11-2014

The Mini-Mental State Examination as a diagnostic and screening test for delirium: systematic review and meta-analysis

Alex J. Mitchell

Deepak Shukla

Hafsa A. Ajumal

Brendon Stubbs

Tayyeb Tahir

General Hospital Psychiatry 36 (2014) 627–633

Contents lists available at ScienceDirect

General Hospital Psychiatry

journal homepage: http://www.ghpjournal.com

The Mini-Mental State Examination as a diagnostic and screening test for delirium: systematic review and meta-analysis

Alex J. Mitchell, M.D. [a,*], Deepak Shukla, M.B.B.S. [a], Hafsa A. Ajumal [b], Brendon Stubbs, M.Sc., MCSP [c], Tayyeb A. Tahir, M.B.B.S., M.D. [d,e]

[a] Leicester General Hospital, Leicester Partnership Trust, Leicester LE5 4PW, UK
[b] University of Leicester, Leicester, UK
[c] University of Greenwich, London, UK
[d] University Hospital of Wales, Cardiff, UK
[e] Vale University Health Board, Heath Park, Cardiff, UK

ARTICLE INFO

Article history:
Received 13 March 2014
Revised 1 September 2014
Accepted 3 September 2014

Keywords:
Delirium
Mini-mental state
Confusion
Screening
Older adults

ABSTRACT

Objective: To analyse the evidence concerning the accuracy of the Mini-Mental State Examination (MMSE) as a diagnostic and screening test for the presence of delirium in adults.
Method: Two authors searched MEDLINE, PsychINFO and EMBASE from inception till March 2014. Articles were included that investigated the diagnostic validity of the MMSE to detect delirium against standardised criteria. A diagnostic validity meta-analysis was conducted.
Results: Thirteen studies were included representing 2017 patients in medical settings of whom 29.4% had delirium. The meta-analysis revealed the MMSE had an overall sensitivity and specificity estimate of 84.1% and 73.0%, but this was 81.1% and 82.8% in a subgroup analysis involving robust high quality studies. Sensitivity was unchanged but specificity was 68.4% (95% CI=50.9–83.5%) in studies using a predefined cutoff of <24 to signify a case. In high-risk samples where delirium was present in 25% of patients, then the Positive predictive value and Negative predictive value would be 50.9% (48.3–66.2%) and 93.2% (90.0–96.5%).
Conclusion: The MMSE cannot be recommended as a case-finding confirmatory test of delirium, but may be used as an initial screen to rule out high scorers who are unlikely to have delirium with approximately 93% accuracy.

1. Introduction

Delirium is a common and pervasive neuropsychiatric condition [1] and the term has been used for acute confusion in the *International Classification of Diseases, 10th Revision* (*ICD-10*) [2]and the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition* (*DSM-IV*) [3]. A number of features defining delirium include rapid onset of symptoms that tend to fluctuate even during the same day with an altered level of consciousness, global disturbance of cognition or perceptual abnormalities with evidence of a physical cause, substance intoxication/withdrawal or multiple etiologies. The presence of delirium causes great concern since people affected have worse outcomes including longer hospital stays,[4,5] high risk of dementia,[6]higher rate of hospital-acquired complications, such as, falls and pressure sores[7,8] and increased mortality [9–11]. In addition, delirium complicates between 17–61% of major surgical procedures [12].

Many older adults are affected by delirium, for instance up to 50% of hospitalized patients can be diagnosed with delirium [13]. The prevalence of delirium on medical wards in hospital is about 3% to 30% [14,15] whilst it other research has demonstrated it may affect between 11–42% of general medical inpatients [13]. Delirium is also problematic at end of life care and may affect up to 83% of older adults [12]. Within the literature, there is a large variation in reporting incidence and prevalence rates of delirium [16–19]. There are numerous reasons that may account for this variability in rates including the source of sample, nature and variety of symptoms, diagnostic criteria and methods used.

Delirium risk is higher in palliative care, intensive care and in patients undergoing cardiothoracic surgery, emergency orthopedic procedures (repair of a hip fracture), vascular surgery or cataract removal [20,21]. Despite the pronounced prevalence and impact of delirium, healthcare professionals ability recognize it is poor with around 50% of cases of delirium going unrecognized [12,22,23]. This is exemplified in one recent study where emergency physicians missed delirium in 76% of cases [24]. In another study in an intensive care unit, nurses' detection sensitivity was 27% and specificity 92%, compared with the Confusion Assessment Method (CAM) for the intensive care unit (ICU) [25]. The fact that delirium is common, troublesome but under-recognized, suggests a role for screening instruments [26,27].

In recognition of this, recent guidelines (NICE, 2010) [28] stipulate that all elderly people admitted to hospital or in long-term care units should be screened for risk factors of developing delirium and cognitive

* Corresponding author. Tel.: +44 116 2463470.
E-mail addresses: ajm80@le.ac.uk (A.J. Mitchell), Deepak.Shukla@leicspart.nhs.uk (D. Shukla), haa15@student.le.ac.uk (H.A. Ajumal), b.stubbs@greenwich.ac.uk (B. Stubbs), tayyeb.tahir@wales.nhs.uk (T.A. Tahir).

http://dx.doi.org/10.1016/j.genhosppsych.2014.09.003
0163-8343/

impairment, using a brief cognitive test. Recently, several reviews of screening instruments to detect delirium have been published. A recent review of 11 instruments in 25 studies highlighted potentially favourable accuracy for Global Attentiveness Rating, Memorial Delirium Assessment Scale (MDAS), Delirium Rating Scale-Revised-98 (DRS-R-98), Clinical Assessment of Confusion, Delirium Observation Screening Scale and Nursing Delirium Screening Scale. The CAM was the most thoroughly investigated but notable the Mini-Mental State Examination (MMSE) was partially omitted from this review. Although the MMSE is designed to assess global cognitive impairment, and it currently under licence (pay per use), it may prove potentially useful to detect delirium and is already commonly used in a range of clinical settings. Many studies have looked at the diagnostic value of the MMSE in cognitive disorders but mostly in context of dementia, not delirium [29]. The MMSE has been used extensively in different clinical and non-clinical settings [30]. It is a brief test consisting of 20 individual tests covering 11 domains including orientation, registration, attention and calculation, recall, naming, repetition, comprehension, writing and construction. Many validation studies exist, but most are underpowered and many lack an adequate criterion standard and hence can give a misleading impression of accuracy [31]. The MMSE is a valid test of cognitive functions and is reliable for 24-h and 28-day assessment for single or multiple raters (Pearson Coefficient 0.877). Internal consistency appears to be moderate with Cronbach alpha scores ranging from 0.6 to 0.89 [32,33]. However, its utility in detecting delirium is uncertain although a large study regarding the MMSE and delirium found a mean MMSE score of 12.6 in those with delirium and 25.7 in those without [34]. Despite the fact the MMSE is widely used to screen for cognitive impairment, its value in diagnosing delirium is uncertain and requires investigation. Thus, the aim of this paper was to systematically review and analyze the evidence concerning the accuracy of the MMSE as a diagnostic (case-finding) and screening test for the presence of delirium in adults.

