



HAL
open science

Troubles cognitifs légers dans le cadre des maladies neurodégénératives : dépistage ou repérage ? Effet de spectre et biais spectral

Gilles Chopard

► **To cite this version:**

Gilles Chopard. Troubles cognitifs légers dans le cadre des maladies neurodégénératives : dépistage ou repérage ? Effet de spectre et biais spectral. Neurosciences. Université de Franche-Comté, 2012. Français. NNT : 2012BESA3007 . tel-01450692

HAL Id: tel-01450692

<https://theses.hal.science/tel-01450692>

Submitted on 31 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

Pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITÉ de FRANCHE-COMTE

Discipline : Sciences de la Vie et de la Santé

Spécialité : Neurosciences

Présentée et soutenue le 21 septembre 2012 par

Gilles CHOPARD

Né le 5 décembre 1969 à Besançon

**Troubles cognitifs légers dans le cadre des
maladies neurodégénératives :**

Dépistage ou repérage ?
Effet de spectre et biais spectral

Membres du jury :

Pr. Pierre Vandell : INSERM CIC-IT 808, EA 481, Université de Franche-Comté	<i>Directeur</i>
Pr. Lucien Rumbach : UMR CNRS 7237, Université de Strasbourg	<i>Co-Directeur</i>
Pr. Bernard Laurent : INSERM 879, Université de St Etienne	<i>Rapporteur</i>
Pr. Alain Vighetto : INSERM 1028, UMR CNRS 5292, Université de Lyon	<i>Rapporteur</i>

A ma mère, in memoriam,

A mon père, merci pour tout ce que tu m'as appris et donné,

A Lily,

A mes beaux-parents,

A Mi Ja, pour ton amour et ton soutien constant,

A Yann, notre petit trésor,

A ma sœur, à mes frères,

Mes belles-sœurs, mes beaux-frères, ma nièce et mes neveux,

A mes amis.

REMERCIEMENTS

Je tiens en premier lieu à remercier mes directeurs de thèse, les professeurs Pierre Vandel et Lucien Rumbach, pour leur disponibilité et leur soutien sans faille qu'ils m'ont témoigné.

Mes remerciements s'adressent également aux professeurs Alain Vighetto et Bernard Laurent, pour m'avoir fait l'honneur de juger ce travail.

J'aimerais adresser un remerciement particulier à Jean Galmiche qui m'a énormément aidé tout au long de ces années : je te remercie pour ta disponibilité et tes précieux conseils pour ce travail dont tu es à l'origine, mais aussi pour les réflexions et les nombreuses discussions que nous avons eues dans les domaines de la neuropsychologie appliquée. Merci de m'avoir accordé ta confiance depuis cette année 1997 et reçois, en ces quelques mots, la marque de ma sincère amitié.

Merci à Géraldine Sylvestre Vanholsbeeck pour son rôle majeur dans le développement du logiciel et de la base RAPID mais aussi pour sa relecture attentive de ce travail et ses encouragements constants.

Je tiens également à remercier le Professeur Frédéric Mauny ainsi que Marc Puyraveau, du centre de méthodologie clinique du CHU de Besançon, pour leur aide précieuse concernant certains aspects méthodologiques et statistiques de ce travail.

Merci à Alexandre Pitard et Grégory Tio, pour la réalisation des analyses statistiques dans le logiciel Stata.

Un grand merci à Julie Hidalgo et Nicolas Carvalho pour leur aide logistique, leur efficacité et leur gentillesse.

Un merci tout aussi grand à Madame Ginette Galmiche et Monsieur Gérard Baroin, pour la correction de mon français et à Monsieur Richard Medeiros, éditeur médical au CHU de Rouen, pour ses nombreuses relectures et corrections de mon anglais.

RESUME EN FRANÇAIS

Titre: Troubles cognitifs légers dans le cadre des maladies neurodégénératives : dépistage ou repérage ? Effet de spectre et biais spectral.

De nombreux tests de dépistage des troubles cognitifs légers ont été élaborés ces dernières années afin d'identifier des personnes plus à risque de développer une maladie neurodégénérative telle que la maladie d'Alzheimer. Plusieurs indices (sensibilité, spécificité, valeurs de prédiction et rapports de vraisemblance) permettent d'évaluer la performance globale et l'utilité d'un test en pratique courante. L'objectif de cette thèse était d'évaluer le degré de variation de la performance d'un outil de dépistage des troubles cognitifs légers et ses conséquences sur la prise de décision clinique. Nos résultats montrent une variation importante des indices de performance de l'outil étudié en fonction des caractéristiques démographiques des individus, de la nature et la sévérité du trouble cognitif. La performance globale d'un test d'évaluation cognitive ne serait donc pas constante et exposerait à un risque de sous-estimer ou de surestimer la présence de troubles cognitifs dans certains sous-groupes de l'échantillon d'étude. Cette variation des performances ou « effet de spectre » est susceptible, dans certaines conditions, de modifier également la probabilité post-test d'avoir des troubles cognitifs légers, et constitue ainsi un « biais spectral » pouvant affecter potentiellement la prise de décision clinique. La mesure de ces deux phénomènes de variation devrait faire partie des étapes obligatoires de la validation d'un test de dépistage des troubles cognitifs. Elle implique l'analyse d'échantillons larges et diversifiés de sujets permettant de rendre compte de la complexité de l'interprétation d'un test et son application au niveau pratique. L'utilisation de valeurs seuils ajustées selon l'âge et le niveau scolaire, est proposée afin d'améliorer la prise de décision clinique. Nous discutons également de l'intérêt d'utiliser le concept épidémiologique de dépistage dans la recherche et l'identification des troubles cognitifs et proposons le terme plus approprié de repérage qui ne préjuge pas de leurs étiologies.

Mots clés: *troubles cognitifs légers, effet de spectre, biais spectral, valeur seuil, sensibilité, spécificité, rapports de vraisemblance, dépistage, repérage, maladies neurodégénératives, maladie d'Alzheimer.*

RESUME EN ANGLAIS

Title: Mild Cognitive Impairment in the context of neurodegenerative pathologies: to screen or not to screen? Spectrum effect and spectrum bias.

In recent years, numerous tests have been proposed to screen Mild Cognitive Impairment (MCI) in order to identify individuals who are at risk for developing a neurodegenerative pathology such as Alzheimer's disease. Several indices (sensitivity, specificity, predictive values and likelihood ratios) enable to assess the performance and the utility of a test in clinical practice. The main objective of this thesis was to determine the degree of variation of an MCI screening test performance and its impact on clinical decision. Our results showed that performance indices of the study test may vary depending on the demographic characteristics of the individuals, the type and the degree of cognitive impairment. They also highlight that the overall performance of a screening test is not fixed and may result in failure to identify cognitive impairment in true cases and misclassified non cases as cognitively impaired in some subgroups of the study sample. In some situations, this subgroup variation (or «spectrum effect») may affect the posttest probability of having MCI and potentially modify the clinician's decision, known as a spectrum bias. These findings support the perspective that these two phenomena should routinely be taken into account to assess the quality of a cognitive screening test. One way to minimize the impact of these phenomena would be to focus the assessment on broad and well defined subgroups of patients. The use of adjusted cut-offs values could help clinicians to increase their level of certainty regarding whether a cognitive impairment is present. We also discuss the interest of using the epidemiological concept of screening in the identification of cognitive impairment and propose a more appropriate term of cognitive impairment detection without prejudging its aetiology.

Key words: *mild cognitive impairment, spectrum effect, spectrum bias, cut-off, sensitivity, specificity, likelihood ratios, screening, detection, neurodegenerative diseases, Alzheimer's disease.*

ABBREVIATIONS

ACE : Antigène Carcino-Embryonnaire

ADNA : Association pour le Développement de la Neuropsychologie Appliquée

AMT : Abbreviated Mental Test Score

CMRR : Centre Mémoire Ressources et Recherche

FN : Faux négatif

FP : Faux positif

GPCOG : General Practitioner Assessment of Cognition

IC : Intervalle de Confiance

LEILA 75+ : Leipzig Longitudinal Study of the Aged

MA : Maladie d'Alzheimer

MOCA : Montreal Cognitive Assessment

M@T : Memory Alteration Test

MCI : Mild Cognitive Impairment

MCIa : Mild Cognitive Impairment amnésique

MCIa-MD : Mild Cognitive Impairment amnésique Multi-Domaine

MIS : Memory Impairment Screen

MMSE : Mini-Mental State Examination

OMS : Organisation Mondiale de la Santé

PIO : Pression Intra Oculaire

P : Prévalence

p : Probabilité

PAQUID : Personnes Agées Quid

PSA : Prostatic Specific Antigen (antigène prostatique spécifique)

Rapid-Fr : Réseau d'Aide au diagnostic et à la Prise en charge des Détériorations cognitives et des pathologies neurologiques chroniques en Franche-comté et au niveau national

ROC : Receiver Operating Characteristic

RV : Rapport de Vraisemblance

RVP : Rapport de Vraisemblance Positif

RVN : Rapport de Vraisemblance Négatif

Se : Sensibilité

Sp : Spécificité

SMI : Score de Mémoire avec Indiçage

STI : Set Test d'Isaacs

VN : Vrai Négatif

VP : Vrai Positif

VPP : Valeur de Prédiction Positive

VPN : Valeur de Prédiction Négative

PUBLICATIONS EN RAPPORT AVEC LA THESE

Articles

Chopard, G., Pitard, A., Ferreira, S., Vanholsbeeck, G., Rumbach, L., Galmiche, J. (2007).

Combining the Memory Impairment Screen and the Isaacs Set Test: A practical Tool for Screening Dementias. *J Am Geriatr Soc* 55, 1426-1430. (IF: 3,539).

Chopard, G., Vanholsbeeck, G., Tio, G., Pitard, A., Binetruy, M., Rumbach, L., Galmiche J.

(2009). Rapid Screening of Cognitive Change in Patients with Questionable Dementia Using the Memory Impairment Screen and the Isaacs Set Test. *J Am Geriatr Soc* 57, 703-708. (IF: 3,656).

Chopard, G., Mauny, F., Puyraveau, M., Meyer, A., Binetruy, M., Vandiel, P., Rumbach, L.,

Galmiche J. (2012). Screening for Amnestic Mild Cognitive Impairment: Spectrum Effect and Spectrum Bias. *Alzheimer's & Dementia (article soumis)*.

Communication avec publication de résumé

Chopard, G., Vanholsbeeck, G., Ferreira, S., Magnin, E., Rumbach, L., Galmiche, J.

Diagnostic Norms for Subjects Aged between 60 and 89 Years with Mild Dementia: Single or Adjusted Cut-Off? 62nd Annual Meeting of American Academy of Neurology (Toronto, 10-17 Avril 2010), *Neurology*, 74(9), suppl. 2:A23. Poster Discussion 1.009.

TABLE DES MATIÈRES

INTRODUCTION.....	1
PREMIÈRE PARTIE : CADRE THÉORIQUE.....	4
I. PRINCIPES GÉNÉRAUX SUR LE DÉPISTAGE	4
1. Définition générale.....	4
2. Les différentes formes de dépistage.....	6
3. Les objectifs du dépistage	7
II. LES CONDITIONS REQUISES POUR LA MISE EN PLACE DU DÉPISTAGE	10
1. La pathologie concernée	10
2. Qualités métrologiques reliées au test.....	11
III. PROBLÈMES MÉTHODOLOGIQUES D'ÉVALUATION D'UN TEST.....	24
1. Les différents types de biais.....	25
2. De l'effet de spectre au biais spectral	27
3. Les critères d'un dépistage appliqués aux troubles cognitifs légers dans le cadre des maladies neurodégénératives.	33
IV. OBJECTIFS DE LA RECHERCHE.....	43
DEUXIÈME PARTIE: ÉTUDES CLINIQUES.....	47
I. ARTICLE 1: Combining the Memory Impairment Screen and the Isaacs Set Test: A practical Tool for Screening Dementias.	47
II. ARTICLE 2: Rapid Screening of Cognitive Change in Patients with Questionable Dementia Using the Memory Impairment Screen and the Isaacs Set Test.....	54
III. ARTICLE 3: Screening for Amnestic Mild Cognitive Impairment: Spectrum Effect and Spectrum Bias.....	61
IV. DISCUSSION GENERALE	82
BIBLIOGRAPHIE	97
ANNEXES	116
ANNEXE 1. Consignes de passation et de cotation du Memory Impairment Screen (MIS)...	117
ANNEXE 2. Consignes de passation et de cotation du Set Test d'Isaacs (STI).	121
ANNEXE 3. Feuille de cotation du Mini-Mental State Examination (MMSE).....	123

ANNEXE 4. Caractéristiques démographiques et neuropsychologiques de l'échantillon d'étude (article 1).	124
ANNEXE 5. Sensibilité, spécificité et valeur de prédiction positive du MIS, du STI, du MMSE et du couple MIS-STI sur l'échantillon de validation (article 1).....	125
ANNEXE 6. Organigramme de sélection des sujets (article 3).....	126
ANNEXE 7. Variation de la sensibilité et de la spécificité du couple MIS-STI (<i>valeurs seuils fixes</i>) selon le type de MCI en fonction de l'âge et du niveau scolaire (article 3).....	127
ANNEXE 8. Variation de la sensibilité et de la spécificité du couple MIS-STI (<i>valeurs seuils ajustés</i>) selon le type de MCI en fonction de l'âge et du niveau scolaire (article 3).	128
ANNEXE 9. Variation de la sensibilité et de la spécificité du MMSE (<i>valeur seuil fixe</i>) selon le type de MCI en fonction de l'âge et du niveau scolaire (article 3).....	129
ANNEXE 10. Variation de la sensibilité et de la spécificité du couple MMSE (<i>valeurs seuils ajustés</i>) selon le type de MCI en fonction de l'âge et du niveau scolaire (article 3).	130
ANNEXE 11. Evaluation du biais spectral pour une valeur seuil fixe du MMSE dans le groupe des cas MCI amnésiques avec déficits multiples (article 3).	131
ANNEXE 12. Algorithme de repérage des troubles cognitifs.....	132

INTRODUCTION

Les maladies neurodégénératives constituent un important problème de santé publique en raison notamment des troubles cognitifs qui les caractérisent. La maladie d'Alzheimer (MA) est la pathologie neurodégénérative la plus fréquente chez la personne âgée de plus de 65 ans (Lobo *et al.*, 2000). Elle se manifeste par une détérioration progressive des fonctions cognitives en particulier de la mémoire, avec un retentissement significatif sur les activités de la vie courante (McKhann *et al.*, 2011). En France, près de 800 000 personnes seraient affectées par la MA ou une maladie apparentée (Ankri, 2006). Le diagnostic précoce de ce type de pathologie apparaît donc légitime afin de permettre une meilleure prise en charge du patient et de sa famille. A un stade précoce, les traitements pharmacologiques et non pharmacologiques pourraient notamment être plus efficaces pour ralentir la progression de la maladie et améliorer la qualité de vie des patients ainsi que des aidants (Calahan *et al.*, 2006).

Dans cette optique, de nombreux travaux se sont intéressés ces dernières années aux patients présentant un trouble cognitif léger ou « Mild Cognitive Impairment » (MCI). Ce dernier est caractérisé par la présence d'une atteinte cognitive légère, notamment dans le domaine mnésique, chez des patients qui restent autonome dans la vie quotidienne (Petersen *et al.*, 1999). L'identification des patients MCI revêt un enjeu majeur car ils présentent un risque accru d'évolution vers une MA ou une maladie apparentée (Petersen *et al.*, 1999). Dès lors, le stade du MCI a souvent été considéré comme une phase prodromale de la MA (Dubois, 2000).

Cette approche pose la question, encore débattue, de la pertinence du dépistage des personnes présentant un MCI (www.cnge.fr). Le concept de dépistage est-il approprié dans le cas du MCI ? L'utilisation du terme dépistage comme procédure de recherche systématique d'une maladie spécifique, comme la MA, ne conduit-elle pas à entretenir la confusion avec le diagnostic précoce ? Si tel est le cas, n'est-il pas préférable de substituer au concept de dépistage celui de « repérage », qui ne préjuge pas d'une étiologie particulière du MCI ?

Par ailleurs, le dépistage du MCI nécessite d'utiliser un ou plusieurs tests rapides, simples et dont les qualités métriques requises sont validées (Gifford et Cummings, 1999). Un nombre croissant de tests ont été développés ces dernières années et la plupart des études ont montré que l'utilisation d'une valeur fixe du seuil d'anormalité pouvait donner des résultats intéressants dans le dépistage du MCI (Nasredine *et al.*, 2005 ; Lonie *et al.*, 2009 ; Lonie *et al.*, 2010). Toutefois, l'évaluation de la validité de ces résultats, avant leur généralisation en

pratique clinique, reste insuffisante. L'examen de la littérature médicale suggère en effet que les indices de performance de nombreux tests biologiques pouvaient varier en fonction du stade de la maladie (Hlatky *et al.*, 1984 ; Fletcher, 1986 ; Junod, 1995), de l'intensité des symptômes (Hlatky *et al.*, 1984 ; Lachs *et al.*, 1992), des co-morbidités présentes (Ayala *et al.*, 2003), des caractéristiques démographiques de la population étudiée (Mulherin et Miller, 2002 ; Parentis *et al.*, 2006) ou de son cadre de sélection (Harper et Reeves, 1999). Cette variation des performances ou « effet de spectre » est susceptible de modifier également la probabilité post-test d'avoir la maladie ou l'anomalie recherchée, et constitue ainsi un « biais spectral » pouvant affecter potentiellement la prise de décision clinique (Goerhing *et al.*, 2004). Le degré de variation des performances d'un test et ses conséquences sur l'interprétation des résultats, n'ont encore jamais fait l'objet d'une analyse dans le cadre du dépistage des troubles cognitifs.

Notre travail se déclinera en deux parties :

La première partie est consacrée au contexte théorique de cette thèse en rappelant d'abord les différentes formes du dépistage ainsi que ses objectifs tels qu'ils ont été fixés par l'Organisation Mondiale de la Santé (OMS) (Wilson et Jungner, 1970). Nous avons ensuite décrit les critères de mise en place du dépistage et présenté les différentes propriétés métrologiques requises pour mesurer l'efficacité d'un test. Nous montrons ensuite comment la performance globale d'un test peut être affectée par de nombreux biais en introduisant plus particulièrement deux paramètres supplémentaires : l'effet de spectre et le biais spectral. A la fin de cette première partie, nous abordons la difficulté d'appliquer la notion de dépistage dans la recherche des troubles cognitifs légers dans le cadre des maladies neurodégénératives.

Dans la seconde partie, nous présentons les travaux que nous avons réalisés concernant l'évaluation d'un outil de détection des troubles cognitifs légers. Les résultats obtenus sont présentés sous la forme de trois articles.

Nous discutons finalement des problèmes méthodologiques posés par la validation psychométrique d'un outil d'évaluation rapide des troubles cognitifs légers et son utilisation en pratique clinique.

L'ensemble du travail présenté dans cette thèse s'inscrit dans la thématique du réseau Rapid-Fr (Réseau d'Aide au diagnostic et à la prise en charge des Détériorations cognitives et des pathologies neurologiques en Franche-Comté et au niveau national) (Vanholsbeeck *et al.*, 2007) en partenariat avec le Centre de Mémoire de Ressources et de Recherche (CMRR) du CHU Jean Minjoz de Besançon et les différentes consultations mémoire de Franche-Comté qui assurent depuis de nombreuses années, l'évaluation, la prise en charge et le suivi des personnes présentant des troubles cognitifs.

PREMIÈRE PARTIE : CADRE THÉORIQUE

I. PRINCIPES GÉNÉRAUX SUR LE DÉPISTAGE

1. Définition générale

La prévention des maladies repose sur deux types d'intervention :

- La prévention primaire qui est l'ensemble des moyens mis en œuvre pour empêcher l'apparition d'un trouble, d'une pathologie ou d'un symptôme (Wilson et Jungner, 1970). Elle consiste à diminuer un risque susceptible de provoquer une maladie à l'aide de mesures à l'échelon de l'individu (ex : vaccination pour les maladies infectieuses) ou de la collectivité (campagne d'arrêt du tabagisme pour le cancer du poumon).
- La prévention secondaire qui elle, est la prise en charge du problème au tout début de l'apparition d'un trouble qui peut ainsi être enrayé (Wilson et Jungner, 1970). Elle vise par conséquent à stopper ou ralentir l'évolution d'une maladie avant l'apparition de complications importantes. Elle comprend le dépistage, le diagnostic et le traitement.

Le dépistage est donc un acte majeur de la prévention secondaire. Il cherche à identifier les signes précoces d'une maladie à l'aide d'examens ou de tests. Par exemple, le test Hemocult[®] est utilisé comme test de dépistage de sang dans les selles dans le cadre du cancer du côlon. Si le test est positif, il est complété par un examen endoscopique pour poser le diagnostic ou non de cancer du côlon ou de lésions précancéreuses (polypes) permettant ainsi un traitement de ces lésions.

Pour comprendre les principaux enjeux du dépistage, il apparaît nécessaire de préciser les différents concepts qui recouvrent ce terme, ses objectifs et ses différentes formes.

Plusieurs définitions du dépistage ont été proposées :

Selon l'OMS (Wilson et Jungner, 1970) « *le dépistage consiste à identifier présomptivement à l'aide de tests, d'examens ou d'autres techniques susceptibles d'une*

application rapide, les sujets atteints d'une maladie ou d'une anomalie passées jusque-là inaperçues. Les tests de dépistage doivent permettre de faire le partage entre les personnes apparemment en bonne santé mais qui sont probablement atteintes d'une maladie donnée et celles qui en sont probablement exemptes. Ils n'ont pas pour objet de poser un diagnostic. Les personnes pour lesquelles les résultats sont positifs ou douteux doivent être envoyées à leur médecin pour diagnostic et, si besoin est, traitement ». (Wilson et Jungner, 1970, p. 11). Dans cette définition, le dépistage consiste à détecter rapidement, à un stade précoce, des personnes atteintes d'une maladie ou d'une anomalie pour permettre un diagnostic et un traitement si besoin.

Pour Morrison (1992), le dépistage est *« l'examen d'individus asymptomatiques pour déterminer leur probabilité d'avoir la condition qui fait l'objet du dépistage. Les individus dont la probabilité d'être atteints est suffisamment élevée, sont ensuite soumis à des investigations diagnostiques complètes. Les individus avec la condition sont alors traités ».* Dans cette perspective, le dépistage est défini comme un triple processus : *« Le premier a pour but de déterminer la probabilité qu'un individu présente une condition (le plus souvent, une maladie, mais aussi une condition pré-morbide ou un facteur de risque). Le second processus vise à établir un diagnostic, et le troisième consiste à intervenir en cas de diagnostic positif ».* (Guessous, 2010, p. 1390). On entend par le mot intervention, un traitement curatif (médical ou chirurgical) ou une surveillance en répétant régulièrement le test de dépistage (Guessous, 2010).

La Rosa (1996) définit le dépistage comme une *« action de santé, qui consiste à identifier, dans une population apparemment saine, une maladie (défiance ou facteur de risque) non reconnue ou asymptomatique, en utilisant des tests qui répondent à des critères de sécurité, de simplicité, d'acceptabilité, de validité et de coût satisfaisants ».* La Rosa (1996, p. 7). Pour cet auteur, le dépistage s'effectue sur une population asymptomatique. Il met aussi l'accent sur l'efficacité et la sécurité des tests de dépistage et leurs implications économiques.

Wald (2004) enfin définit le dépistage parmi les populations asymptomatiques, comme *« L'application systématique d'un examen ou d'une enquête pour identifier les individus à risque d'une...maladie, qui bénéficieront d'autres investigations médicales...ou actions préventives...».* Cette définition se rapproche de celle de l'OMS.

