Decline in cognition from mid-life improves specificity of mini-mental state examination: diagnostic test accuracy in Caerphilly Prospective Study (CaPs)

Sam Creavin, Mark Fish, Antony Bayer, John Gallacher, Yoav Ben-Shlomo

Author information

Dr Sam Creavin PhD MRCGP School of Social and Community Medicine, University of Bristol, Bristol. sam.creavin@bristol.ac.uk

Dr Mark Fish MD MRCP Department of Neurology, Royal Devon and Exeter Hospital, Exeter.

Professor Antony Bayer FRCP Division of Population Medicine, Cardiff University, Academic Centre, University Hospital Llandough, Cardiff.

Professor Yoav Ben-Shlomo MSc, PhD, MRCP School of Social and Community Medicine, University of Bristol, Bristol.

Professor John Gallacher PhD john.gallacher@psych.ox.ac.uk Department of Psychiatry Warneford Hospital Oxford OX3 7JX.

Sam Creavin and Yoav Ben-Shlomo had full access to all the data in the study, and take responsibility for the integrity of the data and the accuracy of the data analysis.

Declaration of interest All authors declare no interest

Running title: midlife cognitive measures enhance diagnostic accuracy in late life

Decline in cognition from mid-life improves specificity of mini-mental state examination: diagnostic test accuracy in Caerphilly Prospective Study (CaPs)

Abstract

Background The merit of using baseline cognitive assessments in mid-life to help interpret cross-sectional cognitive tests scores in later life is uncertain. We evaluated how accuracy for diagnosing dementia is enhanced by comparing cross-sectional results to a midlife measure.

Method Cohort study of 2,512 men with repeated measures of Mini Mental State Examination (MMSE) over approximately 10 years. Index test MMSE at threshold of 24 indicating normal, as a cross-sectional measure and in combination with decline in MMSE score from mid-life. Reference standard consensus clinical diagnosis of dementia by two clinicians according to Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV).

Results 1,150 men participated at phase 4 of whom 75 had dementia. A cross-sectional MMSE alone produced a sensitivity of 60% (50% to 70%) and specificity 95% (94% to 97%) with a threshold of ≥ 24 points indicating normal. For lower-scoring men in late life, with cross sectional scores of <22, combining cross-sectional AND a three-point or more decline over time had a sensitivity of 52% (39% to 64%) and specificity 99% (99% to 100%). For higher-scoring men in later life, with cross sectional scores <26 combining cross-sectional OR decline of at least three points had a sensitivity of 98% (92% to 100%) and specificity 38% (32% to 44%).

Conclusion It may be helpful in practice to formally evaluate cognition in mid-life as a baseline to compare with if problems develop in future, as this may enhance diagnostic accuracy and classification of people in later life.

Introduction.

A range of cognitive tests are available to help as part of the diagnostic work up for dementia [1], but even the best known of these carry a significant risk of false positives (FPs) [2]. FP tests may generate unnecessary concern while further investigations, which may themselves cause or risk morbidity, are

required to confirm or refute the diagnosis of dementia. Strategies to help reduce the number of FPs who require further workup are important because some health services have reported a rise in the number of people who are being evaluated for possible dementia, but a fall in the proportion who are true positives (TP)s [3,4]. Intuitively, comparing objective measures of cognition to a previous baseline objective measure should improve diagnostic accuracy. Our aim was to test the hypothesis that we could improve diagnostic accuracy compared to using later-life cross-sectional measures alone by also considering a personal mid-life measure of cognition, and that this would reduce both false-positives and false-negatives. We used data from the Caerphilly Prospective Study (CaPS).

Methods

Participants

CaPS is a cohort study of men that was established to investigate cardiovascular disease [5]. Men aged 45–59 years were identified from the electoral roll and general practice lists in Caerphilly and adjoining villages in South Wales, UK. Initial participation rate was 89% and 2,512 men were examined in phase 1 (July 1979 to September 1983) and then followed-up at regular intervals. Social class was assessed at phase 2 based on current or most recent occupation according to the 2010 United Kingdom Office for National Statistics[6] : I professionals; II managerial and technical occupations; IIIa skilled occupations (non-manual); IIIb skilled occupations (manual); IV partly skilled occupations; V unskilled occupation. Cognition was assessed at phases 3 (November 1989 to September 1993), 4 (October 1993 to February 1997), and 5 (September 2002 to June 2004) using tests including the Mini Mental State Examination (MMSE) [7]. At phase 5 all 300 men (of whom 205 attended) who met screening criteria (CAMCOG <83 or decline in CAMCOG of 10 points or more between any two measurement periods, or failure to complete CAMCOG) [8], as well as the oldest 47 of the 925 men who screened negative (of whom 45 were seen), were invited for a clinical evaluation in their home or research clinic by a neurologist.