2. Methods

This systematic review and meta-analysis was conducted in accordance with the PRISMA guidelines following a predetermined protocol [35].

2.1. Data sources and search

Two independent reviewers searched Medline, PsycINFO and Embase abstract databases from inception to March 2014. This was supplemented by searches of five full text collections (Science Direct, Ingenta Select, Ovid Full Text, Blackwell Online and Wiley Interscience) and the abstract database Web of Knowledge (4.0, ISI). In accordance with the protocol, where necessary, authors were contacted directly for primary data. The following search terms were used: "(Screen* or test or instrument or measure or tool or diagnos*) and (Mini mental state examination or MMSE or Folstein) and (delirium or cogniti*) and ("sensitivity and specificity or accuracy or cutoff or receiver operator or ROC or Youden").

2.2. Eligibility criteria

We included studies that examined the diagnostic validity of the MMSE to detect delirium against the reference standard according to the *DSM-IV* of the American Psychiatric Association (for example *DSM-IV*) or *ICD* (for example *ICD-10*) of the World Health Organization criteria. Studies that did not clearly state the comparator to be *DSM* or *ICD* diagnosis for delirium, or that did not provide sufficient data to be extracted and included in the meta-analysis were excluded. We did not place a language restriction upon eligible studies.

2.3. Methodological quality appraisal

2.3.1. Quality assessment and risk of bias assessment

Two authors (B.S., A.J.M.) conducted the risk of bias assessment using a four point quality rating and a five point bias risk was applied to each study as used in a recent similar study [36]. The quality rating score was based on study sample size, study design, study attrition and method of dealing with possible confounders with the following scale: 1=low quality 2=low–medium quality 3=medium–high quality 4=high quality. The bias rating score evaluated possible bias in assessments of results as influenced by consideration of setting, sampling method, interview method and sampling method. Bias was rated with the following score: 0=no appreciable bias risk 1=low bias risk 2=low–medium bias risk 3=medium–high bias risk 4=high bias risk. A composite score of >3 on study quality+<3 on bias score generated seven robust studies.

2.3.1.1. Analysis. An unweighted pooled meta-analysis of suitable studies was conducted, to give overall test accuracy, sensitivity, specificity, combined Youden score, positive and negative predictive values (PPV, NPV), positive and negative likelihood ratios (LR+, LR−) and positive and negative clinical utility index (CUI+, CUI−). Further details are available here www.clinicalutility.co.uk. The CUI is a proxy for the applied value of a test with a qualitative as well as quantitative interpretation [37–39]. Clinical utility may be more important to clinicians than validity [40]. Clinical utility estimates the clinical value of a diagnostic test taking into account both the accuracy of the test and its occurrence. The positive utility index (for rule-in or case-finding accuracy) is a product of sensitivity and positive predictive value and the negative utility index (for rule-out or screening accuracy) is a product of Sp × NPV. The interpretation of the CUI is 0.93–1.00 near perfect value; 0.81–0.92 excellent; 0.64–0.80 good; 0.49–0.63 adequate; 0.36–0.48 poor; and <0.36 very poor. Publication bias was tested by Harbord method [41]. Comparative accuracy was tested by conducting a relative risk comparison of pooled sensitivity and specificity and by comparing overall accuracy at equivalent prevalence rates of 25% and 50%. In order to assess the influence of the quality of studies on the observed results, we conducted subgroup analysis using most robust (high quality) studies only where the delirium was determined by robust interview methods. As the included studies used a variety of cutoff thresholds we also conducted a subgroup analysis to establish the observed results differed in studies using a predefined cutoff of <24 on the MMSE.

3. Results

3.1. Part 1 systematic review

We identified 13 valid studies of the MMSE for the detection of delirium in medical settings involving a total of 2017 patients of whom 29.4% had delirium [42–54]. Studies were published between 1982 and 2011. The smallest study involved 18 cases of delirium [43] whilst the largest had 142 cases [50]. The prevalence of delirium ranged from 11.7% to 58.3%. All of the studies had acceptable methodological quality and none of the studies were deemed to be at high risk of bias. A full summary of the included articles details including methodological quality and risk of bias is shown in Table 1.

Anthony et al. studied 97 patients, who were admitted consecutively to a General medical ward at John Hopkins Hospital in 1979, aged above 20 years [42]. The sample was predominantly female, black, with little education and from a socioeconomically deprived background. *DSM* criteria were used as the gold standard, applied by a trained psychiatrist. The MMSE was administered within 24 h of admission to the ward. At a cutoff of <24, the MMSE had a sensitivity of 87.0% and a specificity of 82.4% in diagnosing delirium or dementia. This study was atypical in that delirium or dementia. Was the gold standard. The authors also calculated sensitivity and specificity at various cutoff points on the MMSE. Trzepacz et al. examined 108 liver transplantation candidates

Table 1
Methodological Summary of studies.