Il ressort de ces différentes définitions, trois éléments principaux du dépistage :

- Il permet de rechercher et d'identifier une maladie mais aussi une anomalie ou un facteur de risque qui, bien que susceptibles de provoquer une maladie, ne constituent pas à eux seuls une maladie (Canguilhem, 1952).
- Il s'effectue dans une population apparemment saine en l'absence de symptômes. L'intérêt du dépistage des cas non symptomatiques est de permettre notamment d'identifier les éventuelles anomalies durant la phase latente d'une maladie, c'est-à-dire à un stade précoce de son développement.
- Il fait enfin référence à l'utilisation systématique de tests qui obéissent à un certain nombre de critères : rapides, simples, bien acceptés par la population, valides et peu coûteux.

2. Les différentes formes de dépistage

On distingue plusieurs modes de dépistage :

- *de masse* : c'est une mesure de grande envergure qui s'applique à toute une population comme le dépistage de la phénylcétonurie chez tous les nouveau-nés (Guthrie et Susi, 1963).

- *sélectif* : il concerne le plus souvent l'ensemble des individus d'une population cible choisie en raison de risques élevés. Aujourd'hui, par exemple, le cancer du sein est systématiquement dépisté par mammographie (examen radiographique des seins) chez les femmes de plus de 50 ans (Fletcher *et al.*, 1993). Selon l'OMS (Wilson et Jungner, 1970), le dépistage sélectif peut être considéré comme une forme de dépistage de masse lorsqu'il est appliqué à un grand nombre d'individus.

Ces deux types de dépistage dits « organisés » sont initiés par les pouvoirs publics et doivent avoir fait la preuve de leur bénéfice pour la population. Ils s'appuient sur la participation volontaire de sujets non demandeurs.

Mais il existe aussi d'autres formes de dépistage :

- *individuel* : ce dépistage ne suit pas un protocole général déterminé à l'avance. Il est réalisé au « cas par cas » à l'initiative du médecin ou du patient qui bien que non symptomatique, peut ainsi se voir proposer, au cours d'une consultation, un acte médical tels qu'une mesure de la tension artérielle, une palpation des seins ou un toucher rectal en tenant compte des facteurs de risques ou d'antécédents familiaux et personnels. Il peut s'agir d'un dépistage *opportuniste* lorsqu'il est adressé à des personnes consultant pour un autre motif que la maladie ou le problème de santé objets du dépistage. Le dépistage individuel est différent de la pratique médicale courante où des examens sont pratiqués à des patients qui se plaignent de certains symptômes (Wilson et Jungner, 1970).

- *multiple* : il a été défini comme « *l'application simultanée de deux ou plusieurs tests de dépistage à des groupes nombreux* » (Wilson et Jungner, 1970, p. 12), et consiste à rechercher simultanément plusieurs affections lors d'une même visite comme c'est le cas du dépistage prénuptial.

Il existe enfin une action dite de *surveillance* utilisant des outils similaires à ceux utilisés au cours d'un dépistage : ce terme désigne une « *opération de longue durée au cours de laquelle des examens de dépistage sont pratiqués par intervalles* ». (Wilson et Jungner, 1970, p. 12) et s'oppose au dépistage qui tend à être considéré « *comme une opération transversale de courte durée portant sur une population exposée* » (Wilson et Jungner, 1970, p. 12). C'est le cas, par exemple, des examens qui sont faits régulièrement pour dépister l'hypertension artérielle ou un excès de cholestérol ou encore la réalisation d'une mammographie tous les 2 ans chez les femmes à partir de 50 ans. Ce dernier cas est en quelque sorte la prolongation dans le temps d'une même approche clinique après un premier dépistage.

3. Les objectifs du dépistage

Selon l'OMS, « *le dépistage précoce (recherche de cas) a pour objet de découvrir et de traiter des maladies ayant déjà produit une altération pathologique mais n'ayant pas encore atteint le stade auquel l'intéressé vient spontanément se faire soigner* ». (Wilson et Jungner, 1970, p. 14).

L'objectif du dépistage est souvent défini comme un moyen permettant d'identifier une maladie à son stade précoce afin d'améliorer par un traitement l'état de santé des individus qui en sont atteints. Or, parler de dépistage d'une maladie, c'est prendre le risque de confondre dépistage et diagnostic précoce qui sont de nature différente. Le dépistage permet de rechercher des signes qui peuvent faire suspecter (et suspecter seulement) la présence (résultat positif) ou l'absence (résultat négatif) d'une maladie ou d'une anomalie, chez des individus qui ne se plaignent pas. Autrement dit, le dépistage d'un ou de plusieurs signes qui seraient évocateurs d'une maladie donnée n'est pas le reflet direct de cette maladie dans la mesure où les signes peuvent avoir des causes multiples liées à d'autres problèmes (parfois bénins) que la maladie recherchée. Le dépistage permet donc d'estimer la probabilité d'existence d'une maladie ou d'une anomalie (Guessous, 2010). Le résultat d'un test de dépistage n'est pas un diagnostic et seuls les sujets présentant un test positif, pourront bénéficier d'examens complémentaires (ou tests diagnostiques) permettant de déterminer avec une plus grande certitude la présence ou l'absence d'une maladie ou d'une anomalie.

Par exemple, le frottis cervico-vaginal permet d'identifier les femmes présentant un risque plus élevé de cancer du col de l'utérus, mais cet examen ne pose pas le diagnostic. Il faut que les femmes qui présentent un frottis anormal se soumettent à une biopsie du col pour réalisation d'un examen anatomo-pathologique qui permettra de poser un diagnostic.

Si cette différence peut sembler évidente, il n'en reste pas moins que dans la pratique clinique, ces deux termes sont parfois confondus. En cas de dépistage, une réponse positive à un test peut amener à considérer que ce test est la première étape du diagnostic. D'autre part, la confusion faite entre dépistage et diagnostic étiologique peut amener l'idée que dépister un signe (ou des signes) correspond de fait au diagnostic précoce de la maladie dans laquelle on le rencontre le plus souvent.

Le dépistage cherche en fait à sélectionner les individus avec une probabilité plus élevée de présenter une condition (maladie ou anomalie). Les sujets positifs au test de dépistage doivent alors réaliser secondairement un bilan diagnostique destiné à séparer ceux ayant la condition recherchée (diagnostic positif) de ceux qui n'ont pas cette condition (diagnostic négatif) et le cas échéant les faire bénéficier d'un traitement ou d'une intervention dans les meilleurs délais.

La validité d'un test de dépistage est démontrée par l'amélioration de l'état de santé ou de la qualité de vie chez les patients dépistés par rapport à ceux qui ont été identifiés au stade symptomatique. Il permet notamment de réduire l'incidence d'une maladie. Par exemple, selon Sasieni *et al.* (2009) la politique du dépistage du cancer du col utérin par le frottis cervico-utérin permet de réduire de 60 % le nombre de cancer du col utérin chez les femmes âgées de 40 ans, et 80 % chez les femmes âgées de 64 ans.

En pratique, chaque technique de dépistage doit être évaluée dans ses effets positifs et négatifs.

Les avantages :

- un meilleur pronostic pour la plupart des patients malades dépistés grâce au test (cas vrais positifs).
- si possible une diminution de la mortalité ou de la morbidité en rapport avec le dépistage.
- la diminution du coût total de la prise en charge de la maladie.
- rassurer les patients non malades ayant un test négatif (cas vrais négatifs).

Les inconvénients :

- l'inconfort de certains tests de dépistage et de diagnostic.
- le coût lié à la réalisation du test de dépistage et aux examens complémentaires induits par les cas faux positifs (patients non malades ayant un test positif) qui peuvent surcharger inutilement les services de santé.
- inquiétude chez les cas faux positifs qui vont réaliser inutilement des examens complémentaires.
- réassurance à tort des cas faux négatifs (patients malades non dépistés) pouvant différer le diagnostic et le traitement d'une maladie potentiellement curable.

II. LES CONDITIONS REQUISES POUR LA MISE EN PLACE DU DÉPISTAGE

Pour juger de la nécessité de la mise en place d'un dépistage, il importe de répondre à des critères précis qui ont été fixés par l'OMS en 1970 (Wilson et Jungner, 1970). Certains critères sont reliés aux problèmes de santé à dépister, d'autres aux qualités du test utilisé.

1. La pathologie concernée

1. La maladie dont on recherche les cas constitue un problème majeur de santé publique. L'importance d'une pathologie est définie par sa prévalence et son incidence au sein d'une population donnée. Les conséquences en termes de morbidité et de mortalité vont déterminer sa gravité.

2. Il doit exister un traitement efficace pouvant être administré aux sujets chez lesquels la maladie a été identifiée. Le traitement doit permettre d'arrêter ou de ralentir la progression de la maladie, améliorer la survie des patients ou leur qualité de vie par rapport à l'absence de dépistage.

3. Il faut disposer de moyens de diagnostic et de traitement. Il s'agit du personnel et des structures nécessaires pour accueillir la population visée par le dépistage, appliquer les tests de dépistage et les examens de diagnostic, instaurer le traitement et assurer le suivi des patients.

4. La maladie doit être dépistée pendant une phase de latence ou au début de la phase clinique. Le traitement est d'autant plus efficace que les patients sont identifiés précocement.

5. Le test de dépistage doit être efficace. Il est suffisamment performant afin de réduire les erreurs d'interprétation de son résultat.

6. Le test utilisé doit être acceptable par la population dépistée. Il est simple, rapide et facile à réaliser, le moins invasif possible et sans danger pour les individus.

7. *L'histoire naturelle de la maladie doit être bien connue, notamment son évolution de la phase de latence à la phase symptomatique.* La connaissance de cette phase de latence permet notamment de savoir si le dépistage peut diminuer la mortalité ou la morbidité d'une maladie.

8. *Le choix des sujets qui seront traités doit être opéré selon des critères préétablis.* Ces critères doivent permettre d'identifier clairement les sujets qui pourront bénéficier d'un diagnostic et d'un traitement si besoin.

9. *Il faut que les frais engendrés par le dépistage, le diagnostic et le traitement des sujets malades ne soient pas disproportionnés par rapport au coût total des soins médicaux.* Le dépistage doit apporter un bénéfice en termes d'économie de santé.

10. Il faut assurer la continuité d'action dans la recherche des cas et non la considérer comme une opération exécutée « une fois pour toutes ». Il est d'important d'assurer le suivi de certains patients en répétant le test de dépistage.

On voit donc que même si les définitions du dépistage ne se limitent pas au dépistage d'une maladie, cette liste ramène de fait à ce concept.

2. Qualités métrologiques reliées au test

Pour évaluer l'intérêt d'un test à des fins de dépistage, il faut vérifier au préalable qu'il ait les qualités suivantes :

- Il doit être rapide et facile à appliquer.
- La cotation du test doit être simple tout comme l'interprétation de son résultat.
- Le test doit être bien accepté par la population cible à laquelle il est adressé.
- Il doit enfin être de coût modéré.

Outre l'utilité pratique d'un test, il faut rajouter ses qualités métrologiques. L'évaluation métrologique d'un test est essentielle pour justifier de sa pertinence. Les deux principales qualités recherchées pour un test de dépistage sont la fidélité et la validité (Golafshani, 2003).

Un test est fidèle lorsqu'il donne des résultats stables dans des situations comparables. Il existe divers types d'estimation de la fidélité. Le plus courant est la fidélité test-retest qui se réfère à la stabilité des résultats obtenus à un test lorsqu'il est appliqué au même sujet à deux moments différents.

Un outil de dépistage avec une faible fidélité test-retest introduit une erreur de mesure dans le résultat obtenu par le patient dont le score va varier considérablement d'une passation à l'autre. A l'inverse, une bonne fidélité test-retest indique que le changement de score mesuré chez le patient est réel et non lié à une erreur de mesure.

Pour s'assurer que le résultat obtenu à un test est le même quelle que soit la personne qui administre ce test, on mesure la fidélité inter-examineur. Pour la fidélité inter-examineur, on recherche la concordance des résultats obtenus par deux (ou plusieurs) examinateurs qui administrent et cotent les mêmes sujets avec le même test.

Pour améliorer ce type de fidélité, il convient de standardiser l'administration du test et sa cotation en formant le personnel. Au cours de cette formation, il faut s'assurer que les examinateurs respectent rigoureusement les consignes de passation et les méthodes de cotation qui devront être très précises et formulées très clairement.

La validité d'un test de dépistage renvoie à sa capacité de mesurer ce qu'il est censé mesurer (validité constructive) mais aussi de différencier au sein d'une population cible les sujets ayant une forte probabilité d'avoir une maladie ou une anomalie de ceux qui en sont exempts (validité discriminative). Cette capacité dépend à la fois des performances propres (ou intrinsèques) au test de dépistage (sensibilité et spécificité) et des caractéristiques de la population ciblée (notamment la prévalence) qui conditionnent les performances extrinsèques du test (valeurs de prédiction positive et négative) dans cette population.

Pour mesurer la validité d'un test de dépistage, il faut pouvoir le comparer à un test de référence de qualité (ou « gold standard ») qui permet de déterminer avec la plus grande certitude possible la présence ou l'absence de la maladie ou de l'anomalie recherchée.

Plusieurs indices sont utilisés afin d'évaluer la précision d'un test de dépistage. Ces indices sont résumés dans un tableau de contingence (Tableau 1). Dans ce tableau, les sujets sont classés comme étant malades ou non malades avec le test de référence. Le résultat du test de dépistage étudié est positif (anormal) ou négatif (normal).

Test à évaluer	Malade	Non malade	Total
Résultat positif	VP	FP	VP + FP
Résultat négatif	FN	VN	FN + VN
Total	VP + FN	FP + VN	n

Tableau 1 : Tableau de contingence.

Les sujets (n=effectif total) se répartissent ainsi en quatre catégories (Delacour *et al.* 2009a) :

- Les vrais positifs (VP) : sujets malades avec un résultat positif.
- Les vrais négatifs (VN) : sujets non malades avec un résultat négatif.
- Les faux positifs (FP) : sujets non malades avec un résultat positif.
- Les faux négatifs (FN) : sujets malades avec un résultat négatif.

A partir de ces quatre catégories, il peut être défini et calculé une série d'indices permettant de caractériser les performances d'un test.

La sensibilité (Se) est définie comme la proportion de personnes qui ont effectivement la maladie et dont le test de dépistage est positif. La sensibilité exprime donc la probabilité que le test soit positif alors que la personne est malade.

La sensibilité (Se) est donnée par le nombre de vrais positifs divisé par le nombre total de personnes atteintes de la maladie :

$$\textit{Sensibilité} = VP/(VP+FN)$$

Plus la sensibilité est élevée meilleures sont les chances de dépister les personnes malades avec le test. Ainsi, un test avec une très bonne sensibilité manquera peu de cas.

Par opposition, la proportion de personnes atteintes de la maladie et dont le test de dépistage est négatif constitue le taux de faux négatifs.

$$\textit{Faux négatifs} = 1\text{-sensibilité}$$

La spécificité (Sp) correspond à la proportion de personnes qui n'ont pas la maladie et dont le test de dépistage est négatif (vrais négatifs). La spécificité exprime la probabilité que le test soit négatif alors que la personne n'a pas la maladie recherchée.

Elle est donnée par le nombre de vrais négatifs divisé par le nombre total de personnes indemnes de la maladie :

$$\textit{Spécificité} = VN/(VN+FP)$$

La spécificité exprime donc la capacité du test à identifier les personnes indemnes de la maladie.

Par opposition, la proportion de personnes non malades chez qui le test de dépistage est positif constitue le taux de faux positifs.

$$\textit{Faux positifs} = 1\text{-spécificité}$$

La sensibilité et la spécificité sont des mesures imparfaites car elles ne reflètent la capacité discriminative d'un test que pour une valeur seuil donnée. La grande majorité des tests s'expriment par des résultats quantitatifs, il est donc nécessaire de définir une valeur seuil (ou

cut-off en anglais) permettant de classer le résultat du test en positif (anormal) ou négatif (normal).

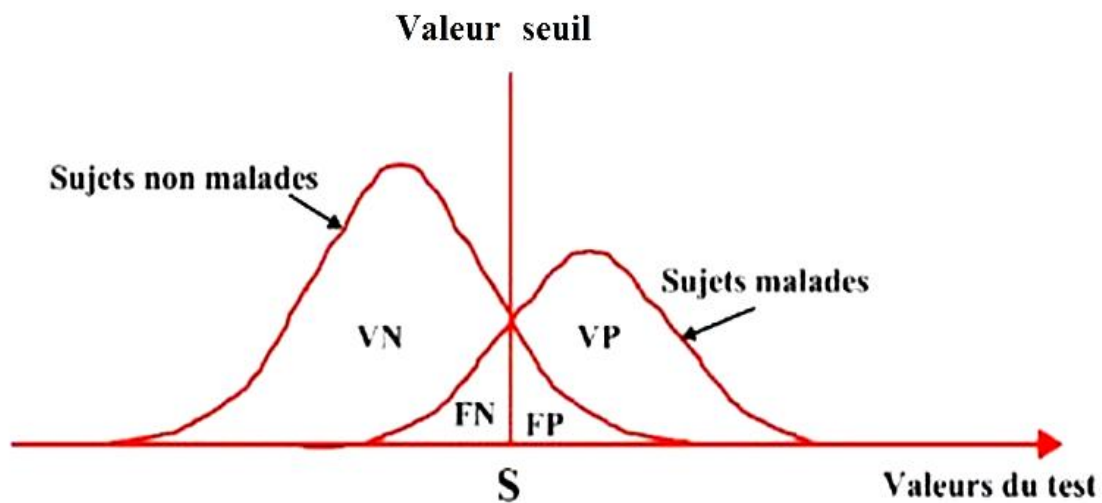


Figure 1 : Représentation des résultats d'un test selon une valeur seuil donnée.

La Courbe ROC (Receiver Operating Characteristics) permet de pallier cet inconvénient (Hanley et McNeil, 1982). Cet outil est une représentation graphique de la relation existante entre le taux de vrais positifs (sensibilité, en ordonnée) et le taux de faux positifs (1-spécificité, en abscisse) obtenus pour toutes les valeurs seuil possibles d'un test (Delacour *et al.* 2005). Cette courbe permet d'exprimer la capacité globale d'un test de faire la distinction entre les personnes saines et les personnes malades. L'aire sous la courbe donne donc une information plus synthétique sur la précision d'un test (Penerger et Perrier, 2004). Elle est égale à 1 si le test est parfait et peut donc identifier tous les malades (sensibilité de 1 ou 100%) sans faux positifs (spécificité de 1 ou 100% ou 1-spécificité=0%), ce qui correspond au coin supérieur gauche du graphique (figure 2). Par conséquent, plus l'aire sous la courbe est proche de 1, plus le test a une précision élevée. Si cette aire est de 0,5, le test est inutile car détectant autant de vrais positifs que de faux positifs. Dans ce cas, la courbe ROC est une courbe à 45 degrés. Pour la plupart des tests, l'aire sous la courbe ROC se trouve entre ces deux extrêmes. Plus la courbe s'éloigne de la diagonale pour rejoindre l'angle supérieur gauche, plus un test est globalement discriminant (figure 2).

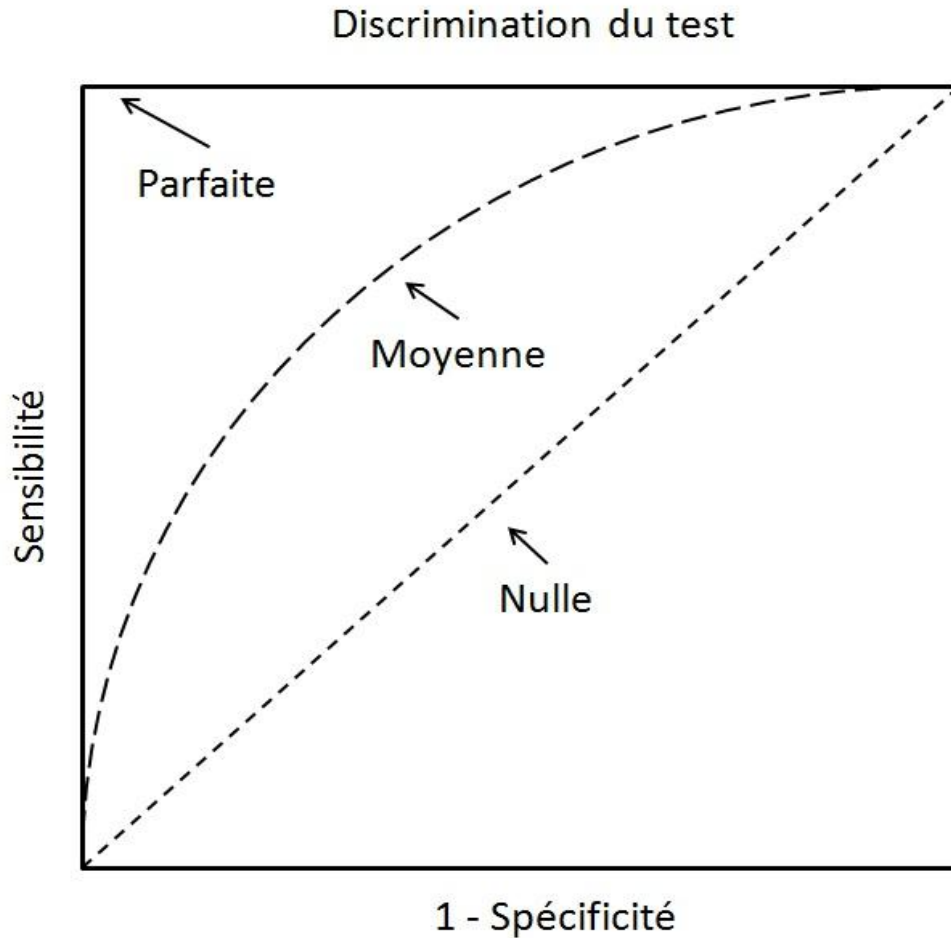


Figure 2 : Courbe de « Receiver Operating Characteristic » ou ROC, pour des tests ayant une capacité de discrimination parfaite, moyenne, et nulle.

La courbe ROC est également employée pour comparer la précision de deux, voire de plusieurs tests sur une même population d'étude. Elle est aussi utilisée pour estimer la valeur seuil d'un test (la valeur à laquelle on décide qu'il devient positif ou négatif).

Le choix du seuil d'un test permet de privilégier :

- soit la sensibilité : presque tous les malades seront détectés (très peu de faux négatifs) pour ne laisser passer aucune personne qui serait affectée par une maladie grave potentiellement curable en cas de diagnostic à un stade précoce (cancer du sein ou du côlon, virus d'immunodéficience acquise).

- soit la spécificité : presque tous les non malades auront un test négatif avec très peu de faux positifs pour ne pas inquiéter les personnes à tort, ne pas leur imposer des tests complémentaires de confirmation diagnostique parfois désagréables (coloscopie), dangereux (biopsie) ou coûteux (imagerie cérébrale).

Cependant, ces deux critères, sensibilité et spécificité, varient en sens inverse. Par conséquent, plus un test est sensible, moins il est spécifique. Par exemple, la valeur du PSA (Prostatic Specific Antigen ou antigène prostatique spécifique) utilisée dans le dépistage du cancer de la prostate est augmentée également au cours d'une prostatite (infection aiguë ou chronique de la prostate) (Fletcher, 2000). A l'inverse, plus un test est spécifique, moins il est sensible. Par exemple, un taux d'ACE (Antigène Carcino-Embryonnaire) est très peu élevé chez des personnes atteintes d'une forme peu évoluée de cancer du côlon qui seront rassurées à tort (Fletcher, 1986).