Written informed consent was obtained from all participants. Approval for the original Caerphilly Prospective Study was obtained from the Research Ethics Committee in South Glamorgan, Wales (01/69).

Approval for the phase five follow-up study was obtained from the Research Ethics Committee in Gwent, Wales.

Mini Mental State Examination

At each of phases 3 to 5, MMSE was conducted in a standardised manner by a trained research assistant who was not aware of the final clinical diagnosis at phase 5. We averaged MMSE score at phases 3 and 4 where possible, to reduce measurement error in assessing mid-life cognition, but did not exclude participants who only completed one mid-life measure. We calculated decline in MMSE by subtracting MMSE at phase 5 from the average mid-life MMSE. We summarised decline in MMSE score and identified men with a decline in MMSE score at least 3 to indicate a reliable change in function [9]. We identified men who attended at phase 3 or phase 4, but who did not participate at phase 5; almost all had died in the interim period.

Index test and reference standard

The typical MMSE threshold is 24 indicating normal[2]. We evaluated the diagnostic accuracy of each MMSE threshold from a range of 20 to 30 indicating normal, since we judged the impact of decline in cognition on diagnostic accuracy might vary across thresholds i.e., a decline of 3 points from midlife might be more important in someone who at phase 5 scored 26 rather than 18. The index tests indicating abnormal were (a) phase 5 MMSE below threshold (b) phase 5 MMSE below threshold *and* mid-life MMSE decline of at least 3 or at least 4 points (c) phase 5 MMSE below threshold *or* mid-life MMSE decline of at least 3 or 4 points (d) decline in MMSE score between 1-5 points regardless of baseline score.

The reference standard was a consensus diagnosis of dementia made by two clinicians with specialist training in memory disorders using the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) [10], after reviewing all relevant information including medical records (including investigations where available), the clinical assessment where available, and MMSE score.

Statistical methods

We calculated the diagnostic accuracy of phase 5 MMSE using standard measures namely sensitivity, and specificity. All analyses were conducted in Stata (version 13). Sensitivity analyses (a) computed the diagnostic accuracy for the three index tests when the four men who participated at phase 5 but had missing values for MMSE were instead scored as 0 for MMSE at phase 5 (b) derived accuracy stratified by mid-life MMSE scores (24 or more) of (i) decline of at least three points and (ii) phase 5 MMSE at a threshold of 24. We also calculated the sensitivity and specificity of using the age-appropriate 5th centile for men [11].

Results

Figure 1 goes here

MMSE was completed by 1,853 of the 2,429 men who were alive at phase 3, 1,498 of the 2,138 men who were alive at phase 4, and 1,221 of the 1,225 men who were alive and contactable at phase 5, of whom 75 men had dementia. Most men (717/1,223) who participated at phase 3 or 4 but not phase 5 had died in the interval. A further 295 men refused contact and 211 could not be traced, for a total of 1,731 men who were still alive. (See figure 1 for flow chart). 387 subjects were seen at a research memory clinic resulting in the prospective identification of 75 dementia cases (25 dementia had already been diagnosed through death certificates and medical records and were not reassessed.)

Table 1 goes here

Table 1 shows the median age at each phase, median MMSE score, and social class of men stratified by participation and presence of dementia at phase 5. The 1,223 men who participated at phase 3 or phase 4 but did not participate at phase 5 were older (median age of 63 years at phase 3), compared to the 1,150 men who participated at phase 5 and did not have dementia (60 years) but younger than the 75 men who participated at phase 5 and had dementia (66 years) At phase 5, the median age for men who had dementia was 78 years compared to 72 years for men who did not have dementia. The median interval between follow-up was similar in men who did or did not have dementia at phase 5.

Although the median MMSE for the whole group of men was 27 at each phase, this concealed a substantial decline in MMSE for men who had dementia at phase 5, who also had lower MMSE scores at baseline than men who did not have dementia at phase 5. The median MMSE at phase 3 was 27 for men who did not participate at phase 5, 25 for men who had dementia at phase 5, and 27 for men who did not have dementia at phase 5.

Table 2 goes here

Table 2 shows the diagnostic accuracy for each of the index tests described in *Methods*. At a threshold of 24 indicating normal, the diagnostic accuracy of MMSE at phase 5 was sensitivity 60% (95% CI 50% to 70%) specificity 95% (95% CI 94% to 97%). Lowering or raising the threshold had the inevitable effect of either reducing sensitivity but increasing specificity or vice versa respectively, and consistent with previous studies the conventional threshold of ≥ 24 was optimal.