Study author (year)	Diagnosis of delirium	Comparisons	Sample, age, gender	Total study size	Quality rating score	Bias rating score	Mean age	Gender	Setting
Anthony et al. (1982)	Delirium OR Dementia *DSM-III* by Psychiatrist	none	97 patients (37 male) 46 over 60 years	97	3	3	60 years	37 male, 60 female	Hospital
Dyer et al. (1994)	CAM	DSI	97% male, mean age 70.1 years	60	2	3	70.1 years	97% male,	Hospital
Fayers et al. (2005)*	*ICD-10* Delirium	Brief 4 items MMSE	80 years, 58% female	305	4	1	80 years	42% male	Hospital
Franco et al. (2010)	Two step: CAM-S then DRS-R98	None	60–99 years	291	2	2	74.4 years	186 females and 105 males	Hospital
Grassi et al. (2001)*	*DSM-III-R* Delirium	CAM), the DRS, the MDAS	55 males 67.7 years	105	3	2	67.7 years	55 males	Hospital
Hart et al. (1996)	*DSM-III-R* Delirium by Psychiatrists	Cognitive test for delirium	NR	103	2	2	62.5years	42.5% female	Hospital (controls included outpatients)
Khurana et al. (2002)*	*ICD-10-DCR* Delirium	CAM DSI	65–89 year	100	3	2	65–89 years	64% males 36% females	Hospital
O'Keeffe et al. (2005)*	CAM	none	79 years	160	3	2	79 years	NR	Hospital
Ringdal et al. (2011)*	CAM	none	84 years, 76% female, 54% with MMSE<24	364	3	2	Over 65 years	76% female	Hospital patients with hip fracture
Rockwood et al. (1996)*	*DSM-III-R* Delirium by Psychiatrists	DRS	79 years	104	3	2	79 years	NR	Hospital
Rolfson et al. (1999)	*DSM-III-R* Delirium by Psychiatrists	CAM, CDT	80% male, mean age 71 years	71	3	3	71 years	80% male,	Hospital inpatients undergoing cardiac surgery
Sharma et al. (2011)*	*DSM-IV-TR* by psychiatrists	none	>18 years	149	3	2	44 years	87 males 62 females	Hospital
Trzepacz et al. (1988)	*DSM-III* Delirium	Trails A, B; EEG	108 consecutive liver transplant candidates	108	2	3	41 years	35% male	Hospital

CDT, clock drawing test; Quality rating scores 1=low quality 2=low-medium quality 3=medium–high quality 4=high quality. Bias rating scores 0=no appreciable bias risk 1=low bias risk 2=low to medium bias risk 3=medium to high bias risk 4=high bias risk. *High quality studies used in subgroup analysis.

with end-stage liver disease from gastroenterology service at Presbyterian-University Hospital, Pittsburgh, PA, USA [43]. They were all English speaking, with 11 or more years of education. Subjects were between 17 and 62 years of age. Psychiatric diagnoses were made using *DSM-III* criteria. A MMSE score of less than 24 had a sensitivity of 55.6% and a specificity of 82.2% in detecting delirium. Further PPV was 38.5% and NPV 90.2%. Comparatively the trail making test B had 66.7% sensitivity and 95.6% specificity.

Dyer et al. conducted a prospective study on the diagnosis post-operative delirium comparing the 107 item Delirium Symptom Interview (DSI) and the MMSE to the CAM [44]. The CAM developed in 1990 was used as the gold standard [55]. The subjects were 60 consecutive patients who underwent general, orthopaedic or urologic surgery. DSI, MMSE and CAM were administered pre-operatively and post-operatively (days 1–7); 12% of subjects had a pre-operative diagnosis of dementia or depression and 58% developed delirium. The MMSE had 77.1% sensitivity, 56.0% specificity 71.1% PPV and 63.6% NPV. Comparatively the DSI had 92% sensitivity and 64% specificity. Hart et al. set out to validate two forms of Cognitive Test for Delirium (CTD) in medical ICU patients [45]. They also compared the performance of CTD to the MMSE and investigated whether these tests can be used to differentiate delirium from other mental illnesses such as dementia in out-patient setting, depression and schizophrenia in in-patient psychiatry service in the Medical College of Virginia. There were less than 30 patients in each group. The *DSM-III-R* was used as the gold standard for diagnosis. An receiver operating characteristic (ROC) analysis indicated that for both CTD and the MMSE, an optimal cutoff score to discriminate delirium from other disorders was <19. At this score, the MMSE had a sensitivity of 100% and specificity of 93.8%. Rockwood et al. compared the MMSE with the DRS in a cross-sectional study in 1992 in Ontario, Canada [46]; 104 inpatients from geriatric medicine and geriatric psychiatry wards of two tertiary referral hospitals participated in the study. *DSM-III-R* was used as the gold standard for diagnosis of delirium. The subjects were administered the Delirium Rating Scale (DRS), MMSE, Barthel Index and Blessed Dementia Scale.. At a cutoff of <24, MMSE showed a sensitivity of 88.5% and specificity of 52.6%. Comparatively the DRS had 82% sensitivity and 94% specificity when 10 is set as the cut-point. Rolfson et al. studied a cohort of 71 consecutive patients undergoing elective coronary artery bypass graft at a tertiary care hospital in Northern Alberta, Canada [47]. The primary objective was to assess the validity of the CAM to detect delirium but the authors also included data on the MMSE. Patients were followed daily until the fourth postoperative day. Delirium was diagnosed using the *DSM-III-R* criteria. The ROC curves were constructed for the CAM and MMSE. At a cutoff of <24, the MMSE had a sensitivity of 34.8% and a specificity of 81.3%. Comparatively the CAM had 70% sensitivity and 100% specificity.

Seven studies have been published since 2000. In Grassi et al. conducted a study which was carried out in 6 centres in Italy, including 4 medical oncology wards and 2 palliative care units [48]; 105 consecutive cancer inpatients presenting with a mental status change that were referred to the consultation-liaison psychiatric service or palliative care unit were evaluated. The objective was to validate the Italian versions of the DRS and the MDAS. The criterion reference was *DSM-III-R* criteria for delirium. Using a cutoff of <24, the MMSE showed a sensitivity of 95.5% and a specificity of 38.5%, PPV of 72.4% and NPV of 83.3%. Comparatively the MDAS had 68% sensitivity and 94% specificity for a cutoff of 13 for delirium. The DRS had 95% sensitivity and 61% specificity for DRS cutoff 10 and 81% sensitivity and 76% specificity for DRS cutoff 12. Khurana et al. studied 100 hospitalised geriatric general medical patients, aged 65 and above, who were admitted under the Department of Internal Medicine, Kasturba Hospital, Manipal, Karnataka [49]. The patients were assessed within 24 h or admission 61 and then on every fourth day thereafter. The assessment was carried out using the MMSE, CAM, DSI against the *ICD-10* criteria for delirium. At a cutoff score of <24, the MMSE showed 100% sensitivity and 45.2% specificity. In comparison, the CAM had 100% sensitivity and 100% specificity. Also, the DSI