Afin de limiter ces erreurs par excès (faux positif) ou par défaut (faux négatif), un test doit offrir un bon compromis entre une sensibilité élevée et une spécificité élevée. La courbe ROC permet de déterminer la valeur seuil optimale correspondant à la meilleure combinaison possible de sensibilité et de spécificité. Cette valeur seuil optimale correspond au point le plus près du coin supérieur gauche (figure 2).

En résumé, la sensibilité et la spécificité permettent d'apprécier les qualités intrinsèques d'un test de dépistage, c'est-à-dire, sa capacité à détecter et séparer correctement des personnes probablement malades de celles probablement non malades. Autrement dit, elles donnent la probabilité d'avoir un test positif quand une personne est malade (sensibilité) et un test négatif dans le cas contraire (spécificité). Cependant, ces deux indices ne permettent pas de répondre à la question qui intéresse le clinicien dans sa pratique quotidienne : quelle est la probabilité que mon patient soit malade (ou non malade) en fonction du résultat positif (ou négatif) de son test ?

Pour y répondre, deux valeurs de prédiction sont utilisées :

Valeur de prédiction positive (VPP) : c'est la probabilité que la personne soit réellement malade en cas de test positif. C'est le nombre de vrais positifs (tests positifs chez les personnes atteintes de la maladie) divisé par le nombre total de personnes dont le test est positif :

$$VPP = VP/(VP+FP)$$

Plus la VPP est proche de 1, plus le patient a de risque d'avoir la maladie ou l'anomalie recherchée si son test est positif.

Valeur de prédiction négative (VPN) : c'est la probabilité que la personne n'ait pas la maladie en cas de test négatif.

C'est le nombre de vrais négatifs (tests négatifs chez les personnes indemnes de la maladie) divisé par la somme totale de personnes dont le test est négatif :

$$VPN = VN/(VN+FN)$$

Plus la VPN est proche de 0, plus le patient a de chance d'être indemne de la maladie si son test est négatif.

Les valeurs prédictives dépendent de la sensibilité et de la spécificité du test ainsi que de la prévalence de la maladie (P). La formule de Bayes (1764) permet de calculer la VPP et la VPN en fonction de ces trois paramètres :

$$VPP = Se \times P / Se \times P + [(1-P) \times (1-Sp)]$$

et

$$VPN = Sp \times (1-P) / Sp \times (1-P) + [P \times (1-Se)]$$

Ainsi, pour une prévalence P de la maladie donnée, plus le test est sensible, meilleure est la VPP et plus le test est spécifique, meilleure est la VPN. De même, pour une sensibilité et une spécificité fixées, plus la prévalence de la maladie est élevée dans une population, plus la VPP tend vers 1 et la VPN vers 0. A l'inverse, plus la prévalence est faible et plus la VPP tend vers 0 et la VPN vers 1 (figure 3).

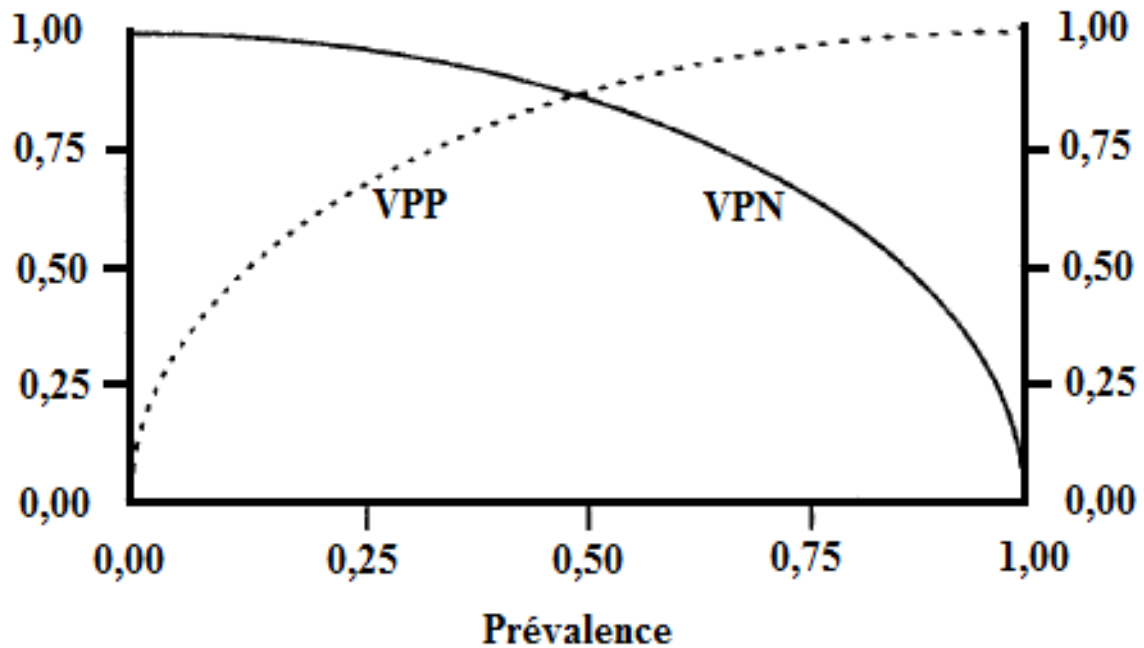


Figure 3 : Variation des valeurs prédictives positive (VPP) et négative (VPN) avec la prévalence de la maladie.

Cette influence de la prévalence sur le calcul des valeurs prédictives a un inconvénient au niveau pratique : la VPP et la VPN retrouvées dans une population d'étude, ne peuvent pas être appliquées dans une autre population où la prévalence de la maladie serait différente (Delacour *et al.*, 2009b). C'est pourquoi les cliniciens sont de plus en plus enclins à utiliser les rapports de vraisemblance (RV) ou Likelihood Ratios en anglais (Jaeschke *et al.*, 1994 ; Delacour *et al.*, 2009b).

Le RV estime le rapport entre la probabilité du résultat du test chez les sujets malades à celle d'avoir ce même résultat chez les sujets sains. Il exprime combien de fois le résultat du test est plus fréquent chez les sujets malades par rapport à ceux qui ne le sont pas.

On distingue deux types de RV selon le résultat du test :

RV d'un test positif (RVP) : il permet d'estimer dans quelle mesure la présence d'une maladie chez un patient est plus plausible après un résultat de test positif. C'est la relation entre la probabilité d'un test positif chez les malades (Se) et celle chez les non-malades (1-Sp).

$$RVP = Se/(1-Sp)$$

Si le RVP est > 1 , cela signifie que le test positif est plus fréquent chez les sujets qui ont la maladie. Plus le RVP est élevé (ou tend vers l'infini) et plus un résultat positif est fiable. A titre d'exemple, un RVP de 6 signifie qu'il y a six fois plus de risque de présenter un test positif lorsque la personne est malade que lorsque la personne n'est pas malade. Autrement dit, plus le RVP est élevé, plus le risque que le patient ait la maladie est élevé.

RV d'un test négatif (RVN) : il permet d'estimer dans quelle mesure une maladie chez un patient est moins plausible après un résultat de test négatif.

C'est la relation entre la probabilité d'un test négatif chez les malades ($1-Se$) et celle chez les non-malades (Sp).

$$RVN = (1-Se)/Sp$$

Un RVN est compris entre 0 et 1. Plus le RVN est faible (ou se rapproche de 0) et plus un résultat négatif est fiable. Par exemple, un RVN de 0,25 signifie qu'il y a quatre fois plus de chance de présenter un test négatif lorsque la personne n'est pas malade que lorsque la personne est malade. Autrement dit, plus le RVN est bas, plus la probabilité que le patient soit exempt de la maladie est élevée.

Un RV égal à 1 signifie que le résultat du test (qu'il soit positif ou négatif) est aussi fréquent chez les malades que chez les non malades et que la réalisation de ce test n'apporte donc aucune information utile au clinicien.

L'intérêt des RV est qu'ils permettent de répondre à la question du clinicien : quelle est la probabilité qu'un patient donné soit malade si son test est positif ou non malade si son test est négatif ? Contrairement aux valeurs de prédiction, les RV sont indépendants de la prévalence de la maladie dans la population et peuvent donc être transposés d'une population à une autre.

Par conséquent, ils permettent de calculer la probabilité post-test de la maladie chez des patients dont le résultat du test est positif (ou négatif) pour des contextes cliniques différents. Cette probabilité post-test dépend des caractéristiques du test (RVP et RVN) et de la probabilité pré-test que la personne ait une maladie. Le clinicien estime cette probabilité pré-test grâce à son expérience, l'anamnèse, les antécédents de son patient, la prévalence de la

maladie dans la population dont est issu le patient ou encore à l'aide d'un système de score clinique dédié à cet effet (Kivipelto *et al.*, 2004 ; Dalziel, 2009).

Les RV se basent sur la théorie des cotes (traduction de l'anglais *odd*) (Bayes, 1763 ; Elwood, 1993). Le théorème de Bayes (Bayes, 1763 ; Elwood, 1993) permet le calcul de la probabilité post-test de la maladie chez des patients d'après les résultats du test. Il utilise pour cela le rapport de la probabilité (p) qu'un évènement se produit sur celle qu'il ne se produit pas, appelé cote :

$$Cote = p/(1-p)$$

Dans le domaine clinique, une cote pré-test est le rapport du nombre de sujets malades divisé par le nombre de sujets non malades, soit :

$$Cote\ pré-test = VP+FN/VN+FP = prévalence/1-prévalence$$

Le théorème de Bayes (Bayes, 1764 ; Elwood, 1993) stipule que la cote post-test est égale au produit de la cote pré-test par le RV du test.

Ainsi, le RVP (ou RVN) obtenu est le facteur par lequel il faut multiplier la cote pré-test pour obtenir la cote post-test chez les patients qui ont un test positif (ou négatif).

Autrement dit, le RVP exprime le degré auquel un test de dépistage positif augmente la probabilité qu'une personne soit atteinte de la maladie ; le RVN exprime le degré auquel un test de dépistage négatif diminue la probabilité qu'elle soit atteinte de cette même maladie.

$$Cote\ pré-test \times RVP = cote\ post-test, test\ positif$$

$$Cote\ pré-test \times RVN = cote\ post-test, test\ négatif$$

Le passage de la cote post-test à la probabilité post-test (Figure 4) s'effectue à l'aide la formule suivante :

$$\text{Probabilité post-test} = \text{cote post-test}/(\text{cote post-test}+1)$$

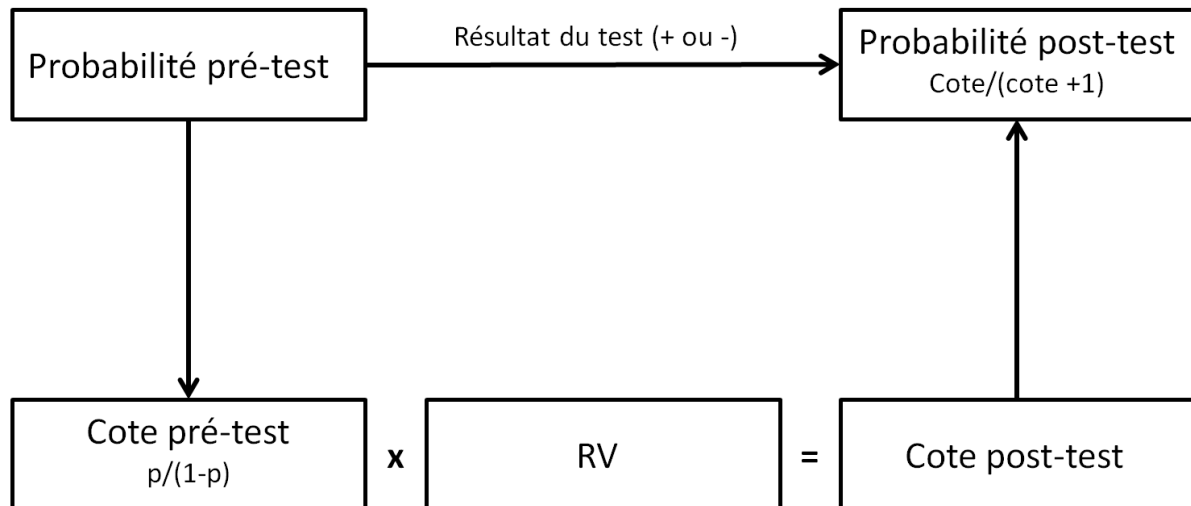


Figure 4 : Représentation schématique du calcul de la probabilité post-test.

Des RVP >10 et des RVN < 0,1 donnent généralement des variations importantes entre la probabilité pré-test et la probabilité post-test. Ils permettent au clinicien, dans le cadre du dépistage, d'orienter ou pas son patient vers des examens diagnostiques complémentaires afin de confirmer ou d'exclure la maladie recherchée.

A titre d'exemple, pour une probabilité pré-test d'une maladie donnée de 20%, la cote pré-test est égale à 0,20/0,80 (0,80 c'est 1-0,20) soit 0,25. Pour un RVP de 6, quand le test est positif, la cote post-test est égale à 1,5 (0,25x6). On obtient une probabilité post-test de 0,6 (1,5/1,5+1) c'est-à-dire 60%. Ainsi la probabilité de la maladie recherchée estimée à 20% chez un patient donné peut désormais être estimée à 60% après un test positif. Si le RVP avait été égale à 12, la probabilité post-test serait de 75%. Pour un RVP de 1,5, la probabilité post-test serait de 27%. Le calcul mathématique, assez fastidieux, de cette variation peut être grandement facilité par l'utilisation du nomogramme de Fagan (1975) (Figure 5). Pour cela, il faut prolonger sur l'axe de la probabilité post-test une droite passant par la probabilité pré-test et le RVP ou RVN.

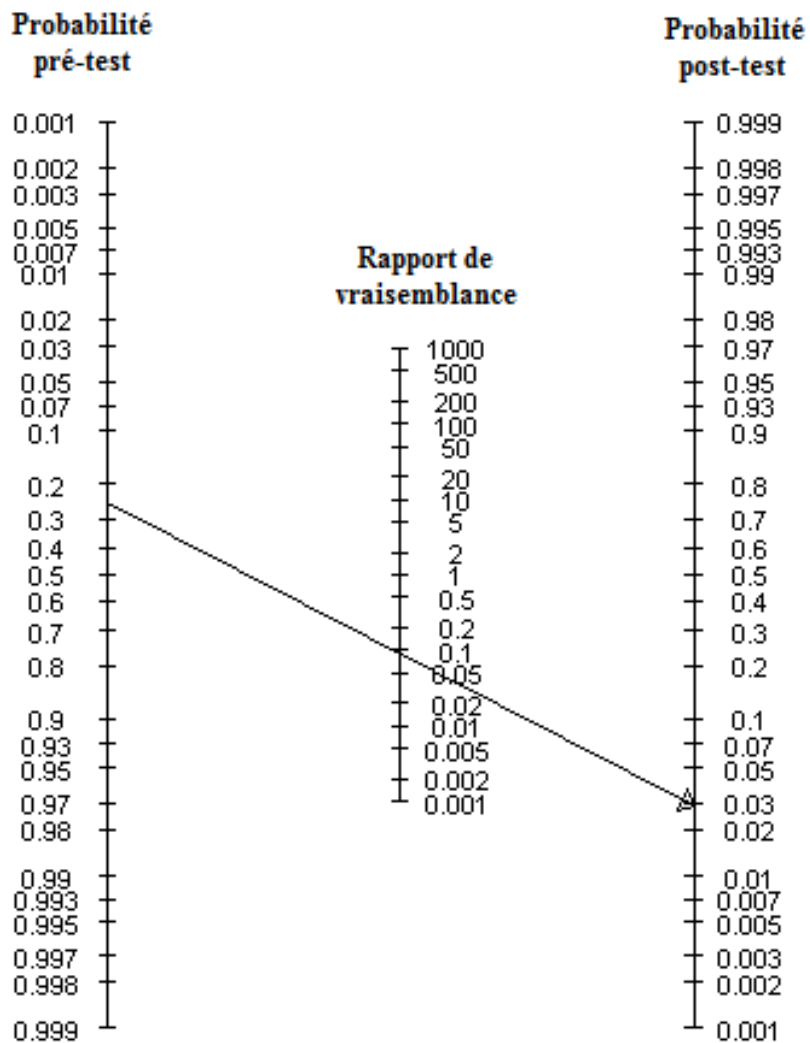


Figure 5 : Nomogramme de Fagan pour le calcul des probabilités post-test.

Le RV permet ainsi de modifier ou pas la croyance que nous avons sur la probabilité d’avoir la maladie avant la réalisation du test (probabilité pré-test) et une fois le test réalisé (probabilité post-test).

III. PROBLÈMES MÉTHODOLOGIQUES D'ÉVALUATION D'UN TEST

La qualité d'évaluation d'un test est déterminée par sa précision et sa validité. Toute étude portant sur l'évaluation de la performance d'un test est basée sur différentes mesures paramétriques inférées à partir d'un échantillon. Les indices de qualités d'un test s'expriment le plus souvent par des valeurs chiffrées (sensibilité, spécificité, VPP, VPN, RVP, RVN) qui sont des estimations à partir desquelles des décisions vont être prises. Il importe donc que ces estimations n'incluent pas des erreurs ou ne soient pas «biaisées».

Ces erreurs sont habituellement classées en 2 catégories : les erreurs aléatoires, et les erreurs systématiques (Choi et Pak, 2005).

La présence d'erreurs aléatoires augmente la variabilité des estimations et par conséquent diminue la précision de la mesure. Ce type d'erreurs peut être minimisé en augmentant la taille de l'échantillon qui va influencer la précision des résultats d'une étude (moyenne, sensibilité, spécificité, VPP, VPN, RVP, RVN, aire sous la courbe ROC...). Ces résultats doivent être exprimés avec des Intervalles de Confiance (IC) (Bulpit, 1987). L'IC à 95% est un intervalle de valeurs qui a 95% de chance de contenir la vraie valeur du résultat estimé. Il donne le degré d'incertitude de l'estimation de la mesure. Plus la taille de l'échantillon est grande, plus l'IC est étroit et plus l'estimation est précise.

La validité d'une étude peut être affectée par des erreurs systématiques (ou biais) entre la valeur de mesure d'un paramètre dans un échantillon et la vraie valeur dans la population. Les biais sont dus à la méthodologie de l'étude. S'ils ne sont pas pris en compte, ils peuvent fausser les résultats (sous ou surestimation). Il y a donc nécessité pour le clinicien qui souhaite valider et utiliser un test dans sa pratique quotidienne de bien connaître l'impact de ces biais au cours des différentes étapes d'évaluation d'un test.

On distingue deux types de validité :

La validité interne d'une étude permet de s'assurer que les résultats obtenus, grâce à l'analyse statistique de l'échantillon de sujets, sont proches de la vraie valeur dans la population d'intérêt. Elle fait référence à la « probabilité que les résultats s'approchent de la vérité ». (Horvath et Pewsner, 2004 ; Watine, 2004).

La validité externe correspond à la possibilité de généraliser les résultats de l'étude obtenus sur l'échantillon étudié à l'ensemble de la population cible, dans des contextes différents, et ce, sans que ses performances en soient modifiées (Irwing *et al.*, 2002 ; Watine, 2004). La validité externe dépend essentiellement des critères de définition de la pathologie étudiée, sa prévalence, le spectre des patients (caractéristiques démographiques, sévérité de la maladie, intensité des symptômes, co-morbidités) et des sujets contrôles inclus dans l'échantillon.

1. Les différents types de biais

Nous présentons ici les principaux biais pouvant affecter la validité interne d'une étude. La présence de ces biais est susceptible de sous-estimer ou de surestimer les performances d'un test (Horvath et Pewsner, 2004 ; Watine, 2004).

Biais de sélection : ce biais apparaît lors de la constitution de l'échantillon c'est-à-dire lors de la sélection des participants qui seront étudiés. Il dépend des critères d'inclusion et d'exclusion des sujets retenus dans l'étude et va se produire lorsque l'échantillon constitué n'est pas représentatif de la population cible à laquelle on veut extrapoler les résultats de l'étude. Cela se traduit par le fait que les individus sélectionnés dans l'échantillon étudié ont des caractéristiques différentes de ceux de l'ensemble de la population cible. A titre d'exemple, est-ce que les sujets inclus dans l'étude ont en moyenne le même âge que la population cible ? S'il existe un biais de sélection, le risque est de sous ou surreprésenter une catégorie de sujets dans l'échantillon (ex : les personnes suspectées de MA âgées de plus de 80 ans ou de moins 60 ans). On peut citer également la constitution d'un échantillon de témoins par le volontariat : les caractéristiques des sujets volontaires pour une étude peuvent être différentes de celles qui refusent d'y participer. Le recrutement des sujets par tirage au sort et la comparaison des caractéristiques des inclus et des non inclus peut permettre de maîtriser ce biais.

Biais de spectre : ce biais survient lorsque l'échantillon de patients ne couvre pas tout le spectre de la maladie. Il peut donc être considéré comme une forme de biais de sélection dans la mesure où l'échantillon étudié n'est pas représentatif de la population cible de patients chez lesquels un test pourra être utilisé en pratique clinique (Watine, 2004). Ainsi, l'inclusion de patients à un stade sévère d'une maladie pour évaluer un test peut conduire à surestimer ses capacités de détection de la maladie recherchée.

Imperfection du test de référence : pour évaluer le pouvoir discriminant d'un test expérimental (ou test index) les sujets doivent être classés comme étant malades ou non-malades à l'aide d'une procédure diagnostique de qualité (ou test de référence). Si le test de référence utilisé est imparfait, il va entraîner une classification erronée des sujets et biaiser la performance du test index. Idéalement, le même test de référence devrait être appliqué à tous les sujets participant à l'étude (Watine, 2004).

Biais de vérification partielle : pour classer un sujet comme étant « malade » ou « non-malade », il est nécessaire d'utiliser une procédure diagnostique (ou test de référence) ayant déjà fait ses preuves. Ce biais apparaît lorsque le résultat d'un test en cours d'évaluation influence la décision d'effectuer ou non le test de référence. Un biais de vérification partielle sera présent si les sujets dont le test index est positif sont classés comme malades, ou non-malades, par le test de référence, tandis que ceux dont le test est négatif sont considérés d'emblée comme non-malades sans qu'aucun test de référence ne soit effectué pour vérifier ce statut. Ce type de biais a pour effet de surestimer la sensibilité et de sous-estimer la spécificité du test étudié (Watine, 2004).

Biais de vérification différentielle : ce biais survient lorsque les sujets dont le test index est positif ne bénéficient pas du même test de référence pour être classés malades ou non-malades que ceux qui ont un test index négatif.

Biais d'incorporation : ce biais est présent si le résultat du test index est combiné à celui du test de référence pour classer les sujets comme étant malades ou non malades. Ce biais surestime l'exactitude du test index.

Biais de lecture: ce biais survient lorsque l'examineur qui réalise le test index connaît le résultat du test de référence ou vice versa entraînant une surestimation de la performance du test index.

Les biais qui viennent d'être décrits peuvent conduire à une distorsion de l'estimation de la performance d'un test. Cette distorsion peut se traduire par une surestimation ou une sous-estimation de ses performances. Les performances d'un test ne seraient donc pas forcément fixes et pourraient varier, de façon plus ou moins importante, en fonction des caractéristiques de la population étudiée. Cette variation des indices de performance a été définie à l'origine comme un biais spectral ou « *spectrum bias* » en anglais (Ransohoff et Feinstein, 1978). Si ce

biais spectral n'est pas évalué, l'extrapolation des résultats d'un test à une population cible risque d'être faussée et peut mener le clinicien à tirer des conclusions erronées.

