313–9.

291. Graham, JE, Rockwood, K, Beattie, BL, McDowell, I, Eastwood, R, and Gauthier, S.
Standardization of the diagnosis of dementia in the Canadian Study of Health and Aging.
Neuroepidemiology 1996;15:246–56.

292. Farrer, LA, Cupples, LA, Blackburn, S, et al.
Interrater agreement for diagnosis of Alzheimer's disease: the MIRAGE study.
Neurology 1994;44:652–6.

293. Borson, S, Scanlan, J, Brush, M, Vitaliano, P, and Dokmak, A.
The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly.
International journal of geriatric psychiatry 2000;15:1021–7.

294. Carnero-Pardo, C, Cruz-Orduña, I, Espejo-Martínez, B, Martos-Aparicio, C, López-Alcalde, S, and Olazarán, J.
Utility of the mini-cog for detection of cognitive impairment in primary care: data from two spanish studies.
International journal of Alzheimer's disease 2013;2013:285462.

295. Fuchs, A, Wiese, B, Altiner, A, Wollny, A, and Pentzek, M.
Cued recall and other cognitive tasks to facilitate dementia recognition in primary care.
Journal of the American Geriatrics Society 2012;60:130–5.

296. Holsinger, T, Plassman, BL, Stechuchak, KM, Burke, JR, Coffman, CJ, and Williams, JW.
Screening for cognitive impairment: comparing the performance of four instruments in primary care.
Journal of the American Geriatrics Society 2012;60:1027–36.

297. McCarten, JR, Anderson, P, Kuskowski, MA, McPherson, SE, Borson, S, and Dysken, MW.
Finding dementia in primary care: the results of a clinical demonstration project.
Journal of the American Geriatrics Society 2012;60:210–7.

298. Jorm, AF and Jacomb, PA.
The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms.
Psychological medicine 1989;19:1015–22.

299. Galvin, JE, Roe, CM, Powlishta, KK, et al.
The AD8: a brief informant interview to detect dementia.
Neurology 2005;65:559–64.

300. Lourenço, RA and Veras, RP.
[Mini-Mental State Examination: psychometric characteristics in elderly outpatients].
Revista de saude publica 2006;40:712–9.

301. Brodaty, H, Pond, D, Kemp, NM, et al.
The GPCOG: a new screening test for dementia designed for general practice.
Journal of the American Geriatrics Society 2002;50:530–4.

302. Cruz-Orduña, I, Bellón, JM, Torrero, P, et al.
Detecting MCI and dementia in primary care: effectiveness of the MMS, the FAQ and the IQCODE [corrected].
Family practice 2012;29:401–6.

303. Lavery, LL, Lu, Sy, Chang, CCH, Saxton, J, and Ganguli, M.
Cognitive assessment of older primary care patients with and without memory complaints.
Journal of general internal medicine 2007;22:949–54.

304. Jorm, AF.

A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation.
Psychological medicine 1994;24:145–53.

305. Jorm, AF.
The Informant Questionnaire on cognitive decline in the elderly (IQCODE): a review.
International psychogeriatrics 2004;16:275–93.

306. Burns, A.
Alistair Burns and 51 colleagues reply to David Le Couteur and colleagues.
BMJ (Clinical research ed.) 2013;347:f6125.

307. Le Couteur, DG, Doust, J, Creasey, H, and Brayne, C.
Political drive to screen for pre-dementia: not evidence based and ignores the harms of diagnosis.
BMJ (Clinical research ed.) 2013;347:f5125.

308. Hunt, H.
Improving the accuracy of brief cognitive assesements when used as part of the process for identifying dementia in general practice.
PhD thesis. University of Exeter, 2018:
345.

309. Iliffe, S, Robinson, L, Brayne, C, et al.
Primary care and dementia: 1. diagnosis, screening and disclosure.
International journal of geriatric psychiatry 2009;24:895–901.

310. NHS England.
Numbers of Patients Registered at a GP Practice - July 2016.
Tech. rep.
2016.
URL: https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice/july-2016.

311. Addicott, R and Ham, C.
Commissioning and funding general practice Making the case for family care networks.
2014:1–49.

312. Digital, N.
NHS Payments to General Practice.
2019.
URL: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-payments-to-general-practice.