had 100% sensitivity and 90% specificity. Fayers et al. recruited 150 patients, diagnosed with delirium, between the ages of 70 and 90, from a general medical unit for somatic diseases in a University Hospital, Norway [50]. Trained nurses administered the MMSE. The authors also studied a separate group of 163 consecutive patients who were admitted at the same hospital and of similar age but with no diagnosis of delirium or other cognitive impairment. At a cutoff of <24, the MMSE had a sensitivity of 89.4% and specificity of 100% in this sample with 100% PPV and 91.6% NPV. O'Keeffe et al. looked at the value of serial MMSEs in diagnosing and monitoring delirium in Ireland [51]. In this prospective study 165 consecutive patients aged 65 and older who were admitted from the accident and emergency department to an acute geriatric medicine service were recruited. Two different examiners blind to each other, administered the MMSE to the subjects on Days 1 and 6. On the same hospital days, an experienced consultant geriatrician examined the subjects and diagnosed delirium using the CAM diagnostic algorithm. A fall of two or more points on the MMSE was the best determinant for detecting the development of delirium. This change score yielded a sensitivity of 91.7% and specificity of 90.0%. A rise of 3 or more points was the best determinant for detecting resolution of delirium with a sensitivity of 77% and specificity of 75%.

Since 2010 a further three studies have been published. Sharma et al. studied 149 consecutive patients who had been referred to the psychiatric department for behavioural abnormalities from various other departments in Shree Krishna hospital, Karamsad, Gujarat, over 1 year. The aim of the study was to assess the optimal cutoff for MMSE to detect delirium, using *DSM-IV-TR* as the gold standard. Diagnoses were made by a psychiatrist blind to the MMSE score. Using the ROC analysis, the optimal cutoff score of the MMSE was 24.5, giving a sensitivity of 97% and a specificity of 69% but at <24 sensitivity was 80.6% and specificity 71.8%. Franco et al. examined 291 patients aged over 60 who were hospitalised in three internal medicine wards in Clinica Universitaria Bolivariana, Columbia. The patients were assessed within 24 h of admission using CAM-Spanish (CAM-S) then DRS-R-98 (two-step procedure). Those who scored 'positive' were excluded and 'negative' were evaluated using the Colombian version of the MMSE, to measure global cognitive status. Using the cutoff score for the MMSE <24.5, a sensitivity of 79.4% and a specificity of 52.1% was found but at <24 sensitivity was 70.6% and specificity 62.6%. The positive and negative predicted values were 20.0% and 94.2%, respectively. However, a limitation of this paper is that the criterion reference was two-step procedure and an important consideration is that the authors appear to measure the incidence and not the prevalence of delirium. Ringdal et al. examined the value of the MMSE for detecting delirium in 364 over 65-year-old Norwegian-speaking subjects [52]. This was the largest study in the literature. Some MMSE questions were modified into Norwegian. The CAM was used as the gold standard with <24 as the cutoff point. The MMSE had a sensitivity of 88.2% and a specificity of 54.2% in detecting delirium (PPV was 33.7% and NPV 94.5%). A summary of the included studies is presented in Table 1.

3.2. Part 2: meta-analytic results

We located 13 studies, all in hospital settings. The total sample size was 2017 of whom 564 giving a pooled prevalence of delirium of 27.9% (25.9–29.9%); corrected to 29.4% (95% CI=21.5–37.9%) on meta-analysis. However, this was 31.6% (95% CI=21.6–42.6%) in robust (high quality) studies using interview based criteria. The statistical summary of the individual results from each study are presented in Table 2.

3.2.1. Sensitivity and specificity

3.2.1.1. Main analysis. Examining sensitivity and specificity, we found a diagnostic validity meta-analysis gave an overall sensitivity estimate of 84.1% (95% CI=75.8–90.9%). It was no different in studies using a predefined cutoff of <24. Regarding specificity meta-analysis gave an

Table 2
Statistical Summary of studies.