2. De l'effet de spectre au biais spectral

Ransohoff et Feinstein (1978) sont les premiers à aborder l'importance de définir le spectre des sujets malades et des non malades comme condition préalable à l'évaluation de la qualité d'un test. Pour ces auteurs, les principaux facteurs qui peuvent influencer les performances d'un test sont le stade de la maladie, la durée et la sévérité des symptômes, les comorbidités et les données sociodémographiques.

Bien plus tard, Van Der Schouw *et al.*, (1995a, 1995b) mettent l'accent sur les conditions de validation d'un test qui ne sont pas toujours proches de ses conditions d'application en milieu clinique. En effet, le pouvoir discriminant d'un test est souvent évalué en confrontant les résultats obtenus par un groupe de patients déjà atteint par la maladie (les cas) à ceux d'un groupe constitué de volontaires sains (les contrôles) (Reid *et al.*, 1995 ; Whiting *et al.*, 2004 ; Rutjes *et al.*, 2005). Or, les études cas-contrôles ne permettent pas toujours de tirer des conclusions fiables sur l'intérêt d'un test. En effet, dans ce type de recherche, sont souvent comparés des patients porteurs d'une maladie à un stade avancé et des sujets sains (Boustani *et al.*, 2003) entraînant ainsi une surestimation du pouvoir discriminant d'un test (Lijmer *et al.*, 1999). Citons comme exemple, l'Antigène Carcino-Embryonnaire (ACE) utilisé comme test de dépistage des carcinomes (tumeurs développées à partir des cellules d'un épithélium). La sensibilité de l'ACE varie entre 36% pour les carcinomes aux stades Dukes A et B (carcinomes locaux, sans extension ganglionnaire, ni métastases) et 75%-83% pour les stades C (extension ganglionnaire) et D (métastases), respectivement (Fletcher, 1986).

Ces études cas-contrôles ne reflètent pas toujours les situations rencontrées en pratique clinique. Un test est avant tout utilisé pour différencier parmi un ensemble de patients partageant un certain nombre de symptômes ou de signes particuliers ceux qui ont ou pas la maladie recherchée (cadre du diagnostic) ou qui sont suspectées d'avoir la maladie ou pas (cadre du dépistage).

Hlatky *et al.*, (1984) ont aussi rapporté la variation du test de l'effort dans la détection de la maladie coronarienne selon sa sévérité (sensibilité de 48%, 68% et 85% pour une atteinte de 1, 2 ou 3 vaisseaux, respectivement) ou la présence typique ou atypique de symptômes telle

que la douleur d'angine de poitrine (sensibilité de 80% vs 53%). Plus récemment, (Muhlerin et Miller, 2002) ont montré que la sensibilité du dosage immuno-enzymatique des anticorps anti-chlamydia trachomatis (bactérie à l'origine d'une infection génitale sexuellement transmissible) pouvait varier selon l'âge avec une bien meilleure sensibilité chez les patients de plus de 24 ans (76% vs 58%).

Dans le domaine cognitif, le Mini-Mental State Examination (MMSE) a été élaboré par Folstein *et al.*, (1975) dans le but d'évaluer rapidement les capacités cognitives. Une valeur seuil fixe inférieure à 24 sur 30 est encore très souvent proposée pour détecter les troubles cognitifs liés à une démence avec une sensibilité et une spécificité de 87% et 82%, respectivement (Anthony *et al.*, 1982). Néanmoins, comme la plupart des tests cognitifs, le MMSE est sensible à l'âge et au niveau d'éducation. Crum (1993) a retrouvé une variation très importante du score total médian du MMSE de 25 chez les sujets les plus âgés (85 ans et plus) à 29 chez les plus jeunes (18-24 ans). Cette variation est encore plus nette en tenant compte du niveau d'études avec un score médian s'échelonnant de 22 pour les sujets ayant moins de 5 ans de scolarisation à 29 pour ceux ayant réalisé des études supérieures. En utilisant une valeur seuil < 24 pour le MMSE, O'Bryant *et al.*, (2008) ont montré ainsi qu'il existe une faible sensibilité (45%, soit un taux de faux négatifs de 55%) chez les patients de haut niveau socio-culturel (au moins 16 années de scolarisation) atteints d'un MCI ou d'une démence. Ces auteurs recommandent d'utiliser un score inférieur à 27 afin d'améliorer la sensibilité (69%) et de limiter ainsi le taux de faux négatifs (31%). Enfin, Flicker *et al.*, (1997) ont mis en évidence une importante variation de la spécificité (et donc du taux de faux positifs) lors de l'application de deux tests, le MMSE et l'Abbreviated Mental Test Score (AMT) (Hodkinson, 1972), sur des échantillons de patients issus, soit d'un service de consultation mémoire, soit d'un centre de gériatrie. Cette étude fait ressortir une augmentation du taux de faux positifs de 20% (consultation mémoire) à 30% (gériatrie) pour le MMSE et de 12% (consultation mémoire) à 47% (gériatrie) pour l'AMT.

Un autre exemple est donné par le test de provocation de la douleur élaboré par Mimori *et al.*, (1999) pour dépister les déchirures musculaires de l'épaule. Ces auteurs rapportent pour ce test une sensibilité parfaite de 100% et une spécificité de 90% dans un échantillon représenté par une très grande majorité d'hommes âgés entre 17 et 29 ans. Or, une autre recherche menée par Parentis *et al.*, (2006) réalisée sur un plus grand nombre de sujets (132 vs 32), appartenant à une tranche d'âge plus vaste de 15 à 71 ans et divisée de manière égale

entre hommes et femmes a montré une chute très nette de la sensibilité de ce test (17%) dans cet échantillon.

Enfin, selon certains auteurs (Fletcher, 1986 ; Begg, 1987) la sensibilité d'un test serait meilleure chez les patients symptomatiques plus à risque de présenter une maladie à un stade avancé. Par exemple, Lachs *et al.*, (1992) révèlent, pour un test de dépistage d'infection des voies urinaires, une diminution de la sensibilité de 92% à 56% lorsqu'il est appliqué à un groupe de patients présentant de nombreux symptômes et par conséquent une haute probabilité, à priori, d'avoir une infection urinaire puis à un groupe de patients très peu symptomatiques avec faible probabilité d'avoir l'infection. De la même manière, la sensibilité du test pour la détection du streptocoque bêta hémolytique du groupe A (responsable de l'angine, surtout chez l'enfant) est beaucoup moins élevée chez les cas avec une faible probabilité de streptocoques que ceux avec une plus forte probabilité (47% vs 90%) (Hall *et al.*, 2004) définie selon les critères cliniques du score de Centor (Centor *et al.*, 1981).

Dans la même perspective, il a été observé que le spectre des sujets non malades pouvait également affecter la spécificité d'un test.

Le syndrome de Cushing (lié à un excès de sécrétion du cortisol par les glandes surrénales) est difficile à diagnostiquer car certains signes évocateurs comme l'hypertension artérielle ne lui sont pas propres. Si le test de la suppression à la dexaméthasone a donné une très bonne sensibilité (95 à 98%) et une bonne spécificité 87% (Ayala *et al.*, 2003), cette dernière pouvait fortement varier (et diminuer) de 70% à 42% chez les sujets contrôles en présence de comorbidités telles que la dépression, la schizophrénie, la MA, les troubles obsessionnels compulsifs ou un éthyisme aigu. D'autres cas faux positifs étaient également observés chez les personnes de plus de 65 ans après une perte de poids, un manque de sommeil, un syndrome de malabsorption ou parce qu'elles étaient traitées par certains médicaments connus pour influencer le métabolisme hépatique de la dexaméthasone tels que les barbituriques, la carbamazépine et la phénytoïne (antiépileptiques) ou encore la rifampicine (antibiotique utilisé dans le traitement de la tuberculose).

Fletcher (2000) souligne de son côté qu'un taux élevé de l'antigène spécifique de la prostate (PSA) ne signifie pas automatiquement la présence d'un cancer puisque qu'il peut être observé également en cas d'hyperplasie (hypertrophie bénigne de la prostate) ou encore d'inflammation de la prostate et donner des cas faux positifs.

Enfin, il nous paraît important de préciser que les tests sont souvent validés dans des populations consultant en milieu hospitalier où la prévalence de la maladie recherchée est bien plus élevée que celle consultant en médecine générale. Leeftang *et al.* (2008a) précisent à cet effet qu'un taux de prévalence plus élevée de la maladie dans une population hospitalière peut potentiellement inclure une proportion plus importante de cas à un stade avancé de la maladie et donc augmenter la sensibilité d'un test.

Les résultats de ces nombreuses études confirment la thèse de Ransohoff et Feinstein (1978) en démontrant que la sensibilité et la spécificité d'un test ne sont pas toujours constantes. En fonction de certaines variables liées aux patients (âge, sexe, éducation...), à la maladie (intensité des symptômes, stade d'évolution, co-morbidités) ou au cadre de sélection des patients (qui va déterminer la prévalence de la maladie), la sensibilité et la spécificité d'un test peuvent être très différentes entre sous-groupes de patients et celles publiées dans une étude sur un échantillon global (Goerhing *et al.*, 2004).

Cependant, si certains facteurs sont susceptibles de faire varier la sensibilité et la spécificité d'un test, ils ne peuvent pas être considérés comme de véritable biais qui conduiraient à une mauvaise interprétation des résultats d'un test (Willis, 2008, Mulherin et Miller, 2002). L'âge, le sexe, le stade d'une maladie ou le contexte de soins étant considérés comme des caractéristiques naturelles liées à la population ou à la maladie. C'est pourquoi Mulherin et Miller (2002) suggèrent de remplacer le terme « biais spectral » par celui d'« effet de spectre » qui traduit le degré de variation de la performance d'un test lorsqu'il est appliqué à différents sous-groupes d'un échantillon en tenant compte d'un certain nombre de caractéristiques. L'auteur suggère pour limiter cet effet de spectre de sélectionner un échantillon suffisamment large afin de pouvoir inclure des patients présentant plusieurs caractéristiques (âge, stade de la maladie, co-morbidité...) et de comparer la performance des tests dans les différents sous-groupes ainsi constitués.

Par conséquent, avant d'intégrer un test dans sa pratique, tout clinicien devrait s'assurer que les caractéristiques des patients inclus dans l'échantillon d'étude soient les mêmes que celles des patients pour lesquels il souhaite utiliser ce test (Chevalier, 2011).

Cette démarche est nécessaire mais pas suffisante car un clinicien est amené à prendre des décisions dans un contexte clinique d'incertitude et doit donc également évaluer le risque de biais du test lorsqu'il l'utilisera dans sa pratique. Quelles sont les conséquences de cette

variation de la sensibilité et de la spécificité sur les RV et donc sur la probabilité post-test que le patient soit malade ou pas ? A titre d'exemple, Harper *et al.*, (2000) rapportent les résultats de 3 études indépendantes utilisant la même valeur seuil (test positif si > 21 mm de mercure) de mesure de la PIO (Pression Intra Oculaire) pour le dépistage du glaucome. Si les sensibilités étaient quasiment identiques d'une étude à l'autre (aux alentours de 50%), la spécificité variait de 49% (Daubs et Crik, 1980) à 84% (Tielsh *et al.*, 1991) et 97% (Harper et Reeves, 1999), correspondant respectivement à des RVP de 1 (le test n'apporte aucune information utile), 3,2 (variation faible de la probabilité post-test) et 17 (modification importante de la probabilité post-test). La faible spécificité retrouvée dans l'étude de Daubs et Crik (1980) s'explique probablement par le mode de sélection de la population contrôle issue d'un service hospitalier de consultation ophtalmologique qui comprend une proportion bien plus importante de personnes présentant une hypertonie oculaire que celle relevée dans la population générale (Harper *et al.*, 2000). Or, si l'hypertension oculaire est considérée comme un facteur de risque, elle ne signifie pas forcément que le patient est atteint d'un glaucome, ce qui peut augmenter considérablement le taux de faux positifs et donc diminuer la spécificité du test de mesure de la PIO. D'autres éléments, nombreux, ont été relevés comme étant susceptibles à un moment donné d'être responsable d'une augmentation de la PIO : le cycle circadien, l'épaisseur de la cornée, la pression artérielle systolique, l'âge, l'exercice physique avant l'examen, la prise de toxiques, le stress, la caféine ou encore le port de la cravate (Harper et Reeves, 1999).

Comme le montre cet exemple, l'effet de spectre peut entraîner un biais spectral, c'est-à-dire une modification pertinente de la probabilité post-test d'avoir un glaucome et ainsi constituer un élément pouvant contribuer de façon importante à la prise de décision du praticien.

Goerhing *et al.*, (2004) démontrent que ce biais spectral peut également être mis en évidence au sein d'une même population d'étude, lorsque la probabilité post-test d'avoir une maladie estimée à partir d'un échantillon global est très différente de celle retrouvée dans un sous-groupe de ce même échantillon. Ainsi, le test dipstick (Lachs *et al.*, 1992) permettant le dépistage d'infection des voies urinaires donne une sensibilité de 83% et une spécificité de 72% sur l'ensemble de l'échantillon étudié, ce qui donne un RVP de 2,86 et un RVN de 0,24. Or, la situation est différente si l'on divise cet échantillon en 2 sous-groupes définis par la probabilité élevée (présence de symptômes typiques) ou faible (présence de symptômes

atypiques) de présenter une infection urinaire. Dans le groupe à risque élevé, le même test, pour une même valeur seuil, donne une sensibilité de 92% et une spécificité de 43% et diminue ainsi le RVP de 2,86 à 1,59. Pour le groupe à faible risque, la sensibilité n'est plus que de 55% pour une spécificité de 78%, ce qui entraîne une augmentation péjorative du RVN de 0,24 à 0,57.

Un autre exemple est donné dans l'étude de Karakaya *et al.*, (2007) concernant l'utilisation du test oral de tolérance au glucose pour prédire le risque de développement d'un diabète. Les performances de ce test en termes de RV ont été analysées sur un échantillon de 196 patients. L'âge et la présence d'une hyperlipidémie avaient une influence significative sur le RVP qui variait de 2,28 à 7,77. Le RVN était directement influencée par le sexe, la consommation de tabac, l'hyperlipidémie, l'inactivité physique, le rapport taille/hanche et l'index de masse corporelle avec des valeurs oscillant entre 0,10 et 0,47.

On retrouve des résultats similaires pour l'ACME (Arrêter, Coupable, Matin, Ennuyé) ou CAGE en anglais (Cut-down, Annoyed, Guilty, Eyes-opener) (Ewing, 1984), un questionnaire rapide (4 questions) parmi les plus utilisés pour détecter les personnes qui présentent un problème de consommation d'alcool. Pour un score total ≥ 2 (considéré comme positif), l'étude de Steinbauer *et al.*, (1998) a rapporté une variation de la sensibilité (21% à 69%) et de la spécificité (75% à 96%) en fonction du sexe et de l'origine ethnique de la population américaine évaluée (population blanche, mexicaine ou africaine). Cet effet de spectre a entraîné une modification substantielle des valeurs du RV qui s'échelonnent de 1,91 à 9,35 pour le RVP et de 0,40 à 0,83 pour le RVN.

Ces données montrent que l'utilisation de la même valeur seuil d'un test (ou valeur seuil fixe) estimée sur un échantillon global, peut engendrer des informations erronées sur sa performance et son interprétation clinique, lorsqu'elle est appliquée à des sous-groupes d'individus ayant des caractéristiques particulières.

Enfin, il apparaît important de souligner qu'un effet de spectre n'engendre pas systématiquement un biais spectral dans la mesure où la sensibilité et la spécificité, qui varient en sens inverse, peuvent parfois ne pas modifier les RV dont le calcul dépend de ces deux indices (Goerhing *et al.*, 2004).

3. Les critères d'un dépistage appliqués aux troubles cognitifs légers dans le cadre des maladies neurodégénératives.

Les maladies neurodégénératives constituent un groupe de maladies altérant le cerveau (Flint Beal *et al.*, 2005). Ces maladies vont se traduire par une perte irréversible des neurones et entraîner un déclin progressif des fonctions cognitives, comportementales voire motrices. Au cours de leur évolution, certaines maladies neurodégénératives, comme la MA, peuvent évoluer vers un état de démence (McKhann *et al.*, 2011). La démence est un syndrome clinique qui se traduit par une détérioration de plusieurs fonctions cognitives suffisamment importantes pour affecter les activités de la vie quotidienne et entraîner une perte d'autonomie (American Psychiatric Association, 1994). La démence constitue un problème majeur de santé publique, ayant un impact important sur des millions de personnes affectées et leurs familles. Le diagnostic d'un syndrome démentiel ne préjuge pas de son étiologie ou de son évolutivité. Le Manuel Diagnostique et Statistique des Troubles Mentaux (American Psychiatric Association, 1994) a proposé une série de critères permettant d'opérationnaliser le diagnostic de démence : troubles de la mémoire et altération d'au moins une autre fonction cognitive (langage, fonctions exécutives, gnosie, praxie) suffisamment sévère pour retentir sur les activités de la vie sociale et professionnelle. La présence de ces troubles cognitifs et la perturbation des activités quotidiennes caractérisent l'état de démence défini comme la perte d'autonomie du sujet et de son indépendance. La démence peut aussi être accompagnée par des troubles du comportement comme l'apathie, la dépression, l'agitation, l'agressivité, l'errance, les idées délirantes ou l'euphorie (Benoît *et al.*, 2005).

La démence peut être liée à une atteinte cérébrale de nature vasculaire, dégénérative, infectieuse, traumatique ou tumorale. Certaines causes de démences sont potentiellement curables (tumeur cérébrale, méningite, hydrocéphalie, infection syphilitique, carence en vitamine B12) (Weytingh *et al.*, 1995) mais d'autres évoluent de façon irréversible comme la MA (de la Torre, 2010). Les causes principales de la démence non curable sont la MA, la démence d'origine vasculaire ou mixte (ex : MA à composante vasculaire), la maladie à corps de Lewy, la chorée de Huntington et la dégénérescence lobaire fronto-temporale notamment (Knopman *et al.*, 2003).

L'objet de ce chapitre n'est pas de détailler la clinique de toutes ces pathologies mais de discuter de la question de leur dépistage et de souligner les ambiguïtés ou les difficultés qu'il peut engendrer.

Pour des raisons pratiques, nous prendrons pour exemple la MA qui est la cause la plus fréquente de la démence (Lobo *et al.*, 2000). La MA se caractérise par un début insidieux et progressif ; typiquement, la mémoire est d'abord affectée. Avec l'évolution de la maladie, d'autres fonctions cognitives vont s'altérer comme l'orientation, le langage, les fonctions exécutives, visuo-spatiales et praxiques pour progresser vers un état de démence (McKhann *et al.*, 1984, McKhan *et al.*, 2011). En raison du vieillissement démographique, la prévalence de la MA et des maladies apparentées va augmenter de façon considérable au cours des prochaines années. Elle touche aujourd'hui 850 000 personnes en France avec, chaque année, environ 150 000 nouveaux cas et une prévision de 1,2 millions de personnes en 2020 (Ankri, 2006). Près d'un cas sur deux serait sous-diagnostiqué en France (Anaes, 2000). Ces maladies sont devenues une priorité nationale en raison de leur fréquence, de leur gravité, de leurs conséquences sur le malade et son entourage et de leur coût pour la société. Les mesures du plan Alzheimer 2008-2012 (Ministère de la Santé, de la Jeunesse, des Sports et de la Vie associative, 2007) ont en particulier fixé pour objectif de favoriser un diagnostic plus précoce et d'améliorer la prise en charge des patients et leurs aidants.

La nécessité de poser un diagnostic précoce de la MA pose la question, comme pour tout autre maladie, de la pertinence de son dépistage. Cette question du dépistage de la MA fait l'objet de débats depuis l'organisation de sa prise en charge dans le cadre des différents plans de santé. En effet, le premier rapport sur cette maladie (Girard et Canestri, 2000) soulève les problèmes éthiques qui consisteraient à vouloir réaliser un dépistage précoce de la MA en raison de difficultés à poser son diagnostic pour les formes débutantes et des possibilités thérapeutiques médicamenteuses limitées. Le rapport Gallez (2005) souligne qu'un dépistage systématique de la MA au stade précoce, comme c'est le cas par exemple pour les cancers ou les maladies infectieuses, n'est pas indiqué. Plus récemment, la Haute Autorité de Santé (2011) n'a pas recommandé le dépistage de la MA en population générale. Dans un communiqué de presse paru en 2011 (www.cnge.fr) le Collège National des Généralistes Enseignants (CNGE) s'est prononcé contre l'intérêt d'un dépistage de la MA et des maladies apparentées en population générale, et à plus forte raison des MCI en évoquant l'inefficacité des traitements sur la maladie à un stade avéré.

Dans ces différentes prises de positions, on voit bien la difficulté d'application des critères de l'OMS (Wilson et Jungner, 1970) pour un dépistage systématique de la MA. Selon l'OMS, le dépistage consiste à appliquer systématiquement un test à une population non symptomatique pour permettre un diagnostic et un traitement des patients atteints de la maladie. On réserve ce terme à des actions telles que le dépistage du cancer du sein, du côlon ou encore des maladies infectieuses. Ainsi, proposer un dépistage systématique de la MA dans une population parfaitement asymptomatique n'est pas d'actualité. La MA comprend probablement une longue phase asymptomatique durant laquelle les premières lésions cérébrales s'installent insidieusement, plusieurs années (10 ans ? 20 ans ?) avant l'apparition des premiers symptômes (Braak et Braak, 1995 ; Price et Morris, 1999). La réalisation d'une imagerie cérébrale, la recherche de l'apolipoprotéine E4 ou l'évaluation des modifications cognitives à l'aide d'une batterie de tests neuropsychologiques au cours du temps, constituent des examens qui présentent, certes, un intérêt dans le domaine de la recherche (Masur *et al.*, 1994 ; Elias *et al.*, 2000 ; Saxton *et al.*, 2004 ; Amieva *et al.*, 2008 ; Barnes *et al.*, 2009) mais qui manquent encore de fiabilité, sont coûteux et prennent du temps (Brayne, 2007).

De surcroît, l'on connaît encore mal l'efficacité des tests de dépistage disponibles appliqués à la population générale. Une étude canadienne (Patterson *et al.*, 1999) a pourtant montré que l'application systématique d'un test rapide et simple comme le MMSE chez des personnes non symptomatiques, âgées de 65 à 74 ans, donnait un taux de résultats faux positifs extrêmement élevé de 93%. Son utilisation dans une population de plus de 75 ans non institutionnalisée lors d'un examen annuel donnait 86% de faux positifs avec une valeur seuil unique de 26 (White *et al.*, 2002).

Un autre argument à l'encontre du dépistage systématique de masse est que les traitements pharmacologiques n'ont pas d'effets curatifs sur la MA. Néanmoins, si ces traitements ont un effet modeste sur la cognition (Courtney *et al.*, 2004), ils permettent une réduction significative du coût de leur prise en charge pour la société (Kiencke *et al.*, 2011). De plus, des études ont démontré que les interventions non pharmacologiques pouvaient permettre de différer le placement des malades en institution (Mittelman *et al.*, 1996) et de diminuer les troubles psychologiques ou comportementaux présentés par les patients et les aidants (Callahan *et al.*, 2006). Il s'agit donc d'arguments qui, bien que souvent débattus, constituent autant de preuves de l'intérêt d'une prise en charge précoce de la MA. Par ailleurs, les éléments développés dans la première partie de ce travail ont permis de préciser que le

dépistage ne se limitait pas à estimer la probabilité d'existence d'une maladie mais aussi d'une condition pré-morbide ou d'un facteur de risque. C'est la raison pour laquelle nous nous focaliserons sur le dépistage des troubles cognitifs légers.