313. Burns, A.

Letter from Alistair Burns to CCG Clinical Leads and CCG Accountable Officers regarding dementia diagnosis rates.
Leeds.
URL: https://www.england.nhs.uk/wp-content/uploads/2017/11/ambulance-handover-letter.pdf.

314. NHS England.
Facilitating timely diagnosis and support for people with dementia.
London, 2013.
DOI: https://www.england.nhs.uk/wp-content/uploads/2013/03/ess-dementia.pdf.

315. Bristol Clinical Commissioning Group.
Recognition and Management of People with Dementia and their Family Carers in General Practices in Bristol.
Bristol, 2013.
URL: http://www.mentalhealth.bristolccg.nhs.uk/media/1279/Dementia-LES.pdf.

316. EMIS Health.
EMIS Web.
Leeds.
URL: https://www.emishealth.com/products/emis-web/.

317. Ballard, C, Alistar, B, Corbett, A, Livingston, G, and Rasmussen, J.
Helping you to assess cognition: A practical toolkit for clinicians.
London: Alzheimer's Society, 2015:
1–42.
URL: http://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2532.

318. Rami, L, Molinuevo, JL, Sanchez-Valle, R, Bosch, B, and Villar, A.
Screening for amnestic mild cognitive impairment and early Alzheimer's disease with M@T (Memory Alteration Test) in the primary care population.
International journal of geriatric psychiatry 2007;22:294–304.

319. Carnero-Pardo, C, Gurpegui, M, Sanchez-Cantalejo, E, et al.
Diagnostic accuracy of the Eurotest for dementia: a naturalistic, multicenter phase II study.
BMC neurology 2006;6:15.

320. Carnero-Pardo, C, Espejo-Martinez, B, Lopez-Alcalde, S, et al.
Effectiveness and costs of phototest in dementia and cognitive impairment screening.
BMC neurology 2011;11:92.

321. Takechi, H and Dodge, HH.

Scenery Picture Memory Test: a new type of quick and effective screening test to detect early stage Alzheimer's disease patients.
Geriatrics & gerontology international 2010;10:183–90.

322. Inouye, SK, Robison, JT, Froehlich, TE, and Richardson, ED.
The time and change test: a simple screening test for dementia.
The journals of gerontology. Series A, Biological sciences and medical sciences 1998;53:281–6.

323. Podsiadlo, D and Richardson, S.
The timed "Up & Go": a test of basic functional mobility for frail elderly persons.
Journal of the American Geriatrics Society 1991;39:142–8.

324. Richards, M, Marder, K, Bell, K, Dooneief, G, Mayeux, R, and Stern, Y.
Interrater reliability of extrapyramidal signs in a group assessed for dementia.
Archives of neurology 1991;48:1147–9.

325. Hummel, T, Sekinger, B, Wolf, S, Pauli, E, and Kobal, G.
'Sniffin' Sticks': Olfactory Performance Assessed by the Combined Testing of Odour Identification, Odor Discrimination and Olfactory Threshold.
Chemical Senses 1997;22:39–52.

326. Hummel, T, Kobal, G, Gudziol, H, and Mackay-Sim, A.
Normative data for the "Sniffin' Sticks" including tests of odor identification, odor discrimination, and olfactory thresholds: an upgrade based on a group of more than 3,000 subjects.
European archives of oto-rhino-laryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery 2007;264:237–43.

327. Pfeffer, RI, Kurosaki, TT, Harrah, CH, et al.
Measurement of Functional Activities in Older Adults in the Community 1.
Journal of Gerontology 1982;37:323–329.

328. Lawton, MP and Brody, EM.
Assessment of older people: self-maintaining and instrumental activities of daily living.
The Gerontologist 1969;9:179–86.

329. Katz, S, Ford, AB, Moskowitz, RW, Jackson, BA, and Jaffe, MW.
Studies of Illness in the Aged. The Index of ADL: A Standardized Measure of Biological and Psychosocial Function.
JAMA 1963;185:914–9.

330. Tsoi, KKF, Chan, JYC, Hirai, HW, Wong, SYS, and Kwok, TCY.
Cognitive Tests to Detect Dementia.
JAMA Internal Medicine 2015.