Study author (year)	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Likelihood ratio+ve (95% CI)	Likelihood ratio+ve (95% CI)	CUI+ve	CUI−ve
Anthony et al. (1982)	23v24	0.870 (0.732–1.00)	0.824 (0.738–0.911)	0.606 (0.439–0.773)	0.953 (0.901–1.00)	4.95 (2.95–8.31)	0.16 (0.05–0.46)	0.527 (0.503–0.551) "fair"	0.786 (0.780–0.791) "good"
Dyer et al. (1994)	NR	0.771 (0.632–0.911)	0.56 (0.365–0.755)	0.711 (0.566–0.855)	0.636 (0.435–0.837)	1.75 (1.09–2.83)	0.41 (0.20–0.82)	0.548 (0.530–0.566) "fair"	0.356 (0.324–0.388) "v poor"
Fayers et al. (2005)	23v24	0.894 (0.844–0.945)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.916 (0.875–0.957)	NA	0.11 (0.07–0.17)	0.894 (0.893–0.896) "good"	0.916 (0.915–0.917) "excellent"
Franco et al. (2010)	<24	0.706 (0.553–0.859	0.626 (0.567–0.686)	0.20 (0.128–0.272)	0.942 (0.906–0.877)	1.89 (1.44–2.47)	0.47 (0.28–0.80)	0.141 (0.133–0.149) "v poor"	0.590 (0.587–0.593) "fair"
Grassi et al. (2001)	23v24	0.955 (0.904–0.100)	0.385 (0.235–0.537)	0.724 (0.630–0.818)	0.833 (0.661–1.00)	1.55 (1.20–2.00)	0.12 (0.04–0.38)	0.691 (0.685–0.698) "good"	0.321 (0.294–0.347) "v poor"
Hart et al. (1996)	18v19	1.00 (1.00–1.00)	0.938 (0.886–0.991)	0.815 (0.668–0.961)	1.00 (1.00–1.00)	16.2 (6.93–37.9)	NA	0.815 (0.801–0.828) "excellent"	0.938 (0.937–0.940) "excellent"
Khurana et al. (2002)	<24	1.00 (1.00–1.00)	0.452 (0.338–0.566)	0.403 (0.286–0.520)	1.00 (1.00–1.00)	1.83 (1.48–2.25)	NA	0.403 (0.387–0.419) "poor"	0.452 (0.438–0.466) "poor"
O'Keeffe et al. (2005)	Fall of 2 points	0.917 (0.826–1.00)	0.900 (0.847–0.953)	0.727 (0.597–0.856)	0.974 (0.945–1.00)	9.17 (5.36–15.7)	0.09 (0.03–0.27)	0.666 (0.653–0.679) "good"	0.876 (0.875–0.878) "excellent"
Ringdal et al. (2011)	23v24	0.882 (0.809–0.954)	0.542 (0.484–0.599)	0.337 (0.271–0.402)	0.945 (0.911–0.980)	1.92 (1.66–2.24)	0.22 (0.12–0.41)	0.297 (0.291–0.302) "v poor"	0.512 (0.509–0.515) "fair"
Rockwood et al. (1996)	23v24	0.885 (0.762–1.00)	0.526 (0.415–0.636)	0.383 (0.260–0.506)	0.932 (0.857–1.00)	1.86 (1.42–2.45)	0.22 (0.07–0.65)	0.339 (0.322–0.357) "v poor"	0.490 (0.478–0.501) "poor"
Rolfson et al. (1999)	23v24	0.348 (0.153–0.542)	0.813 (0.702–0.923)	0.471 (0.233–0.708)	0.722 (0.603–0.842)	1.86 (0.82–4.18)	0.80 (0.58–1.11)	0.164 (0.131–0.197) "v poor"	0.587 (0.575–0.599) "fair"
Sharma et al. (2011)	<24.5	0.806 (0.676–0.935)	0.717 (0.635–0.800)	0.475 (0.350–0.601)	0.921 (0.864–0.977)	2.84 (2.04–3.97)	0.27 (0.14–0.53)	0.383 (0.368–0.398) "poor"	0.660 (0.654–0.665) "good"
Trzepacz et al. (1988)	NR	0.556 (0.326–0.785)	0.822 (0.743–0.901)	0.385 (0.198–0.572)	0.902 (0.838–0.967)	3.13 (1.70–5.73)	0.54 (0.32–0.91)	0.214 (0.181–0.246) "v poor"	0.742 (0.737–0.747) "good"

Footnote: values calculated from raw data using www.clinicalutility.co.uk calculator. The CUI is a proxy for the applied value of a test with a qualitative as well as quantitative interpretation: CUI+ve for case finding and CUI−ve for screening. +ve, positive; −ve, negative.

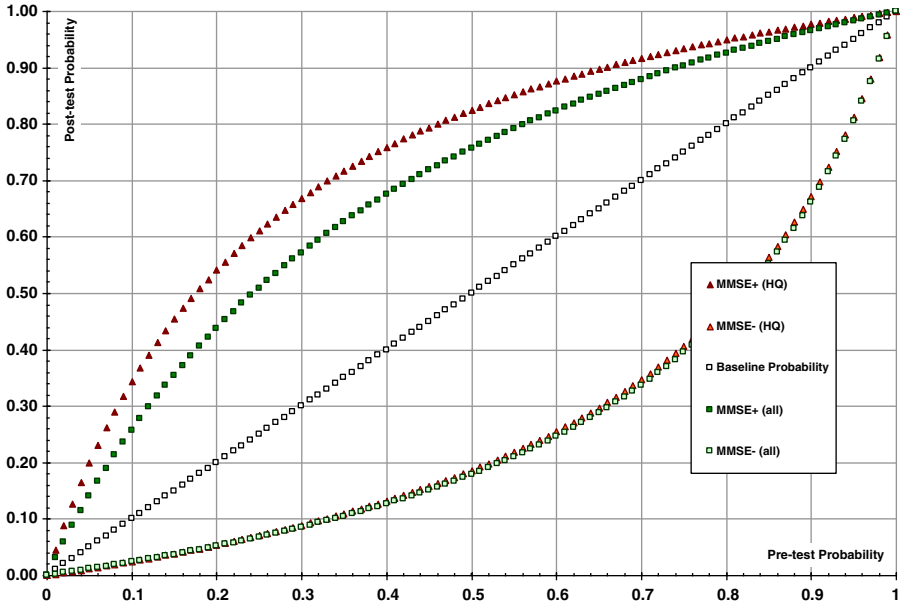

Fig. 1. Bayesian Conditional Probability Plot of MMSE Accuracy. Footnote: The Bayesian Conditional Probability Plot shows the positive and negative predictive values for every possible prevalence value. HQ=high quality studies; All=all studies.

overall sensitivity estimate of 73.0% (95% CI=59.6–84.5%) (Fig. 1). It was 68.4% (95% CI=50.9–83.5%) in studies using a predefined cutoff of <24.

3.2.1.2. Sub-analysis (high-quality studies). Sub-analysis including only robust (high quality) studies using interview-based criteria for delirium was conducted. Seven such studies had a meta-analytic sensitivity of 81.1% (95% CI=65.9–92.6%) and a specificity of 82.8% (95% CI=64.4–95.4%).

3.2.2. Positive and negative predictive value

3.2.2.1. Main analysis. Using the main analysis for sensitivity and specificity, and assuming delirium was present in 10% of patients, then the PPV and NPV would be 25.7% (17.3–39.5%) and 97.6% (95.7–98.8%), respectively, with a positive likelihood ratio of 3.11 (1.88–5.86) and negative likelihood ratio of 0.22 (0.41–0.11) (Fig. 1). Assuming delirium was present in 25% of high-risk patients, then the PPV and NPV would be 50.9% (48.3–66.2%) and 93.2% (90.0–96.5%), respectively, with the same likelihood ratios.