Combining MMSE at phase 5 at a threshold of 24 *and* decline in MMSE of at least 3 points resulted in improved specificity 99% (95% CI 98% to 99%) and a slight decline in sensitivity 58% (95% CI 45% to 70%). Combining MMSE at phase 5 *and* decline led to a slight improvement in both sensitivity and specificity (which was already high) at thresholds lower than 24, compared to the same threshold at phase 5 alone, whereas at thresholds above 24, *and* decline led to an improvement in specificity but a decline in sensitivity. There was a similar pattern for a decline of at least 4 points, being more substantial for scores of above 24 (see Supplemental Table 1).

In contrast, the combined test *either* MMSE at phase 5 with a threshold of 24 *or* a decline of at least 3 points led to a marked improvement in sensitivity 98% (95% CI 91% to 100%) compared to phase 5 alone at the same threshold, but a substantial decline in specificity 18% (95% CI 10% to 27%). Thresholds below 24 combined *or* decline had an improvement in sensitivity but decline in specificity compared to using MMSE at phase 5 alone at threshold of 24, whereas at thresholds above 24 *or* decline had improved in sensitivity, but relatively preserved specificity. There was a similar pattern for *either* MMSE at phase 5 with a threshold of 24 *or* a decline of at least 4 points (see Supplemental Table 1).

Table 3 goes here

7

Table 3 shows that the diagnostic accuracy for decline in midlife cognition of 1 point or more for the whole sample produced a sensitivity of 85% (95% CI 74% to 93%) and specificity 69% (95% CI 66% to 72%), and that specificity increased as decline in midlife cognition worsened with a reduction in sensitivity. In men with a phase 5 MMSE of 24 or more, the diagnostic accuracy of decline of at least one point was sensitivity 67% (30% to 93%) specificity 72% (68% to 75%). In contrast, in men with a phase 5 MMSE of less than 24 the diagnostic accuracy of decline of at least four points was sensitivity 69% (95% CI 55% to 81%) and specificity was 79% (95% CI 63% to 90%). Therefore, in men with normal (≥24) MMSE scores at phase 5, a relatively small decline in MMSE score from mid-life (1 point) improved detection of people with dementia compared to phase 5 measures alone, whereas in men with abnormal (<24) scores at phase 5, the accuracy of decline in MMSE score from mid-life was less favourable than the cross-sectional score alone. In terms of the trade-offs between sensitivity and specificity in natural frequencies, Table 3 shows that the best trade-off is for a 3-point decline in people who score 24 or more at phase 5, where compared to using a cross-sectional MMSE alone there are 10 additional TP at a cost of 36 additional FP (3.6 additional FP per 1 additional TP), and a 3-point decline overall has an additional 2 TP for 0 additional FP. However, using a 3 point decline in people with a score of less than 24 at phase 5 generates many more FP than TP, even with a 4-point decline there are an additional 17.3 FP per 1 additional TP, compared to MMSE at phase 5 alone.

The results of the sensitivity analysis where scores on phase 5 MMSE for four men who had missing data for this item were replaced with zero were identical to the main analysis.

We also did an analysis stratified by scores on *mid-life* MMSE (see Supplemental Table 2). Supplemental Table 2 shows that in an analysis restricted to men who had a midlife MMSE of less than 24, the diagnostic accuracy of decline of at least three points was sensitivity 48% (95% CI 28% to 69%) specificity 96% (95% CI 82% to 100%) and the accuracy of phase 5 MMSE at a threshold of 24 indicating normal was sensitivity 82% (95% CI 63% to 94%) specificity 50% (95% CI 31% to 69%). Therefore, in men with abnormal mid-life MMSE scores, decline from mid-life to late life had lower sensitivity but greater specificity compared to using cross-sectional late life cognition at phase 5 alone. In contrast, in an analysis restricted to men who had a midlife MMSE of more than 24, the diagnostic accuracy of decline of

at least three points was sensitivity 71% (95% CI 53% to 85%) specificity 95% (95% CI 93% to 96%) and the accuracy of phase 5 MMSE at a threshold of 24 indicating normal was sensitivity 56% (95% CI 41% to 71%) specificity 97% (95% CI 97% to 99%). Therefore, in men with normal mid-life MMSE scores, decline from mid-life to late life had higher sensitivity and similar specificity compared to using cross-sectional late life cognition at phase 5 alone.

When we used age-sex specific thresholds the diagnostic accuracy was sensitivity 40% (95% CI 30% to 50%) specificity 99% (95% CI 98% to 99%).

Discussion

Summary

This study found that cross-sectional diagnostic test accuracy of MMSE in men was improved by taking account of decline in cognition using data from an earlier time-point. At higher cross-sectional test thresholds (e.g., 29 indicating normal), sensitivity was improved by combining *or* decline to identify people with a presumed high baseline level of function who might otherwise be missed. At lower cross-sectional test thresholds (e.g., 23 indicating normal) specificity was improved by combining *and* decline to identify people with a presumed low baseline level of cognition who might otherwise be over-diagnosed. Using decline in cognition of at least three points as a single test had similar test accuracy to using MMSE at a threshold of 24 indicating normal. Test accuracy of various definitions of decline in cognition had higher sensitivity but similar specificity in men who scored more than 24 points in midlife than men who scored less than this.