331. Buschke, H, Kuslansky, G, Katz, M, et al.

Screening for dementia with the memory impairment screen.
Neurology 1999;52:231–238.

332. Hodkinson, HM.
Evaluation of a mental test score for assessment of mental impairment in the elderly.
Age and ageing 1972;1:233–8.

333. Nasreddine, ZS, Phillips, NA, Bédirian, V, et al.
The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment.
Journal of the American Geriatrics Society 2005;53:695–9.

334. Davis, DHJ, Creavin, ST, Yip, JLY, Noel-Storr, AH, Brayne, C, and Cullum, S.
Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias.
The Cochrane database of systematic reviews 2015;10:CD010775.

335. Burns, A, Twomey, P, Barrett, E, et al.
Dementia diagnosis and management A brief pragmatic resource for general practitioners.
Tech. rep.
NHS England, 2015:
24.
URL: http://www.england.nhs.uk/wp-content/uploads/2015/01/dementia-diag-mng-ab-pt.pdf.

336. Hsieh, S, Schubert, S, Hoon, C, Mioshi, E, and Hodges, JR.
Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease.
Dementia and geriatric cognitive disorders 2013;36:242–50.

337. Adshead, F, Cody, DD, and Pitt, B.
BASDEC: a novel screening instrument for depression in elderly medical inpatients.
BMJ (Clinical research ed.) 1992;305:397.

338. Bucks, RS, Ashworth, DL, Wilcock, GK, and Siegfried, K.
Assessment of activities of daily living in dementia: development of the Bristol Activities of Daily Living Scale.
Age and ageing 1996;25:113–20.

339. Hsieh, S, McGrory, S, Leslie, F, et al.
The Mini-Addenbrooke's Cognitive Examination: A New Assessment Tool for Dementia.
Dementia and Geriatric Cognitive Disorders 2015;39:1–11.

340. Dodd, E, Cheston, R, Cullum, S, et al.
Primary care-led dementia diagnosis services in South Gloucestershire: Themes from people and families living with dementia and health care professionals.

Dementia 2016;15:1586–1604.

341. Google LLC.
Google Earth Pro 7.3.2.5776 (64-bit).
2019.

342. Harris, PA, Taylor, R, Thielke, R, Payne, J, Gonzalez, N, and Conde, JG.
Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support.
Journal of Biomedical Informatics 2009;42:377–381.

343. Harris, PA, Taylor, R, Minor, BL, et al.
The REDCap consortium: Building an international community of software platform partners.
Journal of biomedical informatics 2019;95:103208.

344. RStudio Team.
RStudio: Integrated Development Environment for R.
Boston, MA, 2016.

345. Richards, M, Stern, Y, and Mayeux, R.
Subtle extrapyramidal signs can predict the development of dementia in elderly individuals.
Neurology 1993;43:2184–2184.

346. Pfeffer, RI, Kurosaki, TT, Harrah, CH, Chance, JM, and Filos, S.
Measurement of Functional Activities in Older Adults in the Community.
Journal of Gerontology 1982;37:323–329.

347. Juva, K, Mäkelä, M, Erkinjuntti, T, et al.
Functional assessment scales in detecting dementia.
Age and ageing 1997;26:393–400.

348. Hummel, T, Rosenheim, K, Konnerth, CG, and Kobal, G.
Screening of Olfactory Function with a Four-Minute Odor Identification Test: Reliability, Normative Data, and Investigations in Patients with Olfactory Loss.
Annals of Otology, Rhinology & Laryngology 2001;110:976–981.

349. Galvin, JE, Roe, CM, Coats, Ma, and Morris, JC.
Patient's rating of cognitive ability: using the AD8, a brief informant interview, as a self-rating tool to detect dementia.
Archives of neurology 2007;64:725–730.

350. Quinn, TJ, Fearon, P, Noel-Storr, AH, Young, C, McShane, R, and Stott, DJ.
Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations.
The Cochrane database of systematic reviews 2014;4:CD010079.

351. Vickers, AJ, Calster, B van, and Steyerberg, EW.

A simple, step-by-step guide to interpreting decision curve analysis.
Diagnostic and Prognostic Research 2019;3:1–8.