3.2.2.2. Sub-analysis (high quality studies). Using the high quality sub-analysis confined to 7 robust (high quality) studies then sensitivity and specificity, and assuming delirium was present in 10% of patients, then the PPV and NPV would be 34.4% (17.1–69.1%) and 97.5% (94.4–99.1%), respectively with a positive likelihood ratio of 4.72 (1.85–20.1) and negative likelihood ratio of 0.23 (0.08–0.53). Assuming delirium was present in 25% of high-risk patients, then the PPV and NPV would be 61.1% (38.2–87.0%) and 92.9% (85.0–97.5%), respectively, with the same likelihood ratios.

3.2.3. Clinical utility

3.2.3.1. Main analysis. Assuming delirium was present in 10% of patients, then the positive clinical utility would be 0.216 (qualitatively poor) and the negative clinical utility would be 0.713 (qualitatively good). Assuming delirium was present in 25% of patients, then the positive clinical utility would be 0.428 (qualitatively poor) and the negative clinical utility would be 0.681 (qualitatively good).

If the MMSE was used in a modest risk setting (prevalence 10%) then it would likely facilitate in the correct detection of 8 delirious patients, missing 2, and correctly ruling out 66 non-delirious patients falsely suggesting 24. If the MMSE was used in a high risk setting (prevalence 25%) then it would likely facilitate in the correct detection of 21 delirious patients, missing 4, and correctly ruling out 55 non-delirious patients but with 20 false positives.

3.2.3.2. Sub-analysis (high quality studies). Using the robust (high quality) sub-analysis confined to 7 studies, and assuming delirium was present in 10% of patients, then the positive clinical utility would be 0.279 (qualitatively poor) and the negative clinical utility would be 0.808 (qualitatively good). Assuming delirium was present in 25% of patients, then the positive clinical utility would be 0.496 (qualitatively poor) and the negative clinical utility would be 0.769 (qualitatively good).

If the MMSE was used in a modest risk setting (prevalence 10%) then it would likely facilitate in the correct detection of 8 delirious patients, missing 2, and correctly ruling out 75 non-delirious patients falsely suggesting 15. If the MMSE was used in a high risk setting (prevalence 25%) then it would likely facilitate in the correct detection of 20 delirious patients, missing 5, and correctly ruling out 62 non-delirious patients but with 13 false positives.

4. Discussion

We located 13 valid diagnostic studies of the MMSE involving 2017 individuals tested for delirium. An inclusive approach (including all qualifying studies) led to a sensitivity and specificity estimate for the MMSE of 84.1% (95% CI=75.8–90.9%) and 73.0% (95% CI=59.6–84.5%). However, only 7 studies were of deemed to be highest quality and used interview based criteria for delirium. In addition, one study used a two-step procedure of the CAM in order to find incident delirium cases during hospitalization, that were then quantified with

the DRS-R98 [54]. and this may have influenced the pooled meta-analysis results. Another included patients with delirium and/or dementia (although the remainder excluded dementia) [42].

Therefore, excluding these and other lower quality studies led to a best estimate of sensitivity and specificity refined to 81.1% (95% CI= 65.9% to 92.6%) and 82.8% (95% CI=64.4% to 95.4%), respectively. Taking this high quality study estimate, in both medium risk and high-risk settings the clinical utility of the MMSE was qualitatively poor for case-finding. However, in both medium risk and high-risk settings the clinical utility of the MMSE was qualitatively good for screening. For example when the prevalence of delirium was 10% the MMSE achieved 97.5% NPV. If the MMSE was used in a modest risk setting (prevalence 10%) as an initial screening tool then it would likely facilitate in the correct detection of 8 out of 10 delirious patients, missing 2. If the MMSE was used in a high risk setting (prevalence 25%) then it would likely facilitate in the correct detection of 20 delirious patients, missing about 5 cases.

The MMSE is the most widely used test of cognitive impairment but its role in assessing delirium has never been adequately clarified. The MMSE was designed to assess broad cognitive impairment whereas other tools have been specifically designed for screening (e.g., CAM and DSI) or ascertaining the severity of delirium (e.g. DI, MDAS and DRS-R-98) [56]. Nevertheless, the MMSE is the most popular tool in clinical practice and the one most often used by clinicians to screen for delirium. Clinicians may, however, assume the MMSE is both an adequate screening and case-finding tool. Few studies have offered a head-to-head comparison of focussed delirium screens against the MMSE. Assuming replication from at least one independent centre is necessary in order to make a judgement about such a comparison we could only find a comparison with the DRS (two studies)[46,48] and the confusion assessment method (CAM) (2 studies) [49,51].

Against the DRS the MMSE had inferior sensitivity and inferior specificity in both studies (DRS SE: 90% Sp 82% vs MMSE SE: 88.5% Sp 52.6%)[48] (DRS SE: 80% Sp 76% vs MMSE SE 66% Sp 38.5%)[50]. Against the CAM the MMSE appeared to have equal or inferior sensitivity and inferior specificity in both studies (CAM SE: 100% Sp 100% vs MMSE SE: 100% SP 45.2%)[51] (CAM SE: 70% Sp 100% vs MMSE SE: 35% SP 81.3%) [49]. Although the sample size is low we can state that the MMSE is probably less accurate that its competitors (CAM and DRS) when diagnosing delirium. However, it is important to note that the differential effect upon missed negatives is very small using either CAM or DRS vs. MMSE. In other words for screening purposes the MMSE is probably acceptable but for case-finding, competitor tools are preferred. Future studies may clarify if specific domains of the MMSE can be used in isolation, for example orientation or spelling. In addition it is likely that accuracy can improved by serial testing [51].

The under-recognition of delirium can be associated with factors such as the fluctuating nature of delirium, its overlap with dementia and depression, the scarcity of formal cognitive assessment in general hospitals by routine, under-appreciation of its clinical consequences, and failure to consider the diagnostic importance. Non-detection of delirium has been also associated with the high prevalence of the hypoactive form of delirium. Four independent risk factors for the under-recognition of delirium by nurses have been identified: hypoactive delirium, advanced age, visual impairment and dementia [57]. It should also be remembered that subtypes of delirium, for example, subsyndromal deliria may be particularly difficult to detect for any screening tool.