Le concept de trouble cognitif léger ou Mild Cognitive Impairment (MCI) selon la dénomination internationale, a été proposé par Petersen (Petersen *et al.*, 1999 ; Petersen *et al.*, 2001) pour définir l'association d'une plainte cognitive (le plus souvent mnésique), d'un déficit d'une ou de plusieurs fonctions cognitives, contrastant avec une autonomie préservée dans la vie quotidienne. Les patients atteints de MCI présentent un risque plus élevé de développer ultérieurement une démence notamment liée à une MA. Ce risque est évalué entre 10 et 15% par an, alors qu'il n'est que de 1 à 2% chez les sujets sans MCI de même âge (Petersen *et al.*, 1999). Toutefois, le taux de conversion du MCI vers la démence varie considérablement (de 2 à 30%) selon les études et pose le problème des tests utilisés, des méthodes employées et des populations utilisées pour définir le MCI (Busse *et al.*, 2006 ; Brooks *et al.*, 2007).

Il est important également de préciser que le MCI est un syndrome clinique qui ne présuppose pas un mécanisme étiologique sous-jacent (Visser, 2006a). Il peut regrouper de nombreuses entités pathologiques (troubles anxieux, dépression, perturbations du sommeil, abus de substances, surdosage médicamenteux, troubles vasculaires, maladies neurodégénératives débutantes, pathologies métaboliques ou endocriniennes) qui peuvent altérer les capacités cognitives.

Le MCI est très souvent considéré, comme la phase symptomatique prodromale de la MA (Dubois, 2000). Or, si certains sujets avec MCI peuvent évoluer vers une MA, d'autres présenteront un état cognitif stable (Visser *et al.*, 2006b) voire même réversible. Des études ont en effet révélé que les performances cognitives de près de 50% des patients avec MCI pouvaient se normaliser après 1 an de suivi (De Rotrou *et al.*, 2005 ; Gauthier *et al.*, 2006).

L'hétérogénéité étiologique et évolutive du MCI pose donc un problème de pronostic et soulève la question de son diagnostic précoce. Le MCI peut constituer en effet, dans un certain nombre de cas, la phase prodromale de la MA. Certains auteurs (Petersen *et al.*, 1994 ; Tounsi *et al.*, 1999) estiment même qu'il est possible, à l'aide de tests neuropsychologiques appropriés, d'objectiver un profil typique de déficit mnésique isolé permettant de porter le diagnostic de MA à un stade préclinique. Cette démarche, qui consiste à vouloir rechercher et

identifier une MA comme étant à l'origine du MCI, a conduit à l'élaboration de nombreux outils dits de dépistage (Lonie *et al.*, 2010). Elle peut mener également à la confusion entre dépistage d'un syndrome cognitif et clinique comme le MCI et diagnostic précoce d'une maladie latente comme la MA (Dubois *et al.*, 2002).

D'autre part, à partir du moment où des symptômes sont exprimés par le sujet, (la plainte cognitive notamment mnésique faisant partie des critères de définition du MCI), on ne peut parler de dépistage du MCI en pratique clinique car ce terme est réservé aux personnes asymptomatiques (Wilson et Jungner, 1970). Le terme dépistage, comme l'a souligné le rapport Gallez (2005) est donc « *probablement mal adapté dans le cas de la maladie d'Alzheimer car il peut être mal compris* » (Gallez, 2005).

La plainte mnésique est fréquente et concernerait près de 50 % des personnes âgées de plus de 50 ans (Chehouilli, 2001). Elle traduit le sentiment d'une diminution de ses capacités mnésiques et peut générer ainsi, en particulier chez la personne âgée, la crainte de développer une MA. Or, les causes potentielles de troubles mnésiques sont très nombreuses et ne correspondent pas toujours à des états de MA. D'autres causes sont possibles comme par exemple la dépression, l'hypothyroïdie ou les perturbations du sommeil (Weytingh *et al.*, 1995) pour lesquelles nous disposons de traitement curatif ou de moyens de prévention. Dans ce contexte, il y a donc un intérêt, face à un sujet plaintif, à proposer plutôt le terme de « repérage » des troubles cognitifs car il ne préjuge pas d'une étiologie particulière.

Plusieurs outils ont été développés ces dernières années pour aider le clinicien à repérer les troubles cognitifs (Mitchell *et al.*, 2010a, Mitchell *et al.*, 2010b). Dans la pratique, l'un des tests les plus utilisés est le MMSE mis au point par Folstein *et al.*, (1975). Ce test composite comprend 30 items qui permettent d'explorer l'orientation temporelle et spatiale, l'attention, l'apprentissage, la mémoire, le calcul mental, le langage et les praxies visuo-constructives. Une valeur seuil < 24/30 est souvent utilisée pour indiquer la présence de troubles cognitifs liés à une démence. Anthony *et al.*, (1982) rapporte une sensibilité de 87% et une spécificité de 82%. D'autres auteurs comme O'Bryant *et al.* (2008) ont proposé d'utiliser une valeur seuil < 27/30 pour améliorer la précision du test chez les patients de très haut niveau d'éducation atteints d'une démence ou d'un MCI. Pour cette valeur, la sensibilité est de 89% (vs 66% pour un seuil < 24/30) et la spécificité de 91% (vs 99% pour un seuil < 24/30). Dans une autre étude (Murden *et al.*, 1991), les auteurs ont proposé d'abaisser ce seuil à 17/30 pour

les personnes ayant moins de 8 ans de scolarité. Le MMSE a une bonne fidélité test-retest (Galasko *et al.*, 1993) et inter-examineurs (O'Connor *et al.*, 1989). L'étude de Giordani *et al.*, (1990) a permis de mettre en évidence une bonne validité constructive du MMSE en retrouvant notamment une bonne corrélation ($r=0,89$) avec le quotient mnésique de l'Echelle Clinique de Wechsler (1945). Toutefois, une des limites du MMSE est son manque de sensibilité aux atteintes des fonctions exécutives. Une autre limite est que son interprétation porte souvent uniquement sur son score total. Dans son étude, Carcaillon *et al.*, (2009) ont montré que les sous-scores d'orientation temporelle et de rappel des 3 mots du MMSE pouvaient constituer des marqueurs efficaces de la présence de troubles mnésiques liés à la MA. Enfin, sa durée de passation (entre 10 et 15 minutes) peut être un obstacle à son utilisation routinière en médecine.

Milne *et al.*, (2008) ont comparé 11 épreuves à l'aide d'une échelle d'appréciation de la faisabilité, de l'aspect pratique et des qualités psychométriques. Trois outils ont obtenu les scores les plus élevés.

Le Memory Impairment Screen (MIS), a été développé par Buschke *et al.*, (1999) pour explorer la mémoire en moins de 5 minutes. Ce test est constitué d'un rappel libre et d'un rappel indicé. Il présente une forme parallèle qui permet le retest. Les auteurs ont mis en évidence pour un score seuil de 4/8, une bonne sensibilité (87%) et une très bonne spécificité (96%) pour la détection des troubles cognitifs de la démence (50 patients avec démence, 433 sujets contrôles). La VPP du MIS s'élevait à 71% pour une prévalence de la MA estimée à 10%. Le MIS présente en outre, une bonne fidélité test-retest ($r=0,69$) et une bonne validité constructive avec un taux d'accord tout à fait acceptable ($kappa=0,62$) avec le rappel total de l'épreuve d'apprentissage verbal de Grober et Buschke (1987). Toutefois, la sensibilité du MIS pour la détection des troubles cognitifs dans les formes légères de démence n'était plus que de 69%. L'introduction d'un rappel différé à 10 min. proposée par de Rotrou *et al.*, (2005) a permis d'augmenter la sensibilité du MIS à 75% pour les formes très légères de démence (i.e. patients ayant un score MMSE ≥ 26).

Le Mini-Cog est un test développé par Borson *et al.*, (2003). Il comprend une épreuve d'apprentissage de 3 mots avec rappel différé et un dessin de l'horloge. La sensibilité (76%) et la spécificité (88%) du Mini-Cog étaient similaires à celles du MMSE sur la même population d'étude (76 patients avec MA légère, 1103 sujets contrôles). Les VPP et VPN

n'ont pas été rapportées dans cette étude. Contrairement au MMSE, le Mini Cog est plus rapide (3 min.) et n'est pas affecté par le niveau d'éducation. Il serait, en plus, sensible aux troubles exécutifs contrairement au MMSE.

Le GPCOG (General Practitioner Assessment of Cognition) (Brodaty *et al.*, 2002) comporte deux parties. La première est remplie par le médecin traitant lorsqu'il interroge le patient en 9 points (orientation, dessin de l'horloge, rappel d'événements récents...). La deuxième partie est un interrogatoire d'un proche en 6 points explorant la capacité du patient à mémoriser, trouver des mots et exécuter les activités instrumentales de la vie quotidienne. C'est un test rapide (6 minutes), facile, acceptable pour le patient, sensible (82%) et spécifique (83%). La VPP était de 67% et la VPN de 92% pour une prévalence de la démence de 71% dans l'échantillon étudié (82 patients avec démence dont 50% de forme légère, 201 sujets contrôles). Le premier questionnaire a montré une bonne fidélité inter-examineur ($r=0,75$) alors que celui destiné au proche avait une fidélité plus modeste ($r=0,56$). Le GPCOG est très bien corrélé avec le MMSE ($r=0,68$). L'inconvénient du GPCOG est qu'il nécessite la présence d'un proche du patient, ce qui n'est pas systématique lors d'une consultation.

Ces trois outils retenus par Milne *et al.*, (2008) présentent de bonnes qualités métrologiques, mais ils ont été validés sur des échantillons de patients trop faibles pour pouvoir tenir compte du large spectre des pathologies démentielles.

En France, l'usage du test des 5 mots (Dubois *et al.*, 2002) est très répandu. Ce test évalue les capacités mnésiques. Il comporte une phase d'apprentissage, de contrôle de l'encodage, une phase de rappel libre, indicé si besoin, puis après une tâche interférente, une phase de rappel différé libre et indicé. Il est noté sur 10. Une étude de validation a été réalisée sur 86 patients atteints de MA et 126 se plaignant de difficultés de mémoire d'origine fonctionnelle (anxiété, dépression, plainte mnésique isolée). Un score inférieur à 10 permettait de détecter 91% des patients atteints de MA probable, avec une spécificité de 87%. Le score de rappel total chez les patients du groupe MA était significativement corrélé ($r=0,56$) au score total de l'épreuve d'apprentissage verbal de Grober et Buschke (1987). La très bonne spécificité de ce test est renforcée par la présence dans le groupe contrôle de sujets se plaignant de difficultés de mémoire d'origine fonctionnelle plus représentatifs des patients reçus en consultation que les sujets volontaires sains. Comme pour le MIS, une des limites de ce test est de n'évaluer

que la mémoire, dont les troubles sont fréquents au cours de la MA, mais qui ne sont pas forcément au premier plan dans les autres pathologies.

Un autre test validé en langue française, le CODEX (Belmin *et al.*, 2007), permet d'explorer en moins de 3 minutes plusieurs fonctions cognitives telles que la mémoire, les fonctions exécutives et les praxies constructives. A l'instar du Mini-Cog, le CODEX comprend une tâche de mémorisation de 3 mots et un dessin de l'horloge auxquels il a été adjoint 5 questions d'orientation spatiale. La sensibilité et la spécificité de ce test étaient de 92% et 85%, respectivement. L'échantillon de patients atteints de démence (n=143) était composé à 56% de MA, 35% de démence vasculaire ou mixte, 7% de démence à corps de Lewy et 5% de démence fronto-temporale. L'efficacité de cet outil doit néanmoins être nuancée compte tenu de la sévérité de la démence des patients inclus (MMSE moyen de 17,9) et du taux plus élevé de sujets de bas niveau d'éducation que dans le groupe des témoins (43% vs 21%).

Concernant les tests visant à détecter le MCI, on peut d'abord citer l'épreuve du MOCA (Montreal Cognitive Assessment) conçue par Nasredine *et al.* (2005). De même que le MMSE, le MOCA est un test composite qui évalue les fonctions cognitives suivantes : la mémoire et l'orientation, l'attention, les fonctions exécutives et d'abstraction, le calcul, le langage et les capacités visuoconstructives. Son temps de réalisation est d'environ 10 min. Le nombre de points maximum est de 30. Pour une valeur seuil < 26 , la sensibilité du MOCA était de 90% et sa spécificité de 87%, pour une VPP et une VPN de 89% et de 91%, respectivement. Malgré ces très bonnes performances, il faut cependant relever la faiblesse des échantillons à partir desquels elles ont été estimées puisque seuls 94 patients aMCI (dont 35% de MCIa multi-domaine) et 90 sujets contrôles volontaires sans plainte cognitive ont été inclus dans l'étude de Nasredine *et al.*, (2005). Le MOCA présente une bonne fidélité test-retest ($r=0.92$) et son score total est bien corrélé avec celui du MMSE ($r=0.87$). Par comparaison sur les mêmes échantillons d'étude, la sensibilité du MMSE était de seulement 18% et la spécificité de 100%.

Un autre test, le DemTect, a été mis au point par Kalbe *et al.* (2004). Cet outil, d'une durée de 8 à 12 min., comprend une épreuve d'apprentissage d'une liste de mots, un exercice de conversion de nombres en lettres (et de lettres en nombres), un test de fluence verbale sémantique et d'empans de chiffres envers. L'étude de validation incluait 97 patients MCI et

145 sujets contrôles. Le score du DemTect corrigé pour l'âge et le niveau d'éducation donnait une sensibilité de 80% et une spécificité de 92%, supérieures à celles retrouvées pour le MMSE (69% et 77%).

Plus récemment, Rami *et al.*, (2007) ont développé le Memory Alteration Test (M@T). De passation plus rapide (5 min.) que le MOCA et le DemTect, le M@T a été validé sur un échantillon composé de 50 patients MCIa issus de la consultation mémoire d'un l'hôpital de Barcelone et de 400 sujets contrôles volontaires, sans plainte cognitive, provenant de la clientèle de 6 cabinets de médecine générale implantés dans les différents quartiers de la ville afin de représenter l'ensemble des couches sociales de la population âgée de plus de 60 ans. Pour une valeur seuil < 37, la sensibilité globale de ce test était de 96% et la spécificité de 79%. Pour cette même valeur seuil, la spécificité était de 74% (soit 26% de faux positifs) pour les sujets de bas niveau scolaire (moins de 9 années d'études). Les auteurs ont donc proposé d'abaisser le seuil du M@T d'un point (< 36) afin de minimiser le taux de faux positifs (19% soit une spécificité de 81%) tout en gardant une sensibilité correcte (93%).

De la même manière, O'Bryant *et al.* (2008) ont proposé de modifier la valeur seuil du MMSE (< 24) considérée par Folstein *et al.* (1975) et encore aujourd'hui par de très nombreux cliniciens comme le seuil devant faire suspecter un déficit cognitif. Ces auteurs soulignent le manque de sensibilité (45%) de ce seuil pour détecter des troubles cognitifs chez des patients ayant un haut niveau d'études (307 avec démence et 176 avec MCI avec au moins 16 années de scolarisation). Ils recommandent d'élever cette valeur seuil à 27 (< 27) pour une meilleure sensibilité du MMSE (69%). Cependant, le calcul de la sensibilité du MMSE sur l'échantillon composé des seuls patients MCI (calcul non publié dans l'article de O'Bryant *et al.*) n'est que de 8% pour un seuil inférieur à 24 et 34% pour un seuil < 27.

Le test du cadran de l'horloge (Freedman *et al.*, 1994) présente l'avantage d'être peu influencé par le niveau d'éducation (Brodsky et Moore, 1997). Il est rapide (3 à 5 min.) et simple à administrer et permet d'explorer différentes fonctions cognitives parmi lesquelles les praxies visuoconstructives, l'attention, les capacités exécutives et les connaissances sémantiques. Un cercle avec un point en son centre représentant le cadran d'une horloge (ou d'une montre) est montré au patient à qui il est demandé de placer les chiffres des heures de 1 à 12, puis de dessiner les aiguilles pour indiquer une heure précise (ex : 10H10). Dans l'étude de Ehreke *et al.* (2011), différentes méthodes de cotation du test de l'horloge ont été

comparées afin d'évaluer celle qui permet de détecter avec le plus d'efficacité la présence d'un MCI dans un échantillon de 58 patients (vs 370 sujets contrôles) extrait de la cohorte de suivi LEILA 75+ (Leipzig Longitudinal Study of the Aged) (Riedel-Heller *et al.*, 2000). Cette étude montre que quelle que soit la méthode de cotation employée, le test de l'horloge s'avère peu efficace avec une sensibilité et une spécificité qui varie de 48% à 76% et de 49% à 79%, respectivement. De plus, la cotation est plus ou moins élaborée selon la méthode : de 3 points (Lin *et al.*, 2003) à 20 points (Mendez *et al.*, 1992), ce qui peut rendre l'interprétation de ce test plus ou moins longue et/ou difficile.

Une revue de la littérature effectuée par Lonie *et al.* (2010) a permis de mettre en évidence que sur les 15 tests retenus, quatre avaient des sensibilités et spécificités supérieures à 80%, alors que la valeur pronostique pour la conversion en démence avait été rapportée pour seulement deux d'entre eux. Dans la grande majorité des études rapportées dans cette revue (9 sur 15), l'effectif de cas MCI inclus était faible ($n < 100$) et par conséquent insuffisant pour apprécier l'impact de l'effet de spectre et du biais spectral sur les performances rapportées. La sensibilité et la spécificité de ces 15 tests ont été estimées à partir d'une valeur seuil fixe sans tenir compte notamment des variables sociodémographiques telles que l'âge ou le niveau d'éducation qui sont pourtant connus pour influencer les scores des tests neuropsychologiques utilisés pour la construction de normes (Crum *et al.*, 1993) ou la détermination de valeurs seuils (O'Bryant *et al.*, 2008). Enfin, pour 10 tests sur 15, la durée de passation totale excédait 5 minutes, fraction de temps exigé en général pour la détection rapide des troubles cognitifs (Brodsky *et al.*, 2006).

IV. OBJECTIFS DE LA RECHERCHE

A l'issue de cette analyse de la littérature, un certain nombre de constatations peuvent être mises en lumière.

La validation d'un test, comme nous l'avons vu, est une démarche complexe. Plusieurs travaux ont montré l'influence d'un très grand nombre de variables sur les performances de tests en mettant en évidence d'une part une variation de la sensibilité et/ou de la spécificité (ou effet de spectre), et d'autre part une modification éventuelle de la probabilité post-test de la maladie ou de l'anomalie recherchée (ou biais spectral) pouvant affecter la prise de décision clinique.

Ces dernières années, de nombreux tests ont été développés pour la détection précoce des troubles cognitifs. Ils ont pour but de mieux cibler notamment les patients atteints de pathologies neurodégénératives mais aussi les patients atteints d'un MCI qui présentent un risque accru de développer ces pathologies notamment une MA. Cependant, les données sur la généralisation des résultats de ces tests à d'autres populations restent limitées et posent deux questions fondamentales :

- La première question concerne les modalités d'application de ces tests dans les différentes populations concernées par les troubles cognitifs.
- La deuxième question relève des articulations que l'on peut établir entre les qualités métrologiques d'un test et son utilité clinique.

En effet, il n'existe aucune étude, à notre connaissance, concernant l'impact de l'effet de spectre et du biais spectral sur les performances d'un outil de détection des troubles cognitifs.

L'objectif de cette thèse était donc d'évaluer le degré de variation de la performance d'un outil de détection des troubles cognitifs légers et ses conséquences sur la prise de décision clinique

Pour atteindre ce but, plusieurs objectifs ont été fixés :

- Développer et valider un outil rapide de détection des troubles cognitifs en tenant compte de la diversité des maladies neurodégénératives.
- Estimer les qualités psychométriques de cet outil et poser les limites de son application clinique.
- Juger de son efficacité pour prédire l'évolution à court terme d'une cohorte de patients atteints d'un MCI.
- Déterminer l'impact de l'effet de spectre et du biais spectral sur la variation et l'interprétation des performances de l'outil étudié dans la détection du MCI amnésique.
- Proposer les bases cliniques d'une stratégie de repérage des troubles cognitifs en médecine.

L'outil étudié dans ce travail associait deux tests déjà connus : le MIS, élaboré par Buschke *et al.*, (1999), qui est un test d'exploration simple et rapide (4 min.) des capacités mnésiques antérogrades et le Set Test d'Isaacs (STI) (Isaacs et Kennie, 1972), une épreuve rapide (1 min.) d'évocation lexicale sémantique.

Cet outil est le fruit d'une longue réflexion menée en Franche-Comté dès 1998 par des neurologues, psychiatres, gériatres et neuropsychologues libéraux ou hospitaliers en collaboration avec l'ADNA (Association pour le Développement de la Neuropsychologie Appliquée), à l'origine de la création en 2002 du réseau Rapidfr (Vanholsbeeck *et al.*, 2007) et du logiciel RAPID qui est une base de données patients sur la MA et syndromes apparentés, destiné à l'ensemble des consultations mémoire, CMRR, et professionnels libéraux du Grand Est.

En effet, l'apparition des premiers traitements symptomatiques (Davis *et al.*, 1992) de la MA puis la création des premières consultations mémoires au sein de notre région vont inciter les cliniciens à modifier et structurer la prise en charge de cette pathologie qui s'est imposée progressivement depuis comme un véritable enjeu de santé publique.

Dans ce contexte, il était donc devenu primordial de développer une batterie de tests neuropsychologiques commune (batterie RAPID), normée et validée (Ferreira *et al.*, 2010) et

d'un logiciel permettant l'implémentation d'une base de données (Vanholsbeeck *et al.*, 2007) ayant pour objectif l'aide au diagnostic, l'échange entre les différents professionnels de la région et la recherche épidémiologique.

Parallèlement au développement de la batterie RAPID, l'objectif était d'élaborer également un outil rapide commun visant à aider le médecin à repérer les troubles cognitifs afin d'orienter éventuellement ces patients vers une consultation mémoire.

A la fin des années 90, le seul outil utilisé en Franche-Comté était le MMSE (Folstein *et al.*, 1975 ; Derouesné *et al.*, 1999) qui reste encore actuellement le plus répandu au monde. Toutefois, si la détection des troubles cognitifs a été considérablement améliorée par l'utilisation du MMSE, son manque de sensibilité aux troubles des fonctions exécutives, le caractère composite de son score total rendant son interprétation plus difficile qu'elle n'y paraît et sa durée de passation trop longue (entre 10 et 15 min.) constituaient des limites à l'examen cognitif routinier des nombreuses et diverses pathologies neurodégénératives rencontrées au cours de nos consultations.

A la même période, le MIS, un outil focalisé sur l'évaluation de la mémoire, mis au point par Buschke *et al.* (1999), a retenu notre attention du fait de sa réalisation plus simple et rapide (4 min.) que le MMSE mais aussi de sa bonne corrélation avec le score total de l'épreuve d'évaluation des capacités mnésiques de Grober et Buschke (1987) (la version française de ce test faisant partie intégrante de la batterie RAPID) qui s'est avéré être un bon prédicteur du déclin cognitif (Grober *et al.*, 2000).

Après avoir traduit le MIS en langue française (ou SMI, Score de Mémoire avec Indichage) (annexe 1), nous avons cherché dans un premier temps à déterminer les capacités du MIS à détecter les troubles cognitifs liés à une MA et à des pathologies neurodégénératives autres que la MA (Chopard *et al.*, 2003). Pour une valeur seuil fixe optimale < 6 , la sensibilité du MIS était de 91% pour les troubles cognitifs de la MA mais seulement de 53% pour les autres types de pathologies (maladie à corps de Lewy, démence vasculaire et dégénérescence fronto-temporale). Afin de compenser les limites de cet outil focalisé sur l'évaluation de la mémoire, il nous a semblé utile de compléter le MIS par la réalisation du test du cadran de l'horloge (Freedman *et al.*, 1994). Cette épreuve, simple et rapide à utiliser (2 min.), bien acceptée par les patients, permet d'explorer d'autres domaines cognitifs comme les capacités visuo-spatiales et les fonctions exécutives de planification (Royal *et al.*, 1998). Au cours de cette épreuve, il est demandé au patient de placer les chiffres des heures de 1 à 12 dans un

cercle dessiné sur une feuille. Ensuite, si les chiffres des heures sont bien placés, il doit dessiner deux aiguilles indiquant 10H10.