352. Vickers, AJ, Van Calster, B, and Steyerberg, EW.
Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests.
Bmj 2016:i6.

353. Alba, AC, Agoritsas, T, Walsh, M, et al.
Discrimination and calibration of clinical prediction models: Users' guides to the medical literature.
JAMA - Journal of the American Medical Association 2017;318:1377–1384.

354. Flahault, A, Cadilhac, M, and Thomas, G.
Sample size calculation should be performed for design accuracy in diagnostic test studies.
Journal of clinical epidemiology 2005;58:859–62.

355. Vittinghoff, E and McCulloch, CE.
Relaxing the rule of ten events per variable in logistic and Cox regression.
American journal of epidemiology 2007;165:710–8.

356. Creavin, ST, Cullum, SJ, Haworth, J, et al.
Towards improving diagnosis of memory loss in general practice: TIMeLi diagnostic test accuracy study protocol.
BMC Family Practice 2016;17:79.

357. Noel-Storr, AH, McCleery, JM, Richard, E, et al.
Reporting standards for studies of diagnostic test accuracy in dementia: The STARDdem Initiative.
Neurology 2014.

358. McGee, S.
Simplifying likelihood ratios.
Journal of general internal medicine 2002;17:646–9.

359. Oxford English Dictionary.
kin, n.1.
In: *OED Online*.
June 2019.
Web: Oxford University Press.
URL: https://www.oed.com/view/Entry/103433?rskey=gtLLAR&amp%20result=1&amp%20isAdvanced=false.

360. Srivastava, A, Thomson, SB, Barnett-Page, E, et al.
Framework Analysis : A qualitative methodology for applied policy research.

BMC medical research methodology 2009;4:72–79.

361. Ritchie, J and Lewis, J.
Qualitative research practice: a Guide for Social Science Students and Researchers.
First edit.
London,UK: Sage, 2003:
336.

362. Gale, NK, Heath, G, Cameron, E, Rashid, S, and Redwood, S.
Using the framework method for the analysis of qualitative data in multi-disciplinary health research.
BMC Medical Research Methodology 2013;13:117.

363. Thorne, S.
Data analysis in qualitative research.
Evidence-Based Nursing 2000;3:68–70.

364. QSR International Pty Ltd.
NVivo qualitative data analysis software.
2018.

365. Hochbaum, GM and United States Public Health Service Division of Special Health Services.
Public participation in medical screening programs; a socio-psychological study.
Washington, 1958.

366. Ajzen, I.
The theory of planned behavior.
Organizational Behavior and Human Decision Processes 1991;50:179–211.

367. May, C and Finch, T.
Implementing, Embedding, and Integrating Practices: An Outline of Normalization Process Theory.
Sociology 2009;43:535–554.

368. Sekhon, M, Cartwright, M, and Francis, JJ.
Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework.
BMC Health Services Research 2017;17:88.

369. Iliffe, Iliffe, S, Lenihan, P, et al.
The development of a short instrument to identify common unmet needs in older people in general practice.
British Journal Of General Practice: The Journal Of The Royal College Of General Practitioners 20041201;54:914.

370. General Medical Council.

Domain 1: Knowledge, skills and performance.
In: *Good Medical Practice*.
Manchester, UK: Author, 2013:
6–9.
URL: https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/good-medical-practice.

371. Hibbard, JH, Mahoney, ER, Stockard, J, and Tusler, M.
Development and testing of a short form of the patient activation measure.
Health services research 2005;40:1918–30.

372. Bunn, F, Goodman, C, Sworn, K, et al.
Psychosocial factors that shape patient and carer experiences of dementia diagnosis and treatment: a systematic review of qualitative studies.
PLoS medicine 2012;9:e1001331.

373. Prorok, JC, Horgan, S, and Seitz, DP.
Health care experiences of people with dementia and their caregivers: a meta-ethnographic analysis of qualitative studies.
CMAJ: Canadian Medical Association Journal 2013;185:1–12.

374. Robinson, L, Gemski, A, Abley, C, et al.
The transition to dementia–individual and family experiences of receiving a diagnosis: a review.
International psychogeriatrics / IPA 2011;23:1026–43.