The MMSE has some limitations that may have influenced the findings. It has an over-reliance on verbal assessment at the expense of non-dominant hemisphere skills and executive functioning, insensitivity to frontal executive dysfunction and visuospatial deficits, superficial assessment of memory and language and inability to provide qualitative information of cognitive profile [58]. Although, the MMSE has high sensitivity and specificity with a good positive predictive validity and negative predictive validity it is modestly effective in ruling out dementia [29,39]. Scales for cognitive assessment can be influenced by factors including age, educational status, affective changes and fluctuations in cognitive picture, compromising their accuracy. Unfortunately only two studies (see Table 1) examined here looked at younger adults therefore the effect of age remains unaddressed. High inter-observer agreement for the MMSE, the Delirium Symptom Inventory and the CAM suggest that they may different but overlapping assessment of delirium.

Although the brief bedside tools for assessment of cognitive functions have a role, it is important to keep in mind that they should not be used to replace a full clinical appraisal to reach a diagnosis of delirium. Hence, the MMSE can be used as an aid to ascertain the cognitive status to monitor any improvement or deterioration to facilitate the process of making and reviewing a clinical diagnosis and management for early intervention for resolution of delirium.

We conclude that the MMSE should not be used as a case-finding confirmatory test of delirium as it would be accurate in one in four to one in two cases, but it could be used as an initial screen to rule out those who are unlikely to have delirium with approximately 93–97% accuracy.

References

[1] Fong Tamara G, Tulebaev Samir R, Sharon K. Inouye delirium in elderly adults: diagnosis, prevention and treatment. Nat Rev Neurol 2009;5(4):210–20. http://dx.doi.org/10.1038/nrneurol.2009.24.
[2] WHO. The *ICD-10* classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines; 1992.
[3] American Psychiatric Association. DSM IV; 1994.
[4] Thomas RI, Cameron DJ, Fahs MC. A prospective study of delirium and prolonged hospital stay. Arch Gen Psychiatry 1988;45:937–40.
[5] Cole M, Primeau F. Prognosis of delirium in elderly hospital patients. Can Med Assoc J 1993;149:41–6.
[6] Rockwood K, Cosway S, Carver D, Jarret P, Stadnyk K, Fisk J. The risk of dementia and death after delirium. Age Ageing 1999;28:551–6.
[7] O'Keeffe S, Lavan J. The prognostic significance of delirium in older hospital patients. J Am Geriatr Soc 1997;45(2):174–8.
[8] Andrew MK, Freter SH, Rockwood K. Incomplete functional recovery after delirium in elderly people: a prospective cohort study. BMC Geriatr 2005;5:5.
[9] Pisani Margaret A, Yeon Joyce Kong So, Kasl Stanislav V, Murphy Terrence E, Araujo Katy LB, Van Ness Peter H. Days of delirium Are associated with 1-year mortality in an older intensive care unit population. Am J Respir Crit Care Med 2009;180(11):1092–7.
[10] McCusker J, Cole M, Abrahamowicz M, Primeau F, Belzile E. Delirium predicts 12 month mortality. Arch Intern Med 2002;162:457–63.
[11] Andrew MK, Freter SH, Rockwood K. Prevalence and outcomes of delirium in community and non-acute care settings in people without dementia: a report from the Canadian Study of Health and Aging. BMC Med 2006;4:4–15.
[12] Vidal EI, Villas Boas PJ, Valle AP, Cerqueira AT, Fukushima FB. Delirium in older adults. BMJ 2013;346:f2031. http://dx.doi.org/10.1136/bmj.f2031.
[13] Cole M. Delirium in elderly patients. Focus 2005;3:320–32.
[14] Barron EA, Holmes J. Delirium within the emergency care setting, occurrence and detection: a systematic review. Emerg Med J 2013;30(4):263–8. http://dx.doi.org/10.1136/emermed-2011-200586 [Epub 2012 Jul 25].
[15] Siddiqi N, House AO, Holmes JD. Occurrence and outcome of delirium in medical in-patients: a systematic literature review. Age Ageing 2006;35:350–64.
[16] Folstein MF, Bassett SS, Romanoski AJ, Nestadt G. The epidemiology of delirium in the community: the Eastern Baltimore Mental Health Survey. Int Psychogeriatr 1991;3(2):169–76.
[17] Sandberg O, Gustafson Y, Brännström B, Bucht G. Prevalence of dementia, delirium and psychiatric symptoms in various care settings for the elderly. Scand J Soc Med 1998;26(2):56–62.
[18] Rockwood K. The occurrence and duration of symptoms in elderly patients with delirium. J Gerontol 1993;48(4):162–6.
[19] Rahkonen T, Eloniemi-Sulkava U, Halonen P, Verkkoniemi A, Niinistö L, Notkola IL, et al. Delirium in the non-demented oldest old in the general population: risk factors and prognosis. Int J Geriatr Psychiatry 2001;16(4):415–21.
[20] Saxena S, Lawley D. Delirium in the elderly: a clinical review. Postgrad Med J 2009;85:405–13. http://dx.doi.org/10.1136/pgmj.2008.072025.
[21] Inouye SK. Delirium in hospitalized older patients. Clin Geriatr Med 1998;14(4):745–64.
[22] Lewis LM, Miller DK, Morley JE, Nork MJ, Lasater LC. Unrecognized delirium in ED geriatric patients. Am J Emerg Med 1995;13(2):142–5.
[23] Inouye SK. Delirium in hospitalized older patients: recognition and risk factors. J Geriatr Psychiatry Neurol 1998;11(3):118–25.
[24] Han JH, Zimmerman EE, Cutler N, Schnelle J, Morandi A, Dittus RS, et al. Delirium in older emergency department patients: recognition, risk factors, and psychomotor subtypes. Acad Emerg Med 2009;16:193–200.
[25] Mistarz R, Eliott S, Whitfield A, Ernest D. Bedside nurse-patient interactions do not reliably detect delirium: an observational study. Aust Crit Care 2011;24(2):126–32. http://dx.doi.org/10.1016/j.aucc.2011.01.002 [Epub 2011 Apr 21].