Strengths and limitations

The key strength of this investigation is the repeated standardised measures of cognition between mid-life and later life, and the blinding of the index test to the reference standard. However, CaPS was not designed as a diagnostic test accuracy study and our results should be interpreted with caution. Our results need confirmation in other settings in studies that include women and are more ethnically diverse. The original purpose of the CaPs cohort was to investigate risk factors for cardiovascular disease

rather than to identify or diagnose dementia but as the men became older cognitive function and later screening for dementia was added prospectively to the protocol. The population were relatively homogenous and had generally left school at a relatively early age, typical of birth cohorts of this period.

The index test was conducted blind to the reference test but is subject to incorporation bias and this will tend to increase the diagnostic accuracy of the test. This is not uncommon in studies of diagnostic accuracy in dementia. The cut point for decline in MMSE score between midlife (phase 3 and 4) and later life (phase 5) could have been chosen arbitrarily but we selected a point based on published suggestions from studies that investigated reliability of MMSE in prolonged follow-up [9]. MMSE is now copyright, and a charge is made for using it in clinical practice. Intuitively it seems likely that our approach would translate to the use of other brief cognitive tests, or indeed computer-based tests, but this should be evaluated before large scale adoption of this approach.

The reference test is a recognised standard for dementia in studies of test accuracy in dementia but does not incorporate neuroimaging and biomarker assay, both of which are of arguably of more benefit in identifying the aetiology of dementia than the presence of the syndrome, according to DSM-IV, and neither of which were in typical clinical or research practice in the UK at the time. Incorrect classification of participants regarding dementia would reduce the diagnostic accuracy of the index tests due to increasing random error in the data, but arguably the main limitation of our reference standard is the inability to explore heterogeneity by aetiological subtype.

Men with dementia at phase 5 may have already been sufficiently impaired at phase 3 and 4 to have dementia, though not been diagnosed, and this would increase the specificity of our index tests. However, some of these men might have died before phase 5 since survival in people with dementia is around 6 years from onset [12]. We clinically evaluated some, but not all, surviving men who scored in the normal range on MMSE at phase 5. However, it is fairly common in population based studies to only examine a sample of participants who score in the normal range on cognitive screening tests [2]. Our analysis indicated that the procedures identified 98% of men with dementia. It is possible that the screening procedures may have altered the diagnostic accuracy of both cross sectional and longitudinal

measures on MMSE, since men with high scores who did not decline were less likely to receive the reference standard and there may therefore be incomplete ascertainment of dementia in people who scored well at phase 5 and had not declined, this would tend to lead the estimates of test accuracy that we report to be over-optimistic.

Comparison to existing literature

In a sample of Swedish adults aged 85 the median change in MMSE scores over 3 years was -3.0 and a decline of 4 points or more between the two examinations had sensitivity of 83% and specificity of 80% for diagnosing dementia [13]. Another study in Dutch adults aged 65 to 84 years reported that a decline of 5 points or more over one year had a sensitivity of 43% and specificity of 97% for diagnosing dementia [14]. Neither of these studies reported the accuracy of combining the cross-sectional MMSE and MMSE decline. Other studies have reported that lower scores on tests of cognition at baseline are associated with dementia diagnosis at follow up [15,16], and that accelerated cognitive decline might be more strongly associated with pre-clinical dementia than a low score on a cross sectional test [17]. In another study, memory and executive function declined most in people with pre-clinical AD [18]. However, none of these studies reported how the diagnostic accuracy of a cross sectional test was affected by use of measures of decline.

In our study, use of age-sex specific thresholds for MMSE increased specificity compared to use of a standard threshold, but at a cost to sensitivity. When we used age-sex specific thresholds the diagnostic accuracy was sensitivity 40% (95% CI 30% to 50%) specificity 99% (95% CI 98% to 99%). In contrast keeping specificity constant (99%, 95% CI 98% to 100%) then using 5-point decline in MMSE points had a sensitivity of 54% (95% CI 41% to 67%), and at the optimal threshold (3-point decline) diagnostic accuracy was sensitivity 62% (95% CI 49% to 74%) specificity 95% (95% CI 93% to 96%). While age and sex specific thresholds may initially seem similar to our approach of using a midlife measure to derive decline, they may be subject to a generational effect and less able to distinguish between outliers within generations; that is, different generational cohorts may have higher or lower age specific thresholds due to the effect of education. For any one individual, the age-sex specific threshold for that generational cohort is unlikely

to be known, and indeed even level of education may be unknown if the person does not recall and they have no informant, whereas if midlife measures were recorded systematically these could easily be used in future for individuals. For people with premorbid cognitive function which is exceptionally high or low, using a decline measure from mid-life allows for a self-controlled assessment of cognition, whereas an age-adjusted measure cannot account so well for people who are outliers in mid-life performance.