375. Bunn, F, Sworn, K, Brayne, C, Iliffe, S, Robinson, L, and Goodman, C.
Contextualizing the findings of a systematic review on patient and carer experiences of dementia diagnosis and treatment: A qualitative study.
Health Expectations 2015;18:740–753.

376. Korstjens, I and Moser, A.
Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing.
European Journal of General Practice 2018;24:120–124.

377. Dementia: assessment, management and support for people living with dementia and their carers NICE guideline [NG97].
Tech. rep.
London: (NICE)National Institute for Health and Care Excellence (NICE), 2018.
URL: https://www.nice.org.uk/guidance/ng97/.

378. Office for National Statistics.
How has life expectancy changed over time?
2015.

URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/articles/howhaslifeexpectancychangedovertime/2015-09-09.

379. Office for National Statistics.
Living longer how our population is changing and why it matters 13 How well connected are older people?
2018.
URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/livinglongerhowourpopulationischangingandwhyitmatters/2018-08-13.

380. Sikkes, S, Pera, A, Vries, S de, et al.
Cross-cultural differences in instrumental activities of daily living (IADL): Translations and adaptations of the amsterdam iadl questionnaire.
Alzheimer's & Dementia 2015;11:P293.

381. Fox, C, Lafortune, L, Boustani, M, and Brayne, C.
The pros and cons of early diagnosis in dementia.
The British journal of general practice : the journal of the Royal College of General Practitioners 2013;63:510–2.

382. 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.
The Cochrane database of systematic reviews 2019;12:CD013282.

383. NHS Digital.
Selected CCG information: NHS Bristol, North Somerset and South Gloucestershire CCG.
In: *General Practice Workforce*.
2019.
URL: https://app.powerbi.com/view?r=eyJrIjoiNmY4NGNiMWQtMGVkZi00MzU2LThiZGMtMTFlZjY2NG

384. Davis, DH, Creavin, ST, Noel-Storr, A, et al.
Neuropsychological tests for the diagnosis of Alzheimer's disease dementia and other dementias: a generic protocol for cross-sectional and delayed-verification studies.
The Cochrane database of systematic reviews 2013:CD010460.

385. James, BD, Boyle, PA, and Bennett, DA.
Correlates of susceptibility to scams in older adults without dementia.
Journal of elder abuse & neglect 2014;26:107–22.

386. McCray, GPJ, Titman, AC, Ghaneh, P, and Lancaster, GA.
Sample size re-estimation in paired comparative diagnostic accuracy studies with a binary response.
BMC Medical Research Methodology 2017;17:102.

387. Riley, RD, Ensor, J, Snell, KI, et al.
Calculating the sample size required for developing a clinical prediction model.
BMJ (Clinical research ed.) 2020;368:m441.

388. Kojima, G, Taniguchi, Y, Iliffe, S, and Walters, K.
Frailty as a Predictor of Alzheimer Disease, Vascular Dementia, and All Dementia Among Community-Dwelling Older People: A Systematic Review and Meta-Analysis.
Journal of the American Medical Directors Association 2016;17:881–888.

389. Sharma, M, Nazareth, I, and Petersen, I.
Trends in incidence, prevalence and prescribing in type 2 diabetes mellitus between 2000 and 2013 in primary care: a retrospective cohort study.
BMJ open 2016;6:e010210.

390. 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).
Programme Grants for Applied Research 2015;3:1–596.

391. Lieberman, D, Shvartzman, P, Korsonsky, I, and Lieberman, D.
Diagnosis of ambulatory community-acquired pneumonia.
Scandinavian Journal of Primary Health Care 2003;21:57–60.

392. Bösner, S, Haasenritter, J, Abu Hani, M, et al.
Accuracy of general practitioners' assessment of chest pain patients for coronary heart disease in primary care: cross-sectional study with follow-up.
Croatian medical journal 2010;51:243–249.

393. Carey, M, Jones, K, Meadows, G, et al.
Accuracy of general practitioner unassisted detection of depression.
Australian & New Zealand Journal of Psychiatry 2014;48:571–578.

394. Callahan, CM, Unverzagt, FW, Hui, SL, Perkins, AJ, and Hendrie, HC.
Six-item screener to identify cognitive impairment among potential subjects for clinical research.
Medical care 2002;40:771–81.