[26] Gustafson Y, Brannstrom B, Norberg A, Bucht G, Winblad B. Underdiagnosis and poor documentation of acute confusional states in elderly hip fracture patients. J Am Geriatr Soc 1991;39(8):760–5.
[27] Inouye SK, Foreman MD, Mion LC, Katz KH, Cooney Jr LM. Nurses' recognition of delirium and its symptoms. Arch Intern Med 2001;161:2467–73.
[28] www.nice.org.uk/nicemedia/live/13060/49909/49909.pdf.
[29] Mitchell AJ. The Mini-Mental State Examination (MMSE): an update on its diagnostic validity for cognitive disorders. Cognitive screening instruments a practical approach, J. Larner; 2013. p. 15–46 [ISBN: 978-1-4471-2451-1 (Print) 978-1-4471-2452-8 (Online)].
[30] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12(3):189–98.
[31] Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. J Am Geriatr Soc 1992;40:922–35.
[32] Toglia J, Fitzgerald KA, O'Dell MW, Mastrogiovanni AR, Lin CD. The Mini-Mental State Examination and Montreal Cognitive Assessment in persons with mild subacute stroke: relationship to functional outcome. Arch Phys Med Rehabil 2011;92(5):792–8.
[33] Mystakidou Kyriaki, Tsilika Eleni, Parpa Efi, Galanos Antonis, Vlahos Lambros. Brief cognitive assessment of cancer patients: evaluation of the Mini-Mental State Examination (MMSE) psychometric properties. Psycho-Oncology 2007;16:352–7.
[34] Kiely DK, Bergmann MA, Murphy KM, Jones RN, Orav EJ, Marcantonio ER. Delirium among newly admitted postacute facility patients: prevalence, symptoms, and severity. J Gerontol A Biol Sci Med Sci 2003;58(5):M441–5.
[35] Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 2009;6(7):e1000097. http://dx.doi.org/10.1371/journal.pmed.1000097.
[36] Mitchell AJ, Chan M, Bhatti H, Halton M, Grassi L, Johansen C, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. Lancet Oncol 2011;12(2):160–74 [Epub 2011 Jan 19].
[37] Mitchell AJ. Sensitivity×PPV is a recognized test called the clinical utility index (CUI+). Eur J Epidemiol 2011;26(3):251–2 [author reply 252. Epub 2011 Mar 26].
[38] Mitchell AJ. The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: a meta-analysis. Int J Geriatr Psychiatry 2008;23(11):1191–202.
[39] Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. J Psychiatr Res 2009;43(4):411–31.
[40] Reeve JL, Lloyd-Williams M, Dowrick C. Revisiting depression in palliative care settings: the need to focus on clinical utility over validity. Palliat Med 2008;22(4):383–91.
[41] Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med 2006;25(20):3443–57.
[42] Anthony JC, LeResche L, Niaz U, von Korff MR, Folstein MF. Limits of the "Mini-Mental-State" as a screening test for dementia and delirium among hospital patients. Psychol Med 1982;12:397–408.
[43] Trzepacz PT, Brenner RP, Coffman G, van Thiel DH, et al. Delirium in liver transplant candidates: discriminant analysis of multiple test variables. Biol Psychiatry 1988;24:3–14.
[44] Dyer CB, Ashton CM, Teasdale MPH. Diagnosing postoperative delirium: comparing the delirium symptom interview and the Mini-Mental State to the Confusion Assessment Method. J Am Geriatr Soc 1994;42:SA16.
[45] Hart RP, Levenson JL, Sessler CN, Best AM, Schwartz SM, Rutherford LE. Validation of a cognitive test for delirium in medical ICU patients. Psychosomatics 1996;37:533–46.
[46] Rockwood K, Goodman J, Flynn M, Stolee P. Cross-validation of the Delirium Rating Scale (DRS). J Am Geriatr Soc 1996;44:839–42.
[47] Rolfson DB, McElhaney JE, Jhangri GS, Rockwood K. Validity of the Confusion Assessment Method in detecting postoperative delirium in the elderly. Int Psychogeriatr 1999;11:431–8.
[48] Grassi L, Caraceni A, Beltrami E, Borreani C, Zamorani M, Maltoni M, et al. Assessing delirium in cancer patients. The Italian versions of the Delirium Rating Scale and the Memorial Delirium Assessment Scale. J Pain Symptom Manag 2001;21(1):59–68.
[49] Khurana P, Sharma PSVN, Avasthi A. Prevalence of delirium in geriatric hospitalized general medical population. Indian J Psychiatry 2002;44:41–6.
[50] Fayers PM, Hjermstad MJ, Ranhoff AH, Kaasa S, Skogstad L, Klepstad P, et al. Which mini-mental state exam items can be used to screen for delirium and cognitive impairment? J Pain Symptom Manage 2005;30(1):41–50.
[51] O'Keeffe ST, Mulkerrin EC, Nayeem K, Varughese M, Pillay I. Use of serial Mini-Mental State examinations to diagnose and monitor delirium in elderly hospital patients. J Am Geriatr Soc 2005;53:867–70.
[52] Ringdal GI, Ringdal K, Juliebo V, et al. Using the Mini-Mental State examination to screen for delirium in elderly patients with Hip fracture. Dement Geriatr Cogn Disord 2011;32:394–400.
[53] Sharma H, Desai N, Ganjiwale J. Comparison of MMSE to DSMIV diagnostic criteria for the detection of delirium in medically ill patients with psychiatric referrals. J Ment Health Behav 2011;16(1):37–42.
[54] Franco JG, Valencia C, Bernal C, Ocampo MV, Trzepacz PT, Pablo Jd, et al. Relationship between cognitive status at admission and incident delirium in older medical inpatients. J Neuropsychiatry Clin Neurosci 2010;22(3):329–37. http://dx.doi.org/10.1176/appi.neuropsych.22.3.329.
[55] Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the Confusion Assessment Method. A new method for detection of delirium. Ann Intern Med 1990;113(12):941–8.
[56] Tahir TA. Delirium. Medicine 2012;40(12):658–61.
[57] Inouye SK, Foreman MD, Mion LC, Katz KH, Cooney Jr LM. Nurses' recognition of delirium and its symptoms: comparison of nurse and researcher ratings. Arch Intern Med 2001;161:2467–73. http://dx.doi.org/10.1001/archinte.161.20.2467.
[58] Bak TH, Mioshi E. A cognitive bedside assessment beyond the MMSE: the Addenbrooke's Cognitive Examination. Pract Neurol 2007;7:245–9.