L'adjonction du test de l'horloge au MIS (combinaison MIS/Horloge) a permis d'améliorer considérablement la sensibilité du MIS utilisé de façon isolée (91% vs 53%) pour la population non MA, sans trop diminuer la spécificité (94% vs 100%). La performance de la combinaison MIS/Horloge étant même supérieure à celle du MMSE (sensibilité de 76% ; spécificité de 90%). Cependant, en dépit des bonnes performances rapportées avec ce couple de tests, celles-ci avaient été obtenues sur des patients atteints en grande majorité d'une forme modérée de démence. De plus, il a été démontré que le test de l'horloge avait une faible sensibilité pour déceler les troubles cognitifs des formes légères de démence (Lee *et al.*, 1996). Par ailleurs, il n'existe aucune méthode de cotation de ce test ayant fait l'objet d'un consensus sur le plan international. C'est pourquoi, nous avons décidé de substituer le test de l'horloge par le Set Test d'Isaacs (STI), une épreuve de fluence verbale catégorielle (annexe 2) dans laquelle le patient doit donner en 1 minute au total des noms de couleurs, d'animaux, de fruits et de villes (soit 15 sec. par catégorie). Tout comme le test de l'horloge, le STI permet d'explorer de nombreux domaines cognitifs tels que l'attention et la mémoire de travail (éviter les redites), la flexibilité cognitive (changement de catégorie), les stratégies de recherche dans le lexique, l'accès lexical sémantique, le langage (production orale) et la rapidité du traitement de l'information (le STI étant réalisé dans un temps limité). Le choix de ce test très rapide et simple à coter, a été conforté par les données de l'étude épidémiologique PAQUID (Personnes Agées quid) (Dartigues *et al.*, 1991). Cette étude menée depuis 1988 en Dordogne et en Gironde sur une cohorte de 3777 sujets âgés 65 ans et plus, a montré en effet que le score du STI était associé à un déclin cognitif 3 ans avant l'apparition d'une démence (Jacquin-Gadda *et al.*, 2000 ; Amieva *et al.*, 2005).

L'ensemble de ce travail a donné lieu à des résultats qui sont présentés ci-après sous la forme de trois articles originaux.

DEUXIÈME PARTIE: ÉTUDES CLINIQUES

I. ARTICLE 1: Combining the Memory Impairment Screen and the Isaacs Set Test: A practical Tool for Screening Dementias.

Article publié dans : *J Am Geriatr Soc* 2007;55(9):1426-1430. (IF=3,539).

Cette première étude visait à évaluer la capacité de la combinaison de deux tests déjà connus : le MIS (Busche *et al.*, 1999) et le STI (Isaacs et Kennie, 1973) ainsi que celle du MMSE (annexe 3) pour distinguer des patients atteints de troubles cognitifs liés à une maladie neurodégénérative de sujets contrôles sans troubles cognitifs.

Deux combinaisons de ces tests ont été évaluées : la somme pondérée (obtenue à l'aide d'un modèle de régression logistique) et une règle « et/ou » (combinaison positive s'il y a déficit à au moins l'un des 2 tests, combinaison négative si les 2 tests sont réussis). Les performances de cette combinaison de tests ont été comparées avec celles du MMSE sur le même échantillon d'étude.

L'étude a porté sur un échantillon de 444 patients (60-89 ans) suivis dans le réseau Rapid-Fr et de 311 sujets contrôles (60-89 ans) (annexe 4) sélectionnés par tirage au sort parmi la clientèle de 6 cabinets de médecine générale.

La règle « et/ou » donnait une sensibilité et une spécificité globale de 92% et de 90%, respectivement (92% et 92% pour la somme pondérée). La sensibilité du couple MIS/STI variait selon le type de pathologie (96% pour la maladie d'Alzheimer, stade léger ; 73% pour les autres pathologies, stade léger) et le niveau de sévérité de l'atteinte cognitive (98% pour les formes sévères de démence ; 85% pour les formes légères). Ces résultats ont été confirmés sur un échantillon indépendant de validation composé de 162 patients et de 105 sujets contrôles (annexe 5).

L'efficacité du couple MIS-STI était meilleure que le MMSE, le MIS ou le STI utilisés seuls. Cet outil est fidèle et présente une bonne validité constructive. Il est sensible à l'âge (MIS, STI) mais aussi au niveau scolaire (STI).

Cet outil, intéressant par sa rapidité (moins de 5 min.) et sa simplicité d'usage permet de détecter des troubles cognitifs liés à divers types de maladies neurodégénératives, dès le stade léger de la démence.

Son intérêt dans la détection des troubles cognitifs légers sans démence (ou MCI) reste à démontrer.

Combining the Memory Impairment Screen and the Isaacs Set Test: A Practical Tool for Screening Dementias

Gilles Chopard, MSc,*† Alexandre Pitard, PhD,† Sabrina Ferreira, MSc,*† Geraldine Vanholsbeeck, MSc,*† Lucien Rumbach, MD, PhD,*† and Jean Galmiche, MD†

OBJECTIVES: To assess the accuracy of combining the Memory Impairment Screen (MIS) with the Isaacs Set Test (IST) in distinguishing subjects with dementia from controls without dementia.

DESIGN: The weighted sum of MIS and IST scores was calculated, and the logical "or" rule combination was performed using two different patient and control samples. Combination results were compared with each test used alone and with Mini-Mental State Examination to detect all types of dementias.

SETTING: Database of the Regional Network for Diagnostic Aid and Management of Patients with Cognitive Impairment in the Franche-Comté, France, geographical area.

PARTICIPANTS: A cohort of 444 patients with Alzheimer's disease, non-Alzheimer's dementia, and reversible dementia and 311 controls.

MEASUREMENTS: Sensitivity, specificity, positive predictive value, and negative predictive value were calculated for the combination of these two tests.

RESULTS: The weighted sum had overall sensitivities of 0.92 and 0.84 for mild cases, with a specificity of 0.92. The "or" rule had overall sensitivities of 0.92 and 0.89 for mild cases, with a specificity of 0.90. This pairing requires no more than 5 minutes. Its scoring is simple with the "or" rule combination.

CONCLUSIONS: The combination of MIS and IST could be a useful tool for screening dementias. The "or" rule combination seems particularly adapted to medical practice. *J Am Geriatr Soc* 55:1426–1430, 2007.

Key words: dementia screening; Alzheimer's disease; non-Alzheimer's dementia

From the *Department of Neurology, Besançon University Hospital-Jean Minjoz, Besançon, France; †Rapid-Fr Network (Regional Network for Diagnostic Aid and Management of Patients with Cognitive Impairment in the Franche-Comté geographical area), Besançon, France.

Address correspondence to Dr. Jean Galmiche, ADNA Rapid-Fr Network, 6B Boulevard Diderot 25000 Besançon, France. E-mail: rapid@cegetel.rss.fr

DOI: 10.1111/j.1532-5415.2007.01271.x

Screening for dementia may be justified in some practical situations, such as with symptomatic patients,¹ for older people in whom the risk of therapeutics is possibly not beneficial (e.g., orthopedic surgery), or in some cases of mismanagement or exploitation.²

Rapidity and ease of scoring are important factors for consideration by practitioners.³ Screening tests should initially be validated in subjects with mild dementia³ and be efficient in screening all types of dementia.⁴ The detection of such a wide range of pathologies requires at least an assessment of episodic memory and executive function.³ The Memory Impairment Screen (MIS)⁵ and the Isaacs Set Test (IST)⁶ have each been used alone to screen for Alzheimer's disease (AD) and non-Alzheimer's dementia (NAD).^{5,7}

The goal of this study was to assess the usefulness of combining these two tests for the detection of cognitive impairment in all types of dementia from mild to severe. A logistic regression analysis and the simple logical "or" rule⁸ were used for the combination of these two tests and were compared with each test used alone and the Mini-Mental State Examination (MMSE).⁹

METHOD

Participants

The study was conducted from January to December 2004. The patient group consisted of 444 patients with all types of dementia, aged 60 to 96 and living at home, and who had been referred to local memory impairment consultation centers. Patients were consecutively recruited via the database of the Regional Network for Diagnostic Aid and Management of Patients with Cognitive Impairment in the Franche-Comté geographical area (Rapid-Fr network).¹⁰ The aim of this health network, financed by the French government, is to coordinate memory consultation in the geographical area of Franche-Comté. Exclusion criteria were craniocerebral trauma, progressive psychiatric illness, stroke within 3 months of the study, and acute neurological or somatic pathologies.

The control group consisted of 311 subjects aged 60 to 96 who were fully independent and did not have any

cognitive complaints. These subjects were randomly selected from the registers of six general practitioners' offices. They all met the same exclusion criteria as for the patient group. The ethics committee of Besançon University Hospital gave its official approval to conduct the study, and all participants gave informed consent for study inclusion.

A multidisciplinary team made the clinical diagnosis of dementia using neurological examination results, neuropsychological evaluation, blood analysis, and neuroradiological evaluation according to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV) diagnostic criteria.¹¹ The diagnosis of a specific dementing illness was made based on the National Institute of Neurological and Communicative Disorders and Stroke—Alzheimer's Disease and Related Disorders Association criteria for AD,¹² National Institute for Neurological Disorders and Stroke—Association Internationale pour la Recherche et l'Enseignement en Neurosciences criteria for vascular dementia (VD),¹³ criteria of the Consortium on Frontotemporal Dementia (FTD),¹⁴ and criteria of the Consortium on Dementia with Lewy Bodies (DLB).¹⁵ These diagnoses were assigned without prior knowledge of the score of the MIS, IST, or MMSE. Once dementia was diagnosed, the severity was determined according to the Clinical Dementia Rating scale (CDR).¹⁶ Subjects with questionable dementia were excluded from study.

Of the 444 patients, 421 had irreversible dementia, of whom 280 (66.5%) presented with AD, 100 (23.7%) with VD, 31 (7.4%) with DLB, and 10 (2.4%) with FTD. Twenty-three patients had reversible dementia (11 hydrocephalus under normal pressure, 5 sleep apnea, 3 thyroid dysfunction, 2 vitamin B₁₂ deficiency, and 2 cerebral tumors).

To validate the performance of the weighted sum and the logical "or" rule combination, each group was divided into patients with irreversible dementia and control samples. An original sample included two-thirds of the patients and two thirds of the controls, selected at random from the participant group. Then the scores of the screening tests validated on the original sample were applied to a cross-validation sample of the remaining one-third of the patients and one-third of the controls selected at random from the participant group.

Neuropsychological Examination

All patients and controls first underwent the two screening tests in the following order: MIS, then IST (Appendix 1), and then MMSE.

Subsequently, all controls and patients underwent a complete neuropsychological battery,¹⁰ including the following tests in the following order: Buschke's Cued Recall Test,¹⁷ the Trail-Making Test parts A and B,¹⁸ tests for verbally designating images and copying geometric figures as part of the cognitive evaluation battery,¹⁹ a scale for assessing instrumental activities of daily living,²⁰ the frontal behavior syndrome scale,²¹ and the DSM-IV depression scale.¹¹

Statistical Analysis

The Student *t*-test or chi-square test was used to compare the demographic variables of the patient and control samples. The influence of demographic variables on MIS and IST was investigated by calculating the Spearman rank correlation coefficient. Construct validity of MIS and IST,

alternate form reliability of the MIS, test-retest reliability of the IST, and the weighted sum were assessed using the Pearson correlation coefficient.

With the weighted sum rule, a logistic regression was performed to combine MIS and IST scores. A cutoff score of 0 was used to determine the degree to which the combination of these two tests scores discriminated between controls and patients with dementia. Individuals with a total MIS-IST score less than 0 were classified as not having dementia, and those assigned a value of 0 or higher were classified as having dementia.

With the logical "or" rule combination, the individuals were classified as having dementia if either MIS or IST score was positive (i.e., "abnormal") and as not having dementia if both tests scores were negative (i.e., "normal").⁸ Then optimal cutoffs, the scores at which the sum of sensitivity and specificity was at a maximum at the same time, were calculated by combining all possible scores for the MIS and the IST.

Sensitivity and specificity were calculated to determine the discriminative ability of the weighted sum, the "or" rule, the MIS, the IST, and the MMSE. The sensitivities of each test were compared using McNemar's test.⁸ Probability of dementia was estimated by calculating the positive predictive value (PPV) and the negative predictive value (NPV).²²

Receiver operating characteristic (ROC) analysis was used to examine the relationship between sensitivity and 1-minus specificity for each possible cutoff. The area under the ROC curve (AUC) was calculated to compare the overall accuracy of the tests.²³ Given that the "or" rule produces two scores, it would be impossible to construct a ROC curve from this type of combination.⁸ A chi-square test was used to compare the AUCs. $P < .05$ was considered statistically significant for all above analyses.

All the computations were performed with Stata Software (release 8.0, StataCorp, College Station, TX).

RESULTS

Participant Characteristics

The original sample consisted of 282 patients with irreversible dementia and 206 controls. The mean age \pm standard deviation was 76.2 ± 6.2 for the patient group and 75.3 ± 8.3 for the controls. There were no significant differences in age, sex, or level of education between patients and controls. The dementia distribution according to the level of severity was as follows; 116 (41%) had a CDR of 1, 108 (38%) had a CDR of 2, and 58 (21%) had a CDR of 3. One hundred eighty-seven (66.3%) presented with AD and 95 (33.7%) with NAD.

The cross-validation sample consisted of 139 patients with dementia and 105 controls. This sample was comparable in age ($P = .86$), sex ($P = .96$), and level of education ($P = .89$) with the original sample for the controls. It was also comparable in age ($P = .36$), sex ($P = .93$), and level of education ($P = .80$) for the patient group. Sixty-eight patients (49%) had a CDR of 1, 40 (29%) a CDR of 2, and 31 (22%) a CDR of 3. Ninety-three patients (66.9%) presented with AD and 46 (37.1%) with NAD.

Performance of the Weighted Sum Rule

A logistic regression analysis was used to predict dementia from MIS and IST scores. A logistic model (Log (P_i)/

Table 1. Sensitivity, Specificity, and Probability of Dementia of the Mini-Mental State Examination (MMSE), the Memory Impairment Screen (MIS), the Isaacs Set Test (IST), the "or" Rule and the Weighted Sum of the Original Sample

Variable	MMSE (0-30)	MIS (0-8)	IST	"or" Rule	Weighted Sum
Outpoint	<25	<7	<27	<6 or <25	0
Specificity	0.86	0.86	0.82	0.90	0.92
Sensitivity, all dementia	0.88	0.87	0.83	0.92*	0.92*
Positive predictive value [†]	0.25	0.24	0.19	0.32	0.37
Sensitivity, dementia and CDR score 1	0.79	0.83	0.75	0.89 [‡] *	0.84*
Sensitivity, dementia and CDR score 2	0.93	0.86	0.84	0.93 [§] *	0.96 [§] *
Sensitivity, dementia and CDR score 3	0.96	0.96	0.98	0.98	1.00
Sensitivity, all AD	0.93	0.97	0.83	0.96*	0.97*
Sensitivity, AD and CDR score 1	0.86	0.94	0.76	0.96 [‡] *	0.92*
Sensitivity, all NAD	0.79	0.67	0.84	0.85 [§]	0.83 [§]
Sensitivity, NAD and CDR score 1	0.65	0.59	0.73	0.73 [§]	0.67

*Sensitivity significantly greater than that of the IST.

[†]Based on a 5% prevalence of dementia.

[‡]Sensitivity significantly greater than that of the MMSE.

[§]Sensitivity significantly greater than that of the MIS.

CDR = Clinical Dementia Rating Scale; AD = Alzheimer's disease; NAD = non-Alzheimer's dementia.

$1 - P_i) = \beta_0 + \beta_1 \times \text{MIS}_i + \beta_2 \times \text{IST}_i$) was estimated using these two tests as the predictor variables, with P being the probability of having dementia, MIS the MIS score, and IST the IST score. This model was estimated on the participants in the original sample (282 patients with irreversible dementia and 206 controls). The natural logarithm (Log) of $P/(1 - P)$ was equal to the total MIS-IST score of the logistic regression formula below:

$$\text{Log}(P/1 - P) = 13.87 - 1.08 \times \text{MIS} - 0.26 \times \text{IST}$$

Or

$$P = 1/1 + e^{(13.87 - 1.08 \times \text{MIS} - 0.26 \times \text{IST})}$$

For example, for a subject scoring 4 on the MIS and 28 on the IST, the total MIS-IST score is 2.2. At this value, the probability, P , of having dementia is 90%.

For a test cutoff score of 0, the weighted sum had an overall sensitivity of 0.92, with a specificity of 0.92. At this cutoff score, the sensitivities of the weighted sum were significantly greater than or similar to those of the MIS, the IST, and the MMSE for discriminating all subgroups with dementia from controls (Table 1). At a 5% base rate of dementia, the PPV was 0.37 for the weighted sum, 0.24 for the MIS, 0.19 for the IST, and 0.25 for the MMSE. The NPV of the weighted sum was 0.99 at a 5% base rate.

In the group of patients with any dementia, the weighted sum had a significantly better AUC (0.971) than the MMSE (0.946), the MIS (0.922), or the IST (0.911) (Figure 1).

In the control subjects in the original sample, the weighted sum of total MIS-IST score was weakly correlated with age ($r = 0.24$, $P < .001$) and level of education ($r = -0.12$, $P < .01$) but not with sex ($r = -0.05$, $P = .53$). The reliability of the weighted sum was assessed using an interval range from 2 days to 2 months in an independent sample of 79 individuals (44 controls and 35 patients with dementia). The mean age of this sample was 75.6 ± 7.0 , 56% were female, and 20% had a high level of education. The overall test-retest reliability using the

predicted probability from the logistic regression analysis of the total MIS-IST score was $r = 0.87$ ($P < .001$).

In the cross-validation sample, sensitivities and specificity of the weighted sum rule were comparable with those calculated in the original sample.

Performance of the "or" Rule

The optimal cutoff scores for the "or" rule were MIS less than 6 or IST less than 25. These cutoffs were different from those when each test was used alone (<7 for the MIS and <27 for the IST). The optimal cutoff score was less than 25 for the MMSE.

At these optimal cutoff scores, the "or" rule had an overall sensitivity of 0.92 and a specificity of 0.90. When detecting mild AD, the sensitivity (0.96) was significantly greater than that of the MMSE (0.86) or the IST (0.76).

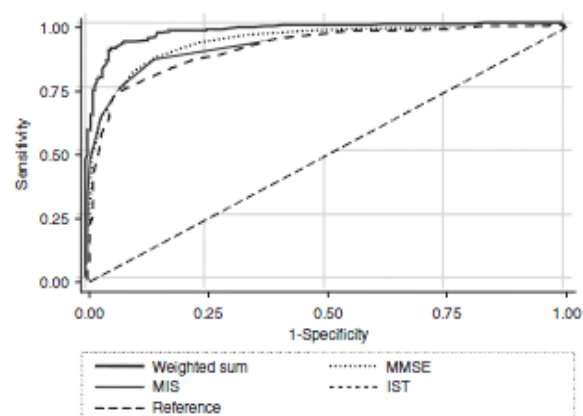


Figure 1. Receiver operating characteristic curves of the Mini-Mental State Examination (MMSE), the Memory Impairment Screen (MIS), the Isaacs Set Test (IST) and the weighted sum for all dementias of the original sample.

It was also significantly more sensitive than the MIS (0.73 vs 0.59) at detecting mild NAD (Table 1).

Of all the individuals who scored positive on both tests (MIS <6 and IST <25), 149 had dementia, of whom only three were controls, whereas of those scoring positive on only one test, 112 had dementia, of whom 18 were controls. These results suggest that individuals who failed on both tests had a high probability of having dementia.

In the cross-validation sample, sensitivities and specificity of the "or" rule combination were comparable with those calculated in the original sample.

In the control group, MIS score was correlated with age ($r = -0.20, P < .001$) but not with sex ($r = 0.03, P = .53$) or level of education ($r = 0.03, P = .53$). IST score was correlated with age ($r = -0.22, P < .001$) and level of education ($r = 0.30, P < .001$) but not with sex ($r = 0.13, P = .02$).

To minimize the practice effect, an alternate form of the MIS was developed. Parallel-form reliability was 0.78 ($P < .001$) for the MIS. Test-retest reliability was 0.91 ($P < .001$) for the IST.

To determine construct validity, a correlation analysis between the MIS and Buschke's Cued Recall Test and between the IST and Trail-Making Test Part B were done. The Pearson correlation coefficient was 0.70 ($P < .001$) and -0.76 ($P < .001$), respectively.

DISCUSSION

In this study, both methods (weighted sum and the "or" rule) used in combining MIS and the IST demonstrated positive results for discriminating patients with dementia (mild to severe stage) from control subjects. They had a high sensitivity (0.92) and a good specificity (0.92 for the weighted sum and 0.90 for the "or" rule). Sensitivity remained optimal to detect mild dementias (0.89) for the "or" rule and for the weighted sum (0.84).

These results demonstrate that the combination improves the sensitivity of the IST to detect mild AD and the MIS to detect mild NAD. The breakdown of dementia detected within the population studied was similar to the percentage observed in another French-speaking memory impairment consultation.²⁴ This pairing can therefore be considered efficient to detect all types of dementia when evaluated in a consultation population.

Administration time and simplicity of scoring are two main advantages of the acceptability of dementia screening.³ Scoring of the MIS and IST is objective and does not depend on the subjectivity of the examiner, which is why the tests were combined. The scoring of the combination can be performed rapidly with the "or" rule. Administration time was fixed, and no more than 5 minutes was required, whatever the type or severity of dementia to be screened.

Although these global results are of clinical interest, some findings may warrant further discussion. Although the MIS-IST pairing was effective in detecting mild dementia, better sensitivities were achieved when screening mild AD (0.96 for the "or" rule, 0.92 for the weighted sum) than mild NAD (0.73 for the "or" rule, 0.67 for the weighted sum). Two explanations for this discrepancy are possible. In contrast to AD, in which the first symptoms are usually cognitive,²⁵ the first symptoms of NAD are more often related to behavior.²⁶ This may explain why cognitive

evaluation is less effective in the early stages. Another possible explanation is that screening is easier when a population is well defined and homogeneous.²⁷ It is most likely the case in the AD population than the NAD one. Despite this moderate sensitivity to detect mild NAD, the "or" rule always had a sensitivity that was greater than or similar to that of each test used alone and the MMSE for all subgroups of dementia. Nevertheless, it allowed the detection of 19 of the 23 (sensitivity = 0.83) reversible dementias but was less sensitive for irreversible dementias (0.92).