395. Brown, J, Pengas, G, Dawson, K, Brown, LA, and Clatworthy, P.
Self administered cognitive screening test (TYM) for detection of Alzheimer's disease: cross sectional study.
BMJ (Clinical research ed.) 2009;338:b2030.

396. Whiting, PF, Davenport, C, Jameson, C, et al.
How well do health professionals interpret diagnostic information? A systematic review.
BMJ open 2015;5:e008155.

397. Zhelev, Z, Garside, R, and Hyde, C.
A qualitative study into the difficulties experienced by healthcare decision makers when reading a Cochrane diagnostic test accuracy review.
Systematic reviews 2013;2:32.

398. Chithiramohan, A, Iliffe, S, and Khattak, I.
Identifying barriers to diagnosing dementia following incentivisation and policy pressures: General practitioners' perspectives.
Dementia 2019;18:514–529.

399. Koch, T, Iliffe, S, and EVIDEM-ED project.
Rapid appraisal of barriers to the diagnosis and management of patients with dementia in primary care: a systematic review.
BMC family practice 2010;11:52.

400. Iliffe, S, Manthorpe, J, and Eden, A.
Sooner or later? Issues in the early diagnosis of dementia in general practice: A qualitative study.
Family Practice 2003;20:376–381.

401. Iliffe, S, Wilcock, J, and Haworth, D.
Obstacles to Shared Care for Patients with Dementia: A qualitative study.
Family Practice 2006;23:353–362.

402. Manthorpe, J, Samsi, K, Campbell, S, 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 2013;63:69–75.

403. Greaves, I, Greaves, N, Walker, E, Greening, L, Benbow, SM, and Jolley, D.
Gnosall Primary Care Memory Clinic: Eldercare facilitator role description and development.
Dementia (London, England) 2013.

404. Lee, L, Hillier, LM, Stolee, P, et al.
Enhancing dementia care: a primary care-based memory clinic.
Journal of the American Geriatrics Society 2010;58:2197–204.

405. Meeuwsen, EJ, Melis, RJF, Van Der Aa, GCHM, et al.
Effectiveness of dementia follow-up care by memory clinics or general practitioners: randomised controlled trial.
BMJ (Clinical research ed.) 2012;344:e3086.

406. Baird, B.
Primary care networks explained.

Tech. rep.
London: Kings Fund, 2019.
URL: https://www.kingsfund.org.uk/publications/primary-care-networks-explained.

407. Maclin, JMA, Wang, T, and Xiao, S.
Biomarkers for the diagnosis of Alzheimer's disease, dementia Lewy body, frontotemporal dementia and vascular dementia.
General Psychiatry 2019;32:e100054.

408. Ranson, JM, Kuzma, E, Hamilton, W, and Llewellyn, DJ.
DECODE dementia: initial development and external validation of clinical prediction models for dementia identification.
In: *Alzheimer's & Dementia*.
Vol. 15.
7.
2019:
P704–P705.
DOI: 10.1016/j.jalz.2019.06.2714.

409. Craig, P, Dieppe, P, Macintyre, S, Michie, S, Nazareth, I, and Petticrew, M.
Developing and evaluating complex interventions: the new Medical Research Council guidance.
BMJ 2008:a1655.

410. O'Cathain, A, Croot, L, Duncan, E, et al.
Guidance on how to develop complex interventions to improve health and healthcare.
BMJ Open 2019;9:e029954.

411. Ferrante di Ruffano, L, Dinnes, J, Sitch, AJ, Hyde, C, and Deeks, JJ.
Test-treatment RCTs are susceptible to bias: a review of the methodological quality of randomized trials that evaluate diagnostic tests.
BMC medical research methodology 2017;17:35.

412. Hyde, C, Peters, J, Bond, M, et al.
Evolution of the evidence on the effectiveness and cost-effectiveness of acetylcholinesterase inhibitors and memantine for Alzheimer's disease: systematic review and economic model.
Age and ageing 2013;42:14–20.

413. NHS.
The NHS long term plan.
2019.
URL: https://www.longtermplan.nhs.uk/.

414. Larner, A.

Impact of the National Dementia Strategy in a neurology-led memory clinic: 5-year data. Clinical Medicine 2014;14:216–216.