Implications

A primary care clinician who applies single thresholds for tests may inappropriately reassure a patient who scores above the threshold for dementia, whereas when considering decline from a previous higher level of functioning the clinician might decide to recommend further tests. Alternatively, a primary care clinician might incorrectly diagnose dementia in a patient who scores below a certain threshold, whereas the clinician would avoid this false-positive diagnosis when taking decline into account.

Though screening for dementia is not recommended [19,20] we believe that these results would support the routine assessment of cognition in mid-life to provide a baseline for future tests to be compared against should symptoms or concerns develop. Decision aids for patients and clinicians who were considering this approach would need to convey the message that the test was being conducted to act as a baseline for future testing, not as a test for diagnosis or prognosis in mid-life. We acknowledge that general health check-ups are of uncertain benefit, and may cause harm [21], but nevertheless have been implemented as policy in both the UK and USA. We believe the potential benefit of this approach is the reduction of misclassifications or incorrect diagnoses, which has benefit to the individual and to society. Since national guidelines do not currently recommend screening for dementia, we focus the impact on a person presenting with concern about cognitive symptoms.

Measures incorporating decline might have particular benefit in identifying additional people with high baseline cognition who might otherwise be missed. Future research could identify the optimal and timing of mid-life cognitive testing, as well as identify public perceptions about this approach.

References

[1]   Tsoi KKF, Chan JYC, Hirai HW, Wong SYS, Kwok TCY (2015) Cognitive Tests to Detect Dementia. *JAMA Intern Med*.

[2]   Creavin ST, Wisniewski S, Noel-Storr AH, Trevelyan CM, Hampton T, Rayment D, Thom VM, Nash KJE, Elhamoui H, Milligan R, Patel AS, Tsivos D V, Wing T, Phillips E, Kellman SM, Shackleton HL, Singleton GF, Neale BE, Watton ME, Cullum S (2016) Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database Syst Rev* **1**, CD011145.

[3]   Bell S, Harkness K, Dickson JM, Blackburn D (2015) A diagnosis for ??55: What is the cost of government initiatives in dementia case finding. *Age Ageing* **44**, 344–345.

[4]   Larner a. J (2014) Impact of the National Dementia Strategy in a neurology-led memory clinic: 5-year data. *Clin Med J R Coll Physicians London* **14**, 216.

[5]   The Caerphilly and Speedwell Collaborative Group (1984) Caerphilly and Speedwell collaborative heart disease studies. The Caerphilly and Speedwell Collaborative Group. *J Epidemiol Community Health* **38**, 259–62.

[6]   Current standard classifications, Last updated 2010, Accessed on 2010.

[7]   Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189–98.

[8]   Fish M, Bayer A, Gallacher JEJ, Bell T, Pickering J, Pedro S, Dunstan FD, Ben-Shlomo Y, Ebrahim S (2008) Prevalence and pattern of cognitive impairment in a community cohort of men in South Wales: methodology and findings from the Caerphilly Prospective Study. *Neuroepidemiology* **30**, 25–33.

[9]   Hensel A, Angermeyer MC, Riedel-Heller SG (2007) Measuring cognitive change in older adults: reliable change indices for the Mini-Mental State Examination. *J Neurol Neurosurg Psychiatry* **78**, 1298–1303.

[10]   American psychiatric Association (1987) *Diagnostic and statistical manual of mental disorders*, Author, Washington, DC.

[11]   Huppert F a, Cabelli ST, Matthews FE (2005) Brief cognitive assessment in a UK population sample --

distributional properties and the relationship between the MMSE and an extended mental state examination. *BMC Geriatr* **5**, 7.

[12]   Rait G, Walters K, Bottomley C, Petersen I, Iliffe S, Nazareth I (2010) Survival of people with clinical diagnosis of dementia in primary care: cohort study. *Bmj* **341**, c3584.

[13]   Aevarsson O, Skoog I (2000) A longitudinal population study of the mini-mental state examination in the very old: relation to dementia and education. *Dement Geriatr Cogn Disord* **11**, 166–175.

[14]   Schmand B, Lindeboom J, Launer L, Dinkgreve M, Hooijer C, Jonker C (1995) What is a significant score change on the mini-mental state examination? *Int J Geriatr Psychiatry* **10**, 411–414.

[15]   Elias MF, Beiser A, Wolf P a, Au R, White RF, D'Agostino RB (2000) The preclinical phase of alzheimer disease: A 22-year prospective study of the Framingham Cohort. *Arch Neurol* **57**, 808–813.