These data also illustrate the difficulty of comparing results from different studies. A higher optimal cutoff was found not only for MIS alone (<7 vs <5), but also for the IST alone (<27 vs <25), than in a previous study.^{5,6} These discrepancies are possibly due to differences in patient populations investigated.³ This illustrates the limits of using single cutoffs and raises the question of using cutoff scores adjusted for age and educational level.³

Another question was to verify how these results can be translated at a low base rate of dementia. With a prevalence of 5%, both combinations had a modest PPV (Table 1). The PPV is an index not only strongly dependent on the prevalence of the disease in the population studied, but also primarily affected by the false-positive rate (i.e., 1-specificity). When the prevalence of the disease is low, PPV is always low, even if a test has good sensitivity and specificity,^{1,27} although some studies have reported relatively high PPVs at low prevalence of dementia.^{5,28} These results could be due to the low proportion of individuals with mild dementia included in the patient group³ or the small size of the dementia sample.^{5,28} The patients with moderate or severe dementia performed worse than mild cases, which led to a decrease in the cutoff score. Under these conditions, a test can correctly classify a greater proportion of controls, producing an increase in specificity and therefore an overestimation of its PPV. For example, a previously reported study⁵ found that the optimal cutoff score for the MIS alone was lower than that in the present study (<5 vs <7). These patients had a mean Blessed Information Memory and Concentration²⁹ test error score of 14.7, indicating moderate dementia severity. In addition, the subjects in that study had the same mean MIS score (2.5) as the patients of dementia group with CDR 3 in the current study. If the cutoff score of less than 5 had been used in the current study sample, the MIS would have had a higher specificity (0.97 vs 0.86) and PPV (0.52 vs 0.24) at a 5% base rate but a much lower sensitivity to detect all dementias (0.64 vs 0.87). As demonstrated, the properties of the same test may vary widely according to the size and severity of the dementia sample. As previously mentioned,²⁷ the MIS-IST pairing is not considered effective to predict dementias in a population with a low prevalence rate. Furthermore, 112 of the 130 individuals who failed only one test (86%) had dementia, whereas 149 of the 152 who failed both tests (98%) had dementia.

An important limitation of this study was the exclusion of questionable dementia. A common case-control approach (i.e., presence or absence of the disease) was used to validate the MIS-IST pairing. Questionable dementia was excluded for this purpose, because it has not been determined whether this population has mild dementia³⁰ or no dementia.²⁸ Nevertheless, care was taken to limit this bias by including a large proportion of mild cases (41%) in the dementia group and selecting the controls at random from

general practice registers and not from community-dwelling volunteers. More study is required before this approach can be recommended for clinical screening, because the decision to employ a case-control method is likely to exaggerate the sensitivity of these measures, which may not perform as well in clinical settings.

In summary, diagnosis of dementia requires a step-by-step procedure. Both combination modes in this study demonstrated a practical usefulness in a focused screening procedure. The "or" rule combination seems to be more adapted to screening in daily medical practice.

ACKNOWLEDGMENTS

The authors thank Richard Medeiros, Rouen University Hospital Medical Editor, for his valuable help in editing the manuscript and Professor Jacques Benichou, Department of Biostatistics, Rouen University Hospital, Rouen, France, for his constant expert advice.

Financial Disclosure: All authors confirm that there was no conflict of interest and that they have not received any payment for research, consultantships, or speakers forum, and do not have any company holdings or patents in connection with this manuscript.

Author Contributions: Gilles Chopard and Jean Galmiche: study concept and design, analysis and interpretation of data, preparation of manuscript. Alexandre Pitard: analysis and interpretation of data. Sabrina Ferreira, Geraldine Vanholsbeeck, and Lucien Rumbach: study concept and design, preparation of manuscript.

Sponsor's Role: None.

REFERENCES

1. Brodaty H, Clarke J, Ganguli M et al. Screening for cognitive impairment in general practice: Toward a consensus. *Alzheimer Dis Assoc Disord* 1998;12:1-13.
2. Kilada S, Gamaldo A, Grant EA et al. Brief screening tests for the diagnosis of dementia: Comparison with the Mini-Mental State Exam. *Alzheimer Dis Assoc Disord* 2005;19:8-16.
3. Lorentz WJ, Scanlan JM, Borson S. Brief screening tests for dementia. *Can J Psychiatry* 2002;47:723-732.
4. Meulen EFJ, Schmand B, van Campen JP et al. The seven minute screen: A neurocognitive screening test highly sensitive to various types of dementia. *J Neurol Neurosurg Psychiatry* 2004;75:700-705.
5. Buschke H, Kuslansky G, Katz M et al. Screening for dementia with the memory impairment screen. *Neurology* 1999;52:231-238.
6. Isaacs B, Kenne AT. The set test as an aid to the detection of dementia in old people. *Br J Psychiatry* 1973;123:467-470.
7. Jacquin-Gadde H, Fabrigoule C, Commenges D et al. A cognitive screening battery for dementia in the elderly. *J Clin Epidemiol* 2000;53:980-987.
8. Zhou XH, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. New York: John Wiley & Sons, 2002.
9. Folstein ME, Folstein SE, McHugh PR. "Mini mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatry Res* 1975;12:189-198.
10. Galmiche J, Chopard G, Ferreira S et al. Evaluation cognitive des demences par la batterie Rapid: Normes, valeurs seuil, seuils ponderes. *Rev Neurol* 2005;159:25171.
11. *Diagnostic and Statistical Manual of Mental Disorders*, 4th Ed. Washington, DC: American Psychiatric Association, 1994.
12. McKhann G, Drachman D, Folstein M et al. 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;34:939-944.
13. Pohjasvaara T, Nantyla R, Ylikoski R et al. Comparison of different clinical criteria (DSM-III, ADDTC, ICD-10, NINDS-AIREN, DMS-IV) for the diagnosis of vascular dementia. *National Institute of Neurological Disorders and Stroke Association-Association Internationale pour la Recherche et l'Enseignement en Neurosciences*. *Stroke* 2000;31:2952-2957.
14. The Lund and Manchester Group. Clinical and neuropathological criteria for frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 1994;57:416-418.
15. McKeith IG, Galasko D, Kosaka K et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies: Report of the consortium on DLB international Workshop. *Neurology* 1996;47:1113-1124.
16. Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* 1993;43:2412-2414.
17. Grober E, Buschke H. Genuine memory deficits in dementia. *Dev Neuropsych* 1987;3:13-36.
18. Reitan R. Validity of the Trail-Making Test as an indicator of organic brain damage. *Percept Mot Skills* 1958;8:271-276.
19. Signoret JL. *Evaluation des troubles de la memoire et des desordres cognitifs associes*. BEC 96. Paris: Ispen, 1989.
20. Lawton M, Brody E. Assessment of older people: Self maintaining and instrumental activities of daily living. *Gerontologist* 1969;9:179-186.
21. Lebert F, Pasquier F, Souliez L et al. Frontotemporal behavior scale. *Alzheimer Dis Assoc Disord* 1998;12:335-339.
22. Elwood RW. Clinical discriminations and neuropsychological tests: An appeal to Bayes' theorem. *Clin Neuropsychol* 1993;7:224-233.
23. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
24. Pasquier F, Lebert F, Petit H. Organisation des centres de la memoire et perspectives. *Rev Neurol* 1999;155:583-592.
25. Perry RJ, Hodges JR. Spectrum of memory dysfunction in degenerative disease. *Curr Opin Neurol* 1996;9:281-285.
26. Pasquier F, Delacourte A. Non-Alzheimer degenerative dementias. *Curr Opin Neurol* 1998;11:417-427.
27. Huppert FA. Guidelines for assessing the predictive value of neuropsychological measures. *Cortex* 2002;38:414-416.
28. Kuslansky G, Buschke H, Katz M et al. Screening for Alzheimer's disease: The memory impairment screen versus the conventional three-word memory test. *J Am Geriatr Soc* 2002;50:1086-1091.
29. Blessed G, Tomlinson R, Roth M. The association between quantitative measures of dementia and of senile changes in the cerebral gray matter of elderly subjects. *Br J Psychiatry* 1968;114:797-811.
30. Galvin JE, Roe CM, Powlishta KK et al. The AD8. A brief informant interview to detect dementia. *Neurology* 2005;65:559-564.

Appendix

The Memory Impairment Screen (MIS)

The MIS is a 4-minute, four-item delayed free- and cued-recall memory test.

1. The subject is informed that he or she would have to learn four words from different semantic categories.
2. The subject is asked to read the four words out loud.
3. The subject is asked to locate, point to, and read out the word once the examiner has given the semantic cue (e.g., "read out and point to the name of the animal").
4. The words are then hidden.
5. After an interfering task (counting backward from 100) lasting at least 2 minutes, the subject is invited to reproduce the words learned in any order over 20 seconds (free recall).
6. A cued recall is proposed for any words not mentioned using free recall: the examiner repeats the semantic cue (e.g., "What was the name of the animal?").

Grading: For each word, the score for a correct answer using free recall is 2 points and for cued recall, 1 point. The total MIS score (total MIS = (2 × free) + cued) ranged from 0 to 8.

The Isaacs Set Test (IST)

The IST is a 1-minute verbal category fluency task.

The subject has to orally produce as many words as possible for the following categories within 15 seconds: colors, animals, fruit, and cities.

Grading: The total number of items named produces the score (any repeated words or intrusions are not included).

II. **ARTICLE 2: Rapid Screening of Cognitive Change in Patients with Questionable Dementia Using the Memory Impairment Screen and the Isaacs Set Test.**

Article publié dans : *J Am Geriatr Soc* 2009;57(4):703-8. (IF=3,656).

Notre seconde étude avait pour objectif d'évaluer l'efficacité de l'utilisation combinée du MIS et du STI pour prédire l'évolution à court terme (6 à 24 mois) de patients présentant un MCI (ou Questionable Dementia).

Nous avons étudié rétrospectivement une cohorte de 106 patients âgés de 66 à 89 ans qui ont été suivis 15 mois en moyenne (6-24 mois). Durant ce suivi, 38 patients (36%) ont développé une démence légère (20 présentaient une MA, 5 une pathologie mixte, 5 une maladie à corps de Lewy, 4 une pathologie vasculaire, 1 une dégénérescence fronto-temporale et 3 un autre type de démence).

Un faible score à l'un ou l'autre des 2 tests était associé à un risque plus accru de conversion à court terme vers une démence. La règle « et/ou » avait une sensibilité de 74% (vs 76% pour le MMSE) et une spécificité de 81% (vs 40% pour le MMSE).

L'utilisation d'une valeur seuil fixe sans tenir compte de l'âge et du niveau scolaire constitue probablement une limite à l'efficacité et à l'interprétation de cet outil.

Rapid Screening of Cognitive Change in Patients with Questionable Dementia Using the Memory Impairment Screen and the Isaacs Set Test

Gilles Chopard, MSc,*† Geraldine Vanholsbeeck, MSc,*† Gregory Tio, MSc,† Alexandre Pitard, PhD,† Mickael Binetruy, MSc,† Lucien Rumbach, MD, PhD,*† and Jean Galmiche, MD*†

OBJECTIVES: To assess the efficacy of combining the Memory Impairment Screen (MIS) and the Isaacs Set Test (IST) in predicting short-term development of dementia in a group of people with questionable dementia (QD) at baseline.

DESIGN: Performances of the weighted sum of MIS and IST and the \llcorner or \lrcorner rule were compared with each other and with the Mini-Mental State Examination.

SETTING: Database of the Regional Network for Diagnostic Aid and Management of Patients with Cognitive Impairment in the Franche-Comté geographical area in France.

PARTICIPANTS: A cohort of 106 patients aged 65 and older with QD were followed up for a mean of 14.9 months (range 6–24 months).

MEASUREMENTS: Sensitivity, specificity, positive predictive value, and negative predictive value were calculated for the combination of these two tests.

RESULTS: The weighted sum had a sensitivity of 0.74 and a specificity of 0.84. The \llcorner or \lrcorner rule (MIS < 6 or IST < 25) had a sensitivity of 0.74 and a specificity of 0.81. When range values were applied, low scores on the MIS and the IST (MIS < 6 and IST < 25) led to a high probability of dementia, whereas high scores (MIS > 7 and IST > 29) suggested a high probability of remaining dementia-free in the study follow-up.

CONCLUSION: This quickly performed tool (5 minutes) is simple to use and score. When including cut-scores (MIS < 6 or IST < 25) or range values, this test could be considered a useful screening procedure for all types of dementias. *J Am Geriatr Soc* 2009.

Key words: dementia screening; questionable dementia

From the *Department of Neurology, Besançon University Hospital-Jean Minjoz, Besançon, France; and †Rapid-Fr Network (Regional Network for Diagnostic Aid and Management of Patients with Cognitive Impairment in the Franche-Comté Geographical area), Besançon, France.

Address correspondence to Jean Galmiche, Rapid-Fr Network, 26 rue de Vesoul, 25000 Besançon, France.
E-mail: adna@orange.fr

DOI: 10.1111/j.1532-5415.2008.02170.x

Early detection of dementia is important for optimal management and optimal outcomes for patients and their families.¹ Neuropsychological tests may be useful in the identification of cognitive impairment that occurs in patients at risk for developing dementia over a wide spectrum of pathologies.² These tests should be validated in comparison with a diagnostic reference standard and applied not only to patients with and without dementia, but also to those with questionable impairment.^{1,3} In patients with ambiguous impairment or questionable dementia (QD) it is more important to predict their cognitive outcome than to discriminate them from normal controls.⁴

An effective screening tool should initially be able to identify cognitive impairment of any etiology while being statistically reliable.³ This type of screening tool should also permit referring clinicians to better describe a patient's symptom profile after a brief consultation.³ Brief screening tests that assess memory and executive functions have demonstrated good performance in detecting mild dementia in a case-control approach.^{5,6}

In a previously reported study,⁶ the combination of the Memory Impairment Screen (MIS)⁷ and the Isaacs Set Test (IST),⁸ which assessed memory and executive functions in 5 minutes, was found to be more sensitive than the Mini-Mental State Examination (MMSE)⁹ in screening all types of dementia at the mild stage. The limitation of this case-control study could be that it did not include patients with QD.

The aim of this study was to assess the efficacy of combining the MIS and the IST in predicting short-term outcomes of patients with QD and to compare its performance with that of the MMSE.

METHOD

Participants

The study was conducted in a cohort of patients aged 66 to 89 who attended local memory impairment consultation centers between November 2003 and December 2007. Pa-

tients were consecutively recruited through the database of the Regional Network for Diagnostic Aid and Management of Patients with Cognitive Impairment in the Franche-Comté geographical area (Rapid-Fr network).⁶

A multidisciplinary team that conducted a neurological assessment, a neuropsychological evaluation, blood tests, and a neuroradiological analysis evaluated the clinical diagnosis of each patient. A diagnosis of QD was made based on the following criteria: presence of cognitive complaints that originated from the patients or their referent; impairment in one or more cognitive tests but insufficiently severe to interfere with daily activities assessed using four instrumental activities of daily living (IADLs) (ability to use the telephone, independence in transportation, self-administration of medication, ability to handle financial affairs);¹⁰ a Clinical Dementia Rating (CDR)¹¹ score of 0.5; and absence of the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM IV), criteria for dementia.¹²

For all patients with QD, the diagnosis at the follow-up visit was considered the main outcome. Conversion to dementia (converters) was defined according to the following criteria: progressive worsening of cognitive function at follow-up severe enough to affect IADLs¹² and progression of CDR score from 0.5 to 1. Standard criteria were used for diagnosis of Alzheimer's disease (AD),¹³ vascular dementia (VaD),¹⁴ dementia with Lewy bodies,¹⁵ and frontotemporal dementia.¹⁶ To avoid a diagnostic circularity, dementia was defined independently of the score results on the MIS, IST, and MMSE. Patients were considered not to have dementia (nonconverters) if their cognitive function remained stable or improved and they scored 0.5 or 0 on the CDR.

To limit the risk of including patients who may have had dementia at the baseline, only patients with QD who progressed to dementia at least 6 months after their baseline assessment¹⁷ and remained without dementia at least 12 months after baseline were enrolled. Patients who were followed up for more than 24 months were not included in this study. It was not considered feasible to attempt to predict long-term outcome based only on brief assessment tests.

Exclusion criteria were craniocerebral trauma, stroke within 3 months from the beginning of the study, acute neurological or somatic pathologies, or progressive psychiatric illness except for major depression. The ethics committee of Besançon University Hospital gave their official approval to conduct the study, and all participants gave their informed consent for study inclusion.

Neuropsychological Evaluation

All participants first underwent the two screening tests (MIS, IST; Appendix A) and then the MMSE. In addition, they all underwent a complete neuropsychological battery of tests, as described previously.⁶ Major depression was diagnosed using the DSM-IV depression scale.¹² All participants underwent the same neuropsychological battery at baseline and at follow-up.

Statistical Analysis

Continuous variables were compared using the Student *t*-test, and categorical data were analyzed using the chi-square test.

The ability of the combination of the MIS and the IST to predict the presence or absence of dementia (as a dichotomous outcome) was investigated using the weighted sum rule and the logical "or" rule.¹⁸

With the weighted sum rule, a logistic regression was performed to combine the MIS and IST scores, yielding a score on a new scale. The optimal cutoff score was derived from a receiver operating characteristic (ROC) analysis¹⁸ to discriminate between converters and nonconverters. Individuals with baseline QD with a total MIS-IST score less than the optimal cutoff were considered to be nonconverters, and those assigned a value equal or higher were considered as converters.

With the logical "or" rule combination, the optimal cutoffs were calculated by combining all possible scores for the MIS and the IST. Individuals with baseline QD were considered to be converters if the MIS or IST score was positive (abnormal) and nonconverters if both tests scores were negative (normal).

To maximize the accuracy of prediction, range values were also used. In this procedure, two combinations of MIS and IST were calculated so as to have no false positives and no false negatives, respectively.

Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)¹⁸ of the MIS-IST combination and MMSE were calculated. Sensitivities and specificities were compared using the McNemar test.¹⁸ The area under the ROC curves¹⁸ was used to measure the overall accuracy of discrimination between converters and nonconverters for each screening tests. *P* < .05 was considered statistically significant for all above analyses.

All computations were performed using Stata software release 8.0 (Stata Corp, College Station, TX).

RESULTS

Participant Characteristics

During the study period, 362 patients with QD were consecutively recruited through the database of the Rapid-Fr Network. Of these, 212 (59%) were not followed up and 150 (41%) were followed up. A cohort of 106 patients were followed up for 6 to 24 months (mean 14.9 ± 4.5). There was no significant difference in age, sex, or depression rate between patients with QD who were followed and those who were not. The patients with QD who were followed were less educated and had better memory performance on the Buschke Cued Recall Test (BCRT).¹⁹ They did not differ on any other neuropsychological tests (Table 1).

To ensure that patients with QD who were followed were within the normal and dementia categories, their cognitive performance was compared with that of 266 patients with cognitive complaints but no dementia (CDR = 0) and those of 207 patients with mild dementia (CDR = 1) who attended the local memory impairment consultation centers over the same study period. The patients with QD performed significantly lower than the group with no dementia and significantly better than the group with mild dementia on all neuropsychological tests (Table 1).

To ensure that differences in the scores on neuropsychological tests did not reflect the older age of the group with mild dementia, a sample that included 90% of the patients was selected at random from this group to match

Table 1. Baseline Characteristics of the Patients According to Diagnosis

Characteristic	QD (CDR = 0.5)			
	Followed Up n = 106	Not Followed Up n = 212	No Dementia (CDR = 0) n = 266	Mild Dementia (CDR = 1) n = 207
Age, mean ± SD	75.7 ± 5.0	75.4 ± 5.4	75.5 ± 5.5	77.6 ± 5.3
Male, n (%)	44 (41.5)	74 (34.9)	112 (42.1)	85 (41.1)
Education ≥12 years, n (%)	9 (8.5) ^{*†}	37 (17.6)	66 (24.9)	30 (14.6)
Depression, n (%)	11 (10.4) [‡]	25 (11.8)	15 (5.6)	5 (2.4)
Memory Impairment Screen score, mean ± SD (range 0–8)	6.7 ± 1.3 ^{*†‡}	6.3 ± 1.7	7.4 ± 0.9	4.8 ± 2.0
Isaacs Set Test score, mean ± SD	28.1 ± 5.7 ^{†‡}	27.2 ± 5.8	32.6 ± 6.0	24.3 ± 5.8
Mini-Mental State Examination score, mean ± SD (range 0–30)	25.2 ± 2.7 ^{†‡}	25.0 ± 2.2	27.1 ± 2.0	23.1 ± 2.3
Buschke Cued Recall Test score, mean ± SD (range 0–48)				
Free recall	16.8 ± 6.1 ^{*†‡}	15.4 ± 5.6	23.8 ± 5.7	9.4 ± 4.5
Total recall	40.9 ± 5.1 ^{*†‡}	38.4 ± 5.9	44.9 ± 2.5	29.4 ± 7.9
Trail-Making Test score, mean ± SD				
Part A (range 0–150)	72.4 ± 31.7 ^{†‡}	75.7 ± 34.6	52.1 ± 20.2	91.5 ± 39.3
Part B (range 0–300)	241.8 ± 79.7 ^{†‡}	246.5 ± 76.2	156.7 ± 63.4	269.5 ± 60.3
Picture naming score, mean ± SD (range 0–30)	28.4 ± 2.2 ^{†‡}	28.2 ± 2.2	29.3 ± 1.3	27.1 ± 3.1
Copy score, mean ± SD (range 0–6)	5.4 ± 1.0 ^{†‡}	5.5 ± 1.0	5.8 ± 0.5	5.0 ± 1.6

*Significantly different from the group with questionable dementia (QD) that was lost to follow-up.

†Significantly lower than the group without dementia.

‡Significantly higher than the group with mild dementia.

SD = standard deviation.

by age the group of patients with QD that was followed. The results showed the same higher performances for patients with QD as the mild dementia group.

Of the 106 patients with QD examined at follow-up, 38 (36%) converted to dementia at a mild stage (CDR = 1), and 68 remained free of dementia. Of the converter patients, 20 had AD, five had mixed dementia (AD and VaD), five had dementia with Lewy bodies, four had VaD, one had frontotemporal dementia, and three had other dementia. Of the nonconverters, 57 still had QD (CDR = 0.5), and 11 had reverted to normal (CDR = 0) at follow-up.

Demographic and cognitive characteristics of converters and nonconverters are presented in Table 2. The age ($P = .24$), the sex ratio ($P = .61$), the level of education ($P = .57$) and the frequency of depression ($P = .46$) at baseline were similar between groups. The converter group performed worse than the nonconverter group on the MIS ($P < .001$), the IST ($P < .001$), and the free recall ($P < .001$) and total recall ($P < .038$) of the BCRT. Performances of the two groups did not differ on any of the neuropsychological tests (Table 2).

Performance of the Weighted Sum Rule

The baseline scores of the MIS and the IST of the 38 converters and the 68 nonconverters were used as predictors. The logistic regression model for this combination was as follows:

$$\text{Log}(P/1 - P) = 12.91 - 1.19 \times \text{MIS} - 0.20 \times \text{IST}$$

Where P was the probability of patients converted to dementia, MIS was the MIS score, and IST was the IST score.

A test cutoff score of -0.23 was considered to be the optimal cutoff point. The sensitivity (i.e., the number of converters with a total MIS-IST score ≥ -0.23) and the

specificity (i.e., the number of nonconverters with a total MIS-IST score < -0.23) of the weighted sum were 0.74 and 0.84, respectively. At the prevalence rate of 25%, the PPV and the NPV were 0.61 and 0.91, respectively. The area under the ROC curve for this model (0.857) was significantly greater than the MMSE (0.590) ($P < .001$).

Performance of the "Or" Rule

Combining a cutoff score of less than 6 on the MIS and less than 25 on the IST produced the optimal combination of sensitivity (0.74) and specificity (0.81). The sensitivity of the "or" rule was similar to that of the MMSE, using an optimal cutoff score of less than 27 (0.74 vs 0.76) ($P = .80$), and the specificity was significantly greater (0.81 vs 0.40) ($P < .001$).

When range values were applied, of all the patients who scored positive on both tests (MIS < 6 and IST < 25), five were converters and none were nonconverters, whereas of those scoring higher on both tests (MIS > 7 and IST > 29), 11 were nonconverters and none were converters.