[16]   Small BJ, Fratiglioni L, Viitanen M, Winblad B, Backman L (2000) The course of cognitive impairment in preclinical Alzheimer disease: three- and 6-year follow-up of a population-based sample. *Arch Neurol* **57**, 839–844.

[17]   Laukka EJ, MacDonald SWS, Bäckman L (2006) Contrasting cognitive trajectories of impending death and preclinical dementia in the very old. *Neurology* **66**, 833–838.

[18]   Chen P, Ratcliff G, Belle SH, Cauley JA, DeKosky ST, Ganguli M (2010) Patterns of cognitive decline in presymptomatic Alzheimer disease: a prospective community study. *Arch Gen Psychiatry* **58**, 853–8.

[19]   Note of the meeting held on the 27 November 2014, Last updated 2014, Accessed on 2014.

[20]   US Preventive Services Task Force, Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, Caughey AB, Doubeni CA, Epling JW, Kubik M, Landefeld CS, Mangione CM, Pbert L, Silverstein M, Simon MA, Tseng C-W, Wong JB (2020) Screening for Cognitive Impairment in Older Adults: US Preventive Services Task Force Recommendation Statement. *JAMA* **323**, 757–763.

[21]   Krogsbøll LT, Jørgensen KJ, Grønhøj Larsen C, Gøtzsche PC (2012) General health checks in adults for reducing morbidity and mortality from disease. In *Cochrane Database of Systematic Reviews*, Krogsbøll LT, ed. John Wiley & Sons, Ltd, Chichester, UK.

[22]   Creavin S, Fish M, Gallacher J, Bayer A, Ben-Shlomo Y (2015) Clinical history for diagnosis of dementia

in men: Caerphilly Prospective Study. *Br J Gen Pract* **65**, e489–e499.

Table 1: Demographics of men and MMSE score at each of phases 3 − 5

| | Men who participated at phase 3 or phase 4 but did not participate at phase 5

n=1,223 | Men who participated at phase 5, with dementia at phase 5

n=75 | Men who participated at phase 5, without dementia at phase 5

n=1,150 |
|---|---|---|---|
| **Median age (years) at phase (IQR)** | | | |
| 3 | 63 (59 to 67) | 66 (63 to 68) | 60 (58 to 64) |
| 4 | 67 (63 to 71) | 70 (69 to 72) | 64 (61 to 68) |
| 5 | - | 78 (75 to 80) | 72 (70 to 76) |
| **Median interval (years)** | | | |
| From phase 3 to phase 5 | - | 11.8 (10.9 to 12.7) | 11.8 (10.8 to 12.8) |
| From phase 4 to phase 5 | - | 8.0 (7.4 to 8.9) | 8.2 (7.4 to 8.9) |
| From phase 3 to phase 4 | 3.8 (3.5 to 3.9) | 3.8 (3.5 to 3.9) | 3.7 (3.4 to 3.9) |
| **Social Class % (n)** | | | |
| I or II | 18 (181) | 14 (10) | 29 (263) |
| IIIa or IIb | 58 (575) | 74 (52) | 59 (538) |
| IV or V | 24 (241) | 11 (8) | 13 (118) |
| **Median (IQR) MMSE** | | | |
| phase 3 | 27 (25 to 28) | 25 (22 to 27) | 27 (26 to 29) |
| phase 4 | 27 (25 to 28) | 25 (22 to 27) | 28 (26 to 29) |
| phase 5 | - | 19 (14 to 23) | 28 (26 to 29) |
| Decline | - | 5.5 (2 to 9.5) | 0 (-1 to 1) |

*four men at phase 5 did not have an MMSE

Table 2: diagnostic accuracy of phase 5 MMSE alone compared to phase 5 MMSE combined with decline of at least 3 points between mid-life and phase 5

MMSE Threshold *	Phase 5 MMSE at threshold		Phase 5 MMSE at threshold AND decline †		Phase 5 MMSE at threshold, OR decline †	
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
20	38 (28 to 48)	99 (99 to 100)	45 (33 to 57)	100 (100 to 100)	94 (83 to 99)	2 (0 to 11)
21	44 (34 to 54)	99 (98 to 99)	48 (36 to 61)	100 (99 to 100)	96 (87 to 100)	6 (1 to 16)
22	48 (38 to 58)	98 (97 to 99)	52 (39 to 64)	99 (99 to 100)	96 (87 to 100)	2 (0 to 10)
23	52 (42 to 62)	98 (97 to 99)	54 (41 to 66)	99 (98 to 100)	92 (82 to 97)	22 (13 to 34)
24	**60 (50 to 70)**	**95 (94 to 97)**	**58 (45 to 70)**	**99 (98 to 99)**	**98 (91 to 100)**	**18 (10 to 27)**
25	64 (54 to 73)	92 (90 to 94)	61 (48 to 73)	98 (97 to 99)	98 (92 to 100)	39 (32 to 47)
26	65 (55 to 74)	84 (81 to 86)	61 (48 to 73)	96 (95 to 98)	98 (92 to 100)	38 (32 to 44)
27	66 (56 to 75)	70 (67 to 73)	61 (48 to 73)	95 (94 to 97)	97 (90 to 100)	35 (31 to 40)
28	69 (59 to 78)	52 (49 to 56)	62 (49 to 74)	95 (94 to 96)	96 (88 to 99)	30 (26 to 34)
29	72 (62 to 81)	30 (28 to 33)	62 (49 to 74)	95 (93 to 96)	100 (95 to 100)	21 (19 to 24)
30	72 (62 to 81)	11 (9 to 13)	62 (49 to 74)	95 (93 to 96)	100 (95 to 100)	10 (8 to 12)

* i.e. for row 24, 24 is the lowest normal value, which is the typical threshold

† Decline = decline of 3 points or more from mid-life (average phase 3/4) to later life (phase 5) MMSE

CI, Confidence Interval

Table 3: diagnostic accuracy of decline in MMSE between mid-life and phase 5

Decline in MMSE *	Sensitivity	Specificity	Additional TP †	Additional FP †
1	85 (74 to 93)	69 (66 to 72)	23	236
2	77 (65 to 89)	86 (84 to 89)	16	82
3	62 (49 to 74)	95 (93 to 96)	2	0
4	59 (46 to 71)	98 (97 to 99)	-1	-27
5	54 (41 to 67)	99 (98 to 100)	-6	-36
Analysis in people who score 24 or more at phase 5 (n=951)				
1	67 (30 to 93)	72 (68 to 75)	62	254
2	22 (3 to 60)	89 (86 to 91)	20	100
3	11 (0.3 to 48)	96 (95 to 98)	10	36
4	0 (0 to 34)	99 (98 to 100)	0	9
5	0 (0 to 34)	99 (99 to 100)	0	9
Analysis in people who score less than 24 at phase 5 (n=105)				
1	89 (77 to 96)	16 (6 to 31)	27	717
2	87 (74 to 94)	42 (26 to 59)	25	481
3	71 (57 to 83)	63 (46 to 78)	10	290
4	69 (55 to 81)	79 (63 to 90)	8	145
5	64 (49 to 76)	87 (72 to 96)	4	73
* Between midlife (phase 3 and 4 average) and later life (phase 5) † compared to 24 indicating normal cross sectional in 1000 people with prevalence of dementia at prevalence of 9.3% (based on this sample)				

Table 4: accuracy of 3 point decline and phase 5 MMSE, stratified by mid-life MMSE scores of ≥24 and < 24

Diagnostic Accuracy stratified by mid-life MMSE score				
	Mid life MMSE score < 24		Mid life MMSE score ≥ 24	
Index test	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
≥ 3 point decline	48 (28 to 69)	96 (82 to 100)	71 (53 to 85)	95 (93 to 96)
Phase 5 MMSE of <24	82 (63 to 94)	50 (31 to 69)	56 (41 to 71)	97 (97 to 99)
CI Confidence interval MMSE Mini Mental State Examination				

Figure 1: flow chart of inclusion of men in the study after [22] with permission from journal
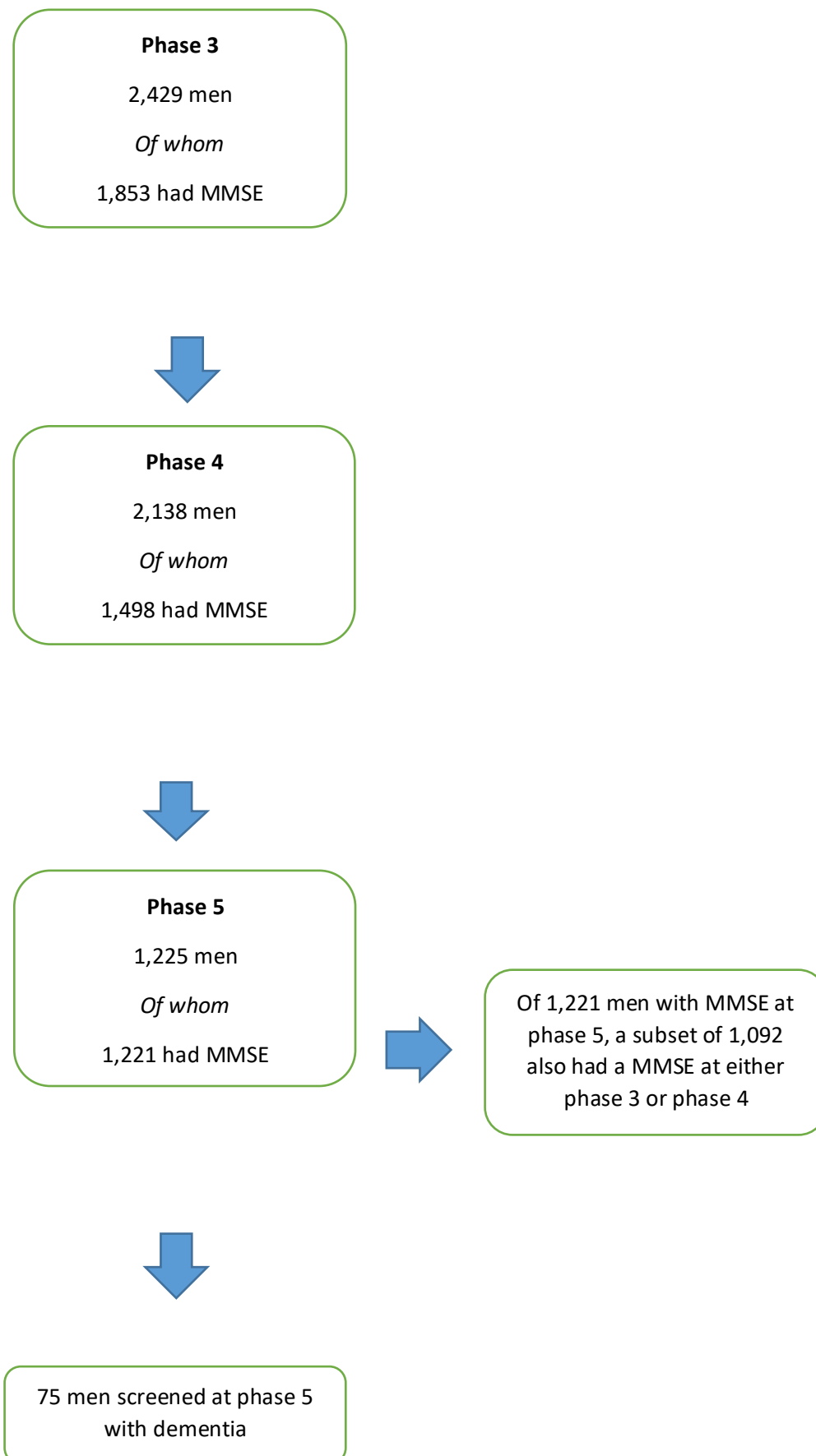
Supplementary Table 1: diagnostic accuracy of phase 5 MMSE alone compared to phase 5 MMSE

combined with decline of 4 points or more between mid-life and phase 5

MMSE 5 Threshold (value indicating normal)*	Phase 5 MMSE at threshold		Phase 5 MMSE at threshold AND decline of 4 points or more from mid-life (average phase 3/4) to later life (phase 5) MMSE		Phase 5 MMSE at threshold, OR a decline of four points or more from mid-life (average phase 3/4) to later life (phase 5) MMSE	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
20	38 (28 to 48)	99 (99 to 100)	45 (33 to 57)	100 (100 to 100)	94 (82 to 99)	8 (1 to 27)
21	44 (34 to 54)	99 (98 to 99)	48 (36 to 61)	100 (99 to 100)	96 (86 to 100)	11 (2 to 29)
22	48 (38 to 58)	98 (97 to 99)	52 (39 to 64)	99 (99 to 100)	96 (87 to 100)	4 (0 to 18)
23	52 (42 to 62)	98 (97 to 99)	54 (41 to 66)	99 (98 to 100)	90 (79 to 96)	41 (28 to 56)
24	**60 (50 to 70)**	**95 (94 to 97)**	**56 (43 to 69)**	**99 (98 to 100)**	**97 (89 to 100)**	**27 (17 to 39)**
25	64 (54 to 73)	92 (90 to 94)	58 (45 to 70)	99 (98 to 99)	98 (92 to 100)	47 (39 to 55)
26	65 (55 to 74)	84 (81 to 86)	58 (45 to 70)	98 (97 to 99)	98 (92 to 100)	41 (35 to 47)
27	66 (56 to 75)	70 (67 to 73)	58 (45 to 70)	98 (97 to 99)	97 (90 to 100)	35 (31 to 40)

28	69 (59 to 78)	52 (49 to 56)	59 (46 to 71)	98 (97 to 99)	96 (88 to 99)	30 (26 to 34)
29	72 (62 to 81)	30 (28 to 33)	59 (46 to 71)	98 (97 to 99)	100 (95 to 100)	21 (19 to 24)
30	72 (62 to 81)	11 (9 to 13)	45 (33 to 57)	100 (100 to 100)	100 (95 to 100)	10 (8 to 12)
* i.e. 24 is the lowest normal value, which is the typical threshold						