DISCUSSION

This study was conducted to assess the accuracy of the MIS-IST combination to discriminate between patients with QD, those who converted to dementia from those who remained free of dementia, over a short period of time. The results indicate that the performances of the weighted sum rule were similar to those of the "or" rule in terms of sensitivity (0.74 for both combinations) and specificity (0.84 and 0.81, respectively). Whatever the method used for combining the MIS and the IST, it was found more effective than the MMSE.

Table 2. Characteristics of Nonconverters and Converters with Questionable Dementia (QD) at Baseline and Follow-Up

Characteristic	Nonconverters* n = 68		Converters [†] n = 38	
	Baseline	Follow-Up	Baseline	Follow-Up
Age, mean ± SD	75.2 ± 5.0	76.4 ± 5.0	76.4 ± 4.8	77.3 ± 4.7
Male, n (%)	27 (39.7)	27 (39.7)	17 (44.7)	17 (44.7)
Education ≥12 years, n (%)	5 (7.4)	5 (7.4)	4 (10.5)	4 (10.5)
Depression, n (%)	7 (10.3)	2 (2.9)	4 (10.5)	2 (5.3)
Weighted sum, mean ± SD	-1.6 ± 1.4 [‡]	-1.8 ± 1.9 [§]	0.8 ± 1.9 [‡]	2.2 ± 2.4
Memory Impairment Screen score, mean ± SD (range 0–8)	7.2 ± 0.9 [‡]	7.3 ± 0.9 [§]	5.8 ± 1.5	5.1 ± 1.9
Isaacs Set Test score, mean ± SD	29.4 ± 5.3 [‡]	30.0 ± 6.0 [§]	25.8 ± 5.7	23.2 ± 5.8
Mini-Mental State Examination score, mean ± SD (range 0–30)	25.5 ± 2.7	25.4 ± 2.5 [§]	24.6 ± 2.8 [‡]	22.2 ± 3.4
Buschke Cued Recall Test score, mean ± SD (range 0–48)				
Free recall	18.4 ± 6.1 [‡]	19.2 ± 6.3 [§]	14.1 ± 5.1 [‡]	9.2 ± 5.6
Total recall	41.6 ± 4.8 [‡]	42.4 ± 5.4 [§]	39.5 ± 5.4 [‡]	32.1 ± 7.4
Trail-Making Test score, mean ± SD				
Part A (range 0–150)	70.1 ± 33.1	71.9 ± 31.9 [§]	76.7 ± 29.1 [‡]	96.7 ± 39.1
Part B (range 0–300)	232.9 ± 83.2	230.5 ± 81.2 [§]	257.6 ± 71.3	279.4 ± 47.9
Picture naming score, mean ± SD (range 0–30)	28.5 ± 1.9	28.8 ± 1.6 [§]	28.2 ± 2.6	27.1 ± 3.0
Copy score, mean ± SD (range 0–6)	5.6 ± 0.9	5.6 ± 1.0 [§]	5.2 ± 1.3	4.7 ± 1.9
Instrumental activity of daily living score, mean ± SD (range 0–4)	0.2 ± 0.5	0.2 ± 0.5 [§]	0.5 ± 0.8 [‡]	1.4 ± 1.1

* Patients with QD who did not convert to dementia during the 6 to 24 months of follow-up.

[†] Patients with QD who converted to dementia during the 6 to 24 months of follow-up.

[‡] Significant difference between nonconverters and converters at baseline.

[§] Significant difference between nonconverters and converters at follow-up.

[‡] Significant difference between baseline and follow-up.

SD = standard deviation.

All patients with QD were followed-up with after a short period. One of the problems related to the short mean time until follow-up was that the patients with QD who converted to dementia in less than 12 months may have had dementia at baseline. This could have led to an overestimation of the sensitivity and the NPV,²⁰ although it was verified that patients with QD who converted rapidly to dementia had similar cognitive performance at baseline to that of those who converted at least 12 months after the follow-up period. Otherwise, it was also verified that the entire converter group had better cognitive performance at baseline than the group with mild dementia. Last, because a short version of the IADL was used, not all daily activities could be assessed with accuracy. Nevertheless, the results demonstrated that the baseline mean IADL score of the converters was similar to that of the nonconverters (Table 2).

Another problem was that some of the patients with QD who were classified as nonconverters, might have developed dementia after a longer follow-up. This could have led to an underestimation of the specificity and the PPV.²⁰ For example, if patients with QD who were classified as nonconverters but were followed-up with in less than 18 months had been excluded from this study, the specificity of the MIS-IST pairing would have increased from 0.81 to 0.88 for the "or" rule and from 0.84 to 1.00 for the weighted sum rule. Moreover, 11 of the nonconverters, reverted to normal, and for this subgroup of patients, the specificity was better than for those who remained QD (0.91 vs 0.79).

Weighted sum based on logistic regression is often used as a method of combining two or more tests scores²¹ but requires complex calculations difficult to use in general daily practice.²¹ In contrast, the logical "or" rule is a simpler method.^{6,18,21} An interesting result of this study was that the optimal cutoff scores (MIS < 6 or IST < 25) found with the "or" rule for discriminating between converters and nonconverters were similar to those used in a previous study for screening mild dementia.⁶ This appears to be important, because for routine use in clinical practice, a screening test must be quick to administer but also simple to use and score.¹

For these reasons, the MIS-IST combination seems to us more practical than other screenings tests such as the MMSE,⁹ the Montreal Cognitive Assessment,²² or the Demenz-Detektion,²³ on which clinicians have to separately analyze subscores in addition to the total composite score.

Also interesting is that non-AD was not excluded from the primary outcome.²⁴ The MIS-IST pairing, by covering episodic verbal memory and executive function, is able to detect signs of cognitive impairment that may be observed in various types of dementia and not only in AD.³ Last, depressive patients were not excluded because this is a factor that may be associated with early dementia.²⁵

QD has been defined as a broad category of patients neither clearly demented nor healthy.¹¹ QD and Mild cognitive impairment (MCI) overlap.²⁶ The term QD was chosen, because the study population was recruited from memory consultation centers and not from a community

sample. Patients presenting to a memory consultation are at high risk of conversion to dementia. This may explain the high conversion rate (36%) found in this study. Nevertheless, 10% of the patients with QD reverted to normal, confirming the heterogeneity of this population.⁴

The main limitation of this study was the low proportion (29%) of follow-up of the QD population recruited in the Rapid-Fr network. Twelve percent were followed-up with before 6 months or after 24 months and were therefore excluded from this study. There were 59% of patients who did not return to the memory consultation, although in the Rapid-Fr network, these patients were advised that it was required that they be reexamined every 6 months or at least once a year. It could not be verified, for ethical reasons, why these patients did not return. All of the patients with QD who were followed-up with had a lower level of education than those who were not. Both groups had similar cognitive performances on all tests except for memory, on which the patients with QD who were followed-up with were better. They did not differ significantly in terms of age, sex, or depression rate (Table 1). It is unlikely that these differences could have contributed to any overestimation of the results.

Another limitation was that the PPV for the "or" rule and the weighted sum were modest at a 25% base rate (0.56 and 0.61, respectively). The PPV is usually low when the prevalence of the disease is low,²⁷ but it can also be overestimated depending on the populations investigated.²⁷ Therefore, a good PPV must be interpreted with caution before being communicated to clinicians.²⁸ In addition to the problem mentioned with the PPV, the generalization of statistical study results applied to clinical practice remain difficult.²⁸ Therefore, to lessen this difficulty, it is suggested that clinicians also include range values,³ which were calculated in this study so as to have no false positives and no false negatives. The results of the current study showed that only converters ($n = 5$) failed both tests ($MIS < 6$ and $IST < 25$) and that only nonconverters ($n = 11$) obtained high scores on both tests ($MIS > 7$ and $IST > 29$). Although this was accurate in 100% (16/16), range values could be performed for only 15% (16/106) of the group with QD. Nevertheless, when range values are applied to patients with CDR scores of 0 and 1 used in this study to compare QD cognitive performance (Table 1), 55 patients with mild dementia scored positive on both tests but none with no dementia. In contrast, 117 patients with no dementia scored high on both tests, compared with only five in the group with mild dementia.

The results of this study, obtained from patients with QD who were recruited from memory consultations, cannot be routinely applied to primary care settings. Although the role of dementia screening in primary care still remains debated, screening is frequently proposed as a first stage in the diagnostic procedure, with specialized consultations conducted in a second stage.⁵ Although a negative screen may reassure patients who are concerned about cognitive decline, the distress caused by a false positive result could be a possible limitation of this procedure.³ Administration of a second cognitive screening within 2 weeks to 3 months was proposed to reduce the false-positive rate.²⁹ This technique needs to be further developed so as to avoid unnecessarily longer and more costly testing in patients without dementia.²⁹ Prospective studies are warranted to establish the op-

timal delay for a repeated screening assessment. General practitioners are currently the best professionals to administer this second screen. When considering positive results, they can, without prejudging the diagnosis, routinely begin their research of curable pathologies. Even when rare, diagnosis and treatment of these curable pathologies remains of primary interest for early screening.³⁰ Based on their knowledge of the clinical and family context general practitioners are also in the best position to subsequently decide and propose a specialized consultation for their patients.

The MIS-IST pairing is quick (5 minutes) and simple to use and score. When including cutscore ($MIS < 6$ or $IST < 25$) or range values, this test could be considered a useful screening procedure for all types of dementias.

ACKNOWLEDGMENTS

The authors are grateful to Richard Medeiros, Rouen University Hospital Medical Editor, for editing the manuscript.

Conflict of Interest: All authors confirm that there was no conflict of interest and that they have not received any payment for research, consultantships, or speakers forum, and do not have any company holdings or patents in connection with this manuscript.

Author Contributions: Gilles Chopard and Jean Gal-miche: study concept and design, statistical analysis and interpretation of data, preparation of the manuscript. Gregory Tio and Alexandre Pitard: statistical analysis and interpretation of data. Geraldine Vanholsbeeck and Mickael Binetruy: acquisition of subjects and data. Lucien Rumbach: preparation of the manuscript.

Sponsor's Role: None.

REFERENCES

- Lorentz WJ, Scanlan JM, Borson S. Brief screening tests for dementia. *Can J Psychiatry* 2002;47:723-732.
- Palmer K, Bäckman L, Winblad B et al. Detection of Alzheimer's disease and dementia in the preclinical phase: Population based cohort study. *BMJ* 2003;326:1-5.
- Cullen B, O'Neill B, Evans JJ et al. A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry* 2007;78:790-799.
- Devanand DP, Folz M, Gorlyn M et al. Questionable dementia: Clinical course and predictors of outcome. *J Am Geriatr Soc* 1997;45:321-328.
- Grober E, Hall C, Lipton RB et al. Primary care screen for early dementia. *J Am Geriatr Soc* 2008;56:206-213.
- Chopard G, Pitard A, Ferreira S et al. Combining the memory impairment screen and the Isaacs set test: A practical tool for screening dementias. *J Am Geriatr Soc* 2007;55:1426-1430.
- Buschke H, Kuslansky G, Katz M et al. Screening for dementia with the memory impairment screen. *Neurology* 1999;52:231-238.
- Isaacs B, Kennie AT. The set test as an aid to the detection of dementia in old people. *Br J Psychiatry* 1973;123:467-470.
- Folstein ME, Folstein SE, McHugh PR. "Mini mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatry Res* 1975;12:129-138.
- Lawton M, Brody E. Assessment of older people: Self maintaining and instrumental activities of daily living. *Gerontologist* 1969;9:179-186.
- Hugues CP, Berg L, Danziger WL et al. A new clinical scale for the staging of dementia. *Br J Psychiatry* 1982;140:566-572.
- Diagnostic and Statistical Manual of Mental Disorders, 4th Ed. Washington, DC: American Psychiatric Association, 1994.
- McKhann G, Drachman D, Folstein M et al. 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;34:939-944.
14. Pohjasvaara T, Nantyla R, Ylikoski R et al. Comparison of different clinical criteria (DSM-III, ADDTC, ICD-10, NINDS-AIREN, DMS-IV) for the diagnosis of vascular dementia. *National Institute of Neurological Disorders and Stroke Association-Association Internationale pour la Recherche et l'Enseignement en Neurosciences. Stroke* 2000;31:2952-2957.
 15. McKeith IG, Galasko D, Kosaka K et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies: Report of the consortium on DLB international Workshop. *Neurology* 1996;47:1113-1124.
 16. The Lund and Manchester Group. Clinical and neuropathological criteria for frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 1994;57:416-418.
 17. Masur DM, Sliwinski M, Lipton RB et al. Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons. *Neurology* 1994;44:1427-1432.
 18. Zhou XH, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. New York: John Wiley and Sons, 2002.
 19. Grober E, Buschke H. Genuine memory deficits in dementia. *Dev Neuropsych* 1987;3:13-36.
 20. Visser PJ. Predictors of Alzheimer Type Dementia in Subjects with Mild Cognitive Impairments. Maastricht: Neuropsych Publishers Maastricht, 2000.
 21. Mackinnon A, Mulligan R. Combining cognitive testing and informant report to increase accuracy in screening for dementia. *Am J Psychiatry* 1998;155:1529-1535.
 22. Smith T, Gildeh N, Holmes C. The Montreal Cognitive Assessment: Validity and utility in a memory clinic setting. *Can J Psychiatry* 2007;52:329-332.
 23. Kalbe E, Kessler J, Calabrese P et al. DemTect: A new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int J Geriatr Psychiatry* 2004;19:136-143.
 24. Tian J, Bucks RS, Haworth J et al. Neuropsychological prediction of conversion to dementia from questionable dementia: Statistically significant but not yet clinically useful. *J Neurol Neurosurg Psychiatry* 2003;74:433-438.
 25. Lyketsos CG, Lopez O, Jones B et al. Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: Results from the Cardiovascular Health Study. *JAMA* 2002;288:1475-1483.
 26. Petersen RC. Mild cognitive impairment or questionable dementia? *Arch Neurol* 1999;57:566-583.
 27. Huppert FA. Guidelines for assessing the predictive value of neuropsychological measures. *Cortex* 2002;38:414-416.
 28. Steurer J, Fischer JE, Bachman LM et al. Communicating accuracy of tests to general practitioners: A controlled study. *BMJ* 2002;324:824-826.
 29. Fujii D, Hishinuma E, Masaki K et al. Dementia screening. Can a second administration reduce the number of false positives? *Am J Geriatr Psychiatry* 2003;11:462-465.
 30. Ashford JW, Borson S, O'Hara R et al. Should older adults be screened for dementia? It is important to screen for evidence of dementia! *Alzheimer's & Dementia* 2007;3:75-80.

APPENDIX A

The Memory Impairment Screen (MIS)

The MIS is a 4-minute, four-item delayed free- and cued-recall memory test.

1. The subject is informed that he or she must learn four words from different semantic categories.
2. The subject is asked to read the four words out loud.
3. The subject is asked to locate, point to, and read out the word once the examiner has given the semantic cue (e.g., "Read out and point to the name of the animal").
4. The words are then hidden.
5. After an interfering task (counting backwards from 100) lasting at least 2 minutes, the subject is invited to reproduce the words learned in any order over 20 seconds (free recall).
6. A cued recall is proposed for any words not mentioned by free recall: the examiner repeat the semantic cue (e.g., "What was the name of the animal?").

Grading: For each word, the score for a correct answer by free recall is 2 points and for cued recall, 1 point. The total MIS score (total MIS = (2 × free) + cued) ranges from 0 to 8.

The Isaacs Set Test (IST)

The IST is a 1-minute verbal category fluency task.

Subject must orally produce as many words as possible for the following categories within 15 seconds: colors, animals, fruits, and cities.

Grading: The total number of items named produces the score (any repeated words or intrusions are not included).

III. ARTICLE 3: Screening for Amnestic Mild Cognitive Impairment: Spectrum Effect and Spectrum Bias.

Cette troisième étude visait à déterminer l'impact de l'effet de spectre et du biais spectral sur la performance de l'outil associant le MIS et le STI dans la détection du MCI amnésique (MCIa).

Un échantillon de 2490 sujets âgés de 60 à 89 ans (1500 patients atteints d'un MCIa, 990 patients avec plainte sans troubles cognitifs) appartenant à la base de données Rapid-Fr et consultant pour une plainte cognitive, a été inclus (annexe 6).

L'effet de spectre a été défini par la variation de la sensibilité et de la spécificité de la combinaison de tests MIS/STI et de l'épreuve du MMSE dans la détection des troubles cognitifs selon l'âge et le niveau scolaire, en fonction du type de MCIa (déficit mnésique isolé ou associé à d'autres déficits cognitifs) et du degré de son atteinte cognitive (atteinte légère ou très légère).

Le biais spectral a été estimé selon la méthode de Goerhing *et al.*, (2004) par comparaison des RVP et des RVN obtenus par l'utilisation d'une valeur seuil fixe appliquée sur l'échantillon global puis sur les différents sous-groupes de cet échantillon définis selon quatre variables : l'âge, le niveau scolaire, le type de MCIa et son degré d'atteinte cognitive.

Ainsi, le rapport $RVP_{(\text{sous-groupe})}/RVP_{(\text{global})}$ ou le rapport $RVN_{(\text{sous-groupe})}/RVN_{(\text{global})}$ comme mesure d'évaluation du biais spectral est considéré comme significatif lorsque l'intervalle de confiance à 95% (IC 95%) ne contient pas la valeur 1. En revanche, ce rapport n'est pas significatif lorsque la valeur 1 est contenue dans l'IC 95%.

Nos résultats ont montré une variation très importante de la sensibilité (0 à 94%) et de la spécificité (52 à 95%) pour l'utilisation d'une valeur seuil fixe optimale (annexe 7). Cette variation est l'expression d'un effet de spectre.

Pour cette même valeur seuil, la sensibilité et la spécificité globale était de 76% et de 78%, respectivement dans le groupe des cas MCIa avec déficits multiples (aMCI-m en anglais dans l'article) avec un RVP de 3,45 et un RVN de 0,31. On retrouve une importante variation de la sensibilité dans ce groupe (54% à 94%) avec un fort taux de faux négatifs (46%) chez les patients âgés de 60 à 69 ans, de haut niveau scolaire et de faux positifs (48%) chez les 80-89 ans, de bas niveau. Le RVP variait de 1,93 à 12,33 et le RVN de 0,10 à 0,48. Cette variation a entraîné un biais spectral dans deux classes d'âge (60-69 ans et 80-89 ans) et pour les trois niveaux scolaires étudiés (bas, moyen, haut).

L'utilisation de valeurs-seuils ajustées selon les variables âge et niveau scolaire dans le groupe des patients MCIa avec déficits multiples (ou aMCI-m) limite considérablement l'effet de spectre observé avec des sensibilités variant de 69% à 79% et des spécificités de 85% à 94% (Annexe 8). Ces valeurs seuils ajustées peuvent permettre au clinicien d'interpréter les résultats avec plus de nuances en limitant les marges d'erreurs en ce qui concerne notamment les taux de faux positifs. Des résultats similaires ont été retrouvés pour le MMSE (Annexe 9, 10 et 11).

L'impact de l'effet de spectre et du biais spectral suggèrent que la performance globale des tests étudiés dans ce travail, estimée à partir d'une valeur seuil fixe, exposerait à un risque de sous-estimer les troubles cognitifs chez les sujets de 60-69 ans de haut niveau d'éducation et de surestimer les troubles cognitifs chez les sujets plus âgés de 80-89 ans, quel que soit le niveau d'éducation. Ce résultat peut avoir des conséquences sur l'interprétation des performances du test et la prise de décision clinique. A titre d'exemple, pour une probabilité pré-test de 20%, un test positif pour un patient âgé de 84 ans, avec un niveau scolaire moyen, donnerait une probabilité post-test de 36% après utilisation d'une valeur-seuil fixe et de 71% avec une valeur seuil ajusté.

Cet article est actuellement soumis pour publication à la revue *Alzheimer's & Dementia* (*IF*=6,373).

Screening for Amnestic Mild Cognitive Impairment: Spectrum Effect and Spectrum Bias.

Chopard G, Msc,^{1,2,3} Mauny F, MD, PhD,⁴ Puyraveau M, Msc,⁴ Meyer A, Msc,³ Binetruy M, Msc,³ Rumbach L, MD, PhD,^{1,3} Vandael P, MD, PhD,^{2,3} Galmiche J, MD.^{1,3}

¹ Department of Neurology, University Hospital of Besançon, France

² Department of Psychiatry, University Hospital of Besançon, France

³ Rapid-Fr Network (Regional Network for Diagnostic Aid and Management of Patients with Cognitive Impairment in the Franche-Comté geographical area), Besançon, France

⁴ Clinical Methodology Center, University Hospital of Besançon, France

Author contributions

G. Chopard: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, statistical analysis.

F. Mauny: analysis or interpretation of data, statistical analysis.

M. Puyraveau: statistical analysis.

A. Meyer: acquisition of data.

M. Binetruy: drafting/revising the manuscript.

P. Vandael: drafting/revising the manuscript.

L. Rumbach: drafting/revising the manuscript.

J. Galmiche: drafting/revising the manuscript, study concept or design, analysis or interpretation of data.

ABSTRACT

Objective: To assess the magnitude of spectrum effect and spectrum bias for the combination of Memory Impairment Screen (MIS) with Isaacs Set Test (IST) in the detection of amnesic Mild Cognitive Impairment (aMCI) and determine whether adjusted cut-offs may improve clinical efficacy.

Methods: Participants recruited from the database of the Regional Network for Diagnostic Aid and Management of Patients with Cognitive Impairment (Rapid-Fr network) were administered the MIS, the IST, The Mini-Mental State Examination (MMSE) and a neuropsychological battery of tests. We calculated the sensitivity, specificity, positive and negative likelihood ratio (LR+, LR-) of single and age-education adjusted cutoff scores in four subtypes of aMCI: single very aMCI, single aMCI, very aMCI multi-domain and aMCI multi-domain. Spectrum effect was measured by the variation of sensitivity and specificity in the different subgroups of aMCI. Spectrum bias was expressed as a ratio of LRs.

Results: A total of 1500 subjects with aMCI and 990 with no cognitive impairment (NCI) were included in the study. Spectrum effect was found using single cut-offs. This led to a spectrum bias in the multidomain aMCI group (aMCI-m) with a relevant impact upon clinical decisions. A lesser degree of spectrum effect was also found with adjusted cut-offs. This latter effect had more balanced discriminative performances with high specificities.

Conclusions: Clinicians should consider single cutoff scores carefully before determining whether performances of a given screening test are generalizable to their own patient population. Face with the challenge to screen aMCI among a broad and heterogeneous population, adjusted cut-offs seem more adapted than single cutoffs.

keywords: *Amnesic mild cognitive impairment, screening, spectrum effect, spectrum bias, single cut-offs, age-education adjusted cut-offs.*

1. Introduction

Amnesic Mild Cognitive Impairment (aMCI) is a heterogeneous group.¹ Screening tests for aMCI require to detect the presence of cognitive impairment among a broad range of patients having cognitive complaints so as to select those individuals who are at higher risk for developing dementia.² Case-healthy control studies, performed to validate these screening tests, may lead to an overestimation of their accuracy.² Sensitivity and specificity are often determined based on a single cut-off score whereas it is well established that most cognitive performances are influenced by age and education.³ Consequently, the performance of a test may vary across subgroups of patients, not only with different demographic characteristics, but also types of disorders or degrees of cognitive impairment.⁴ This phenomenon has been referred to as the spectrum effect.⁵ In some situations, these subgroup variations, may also affect the clinician's decision, known as a spectrum bias,⁶ particularly if clinicians use the overall performance of a single cut-off score test for their own patient population.⁷

More appropriate cut-offs have been advocated as an alternative to the single threshold approach.⁸ To what extent they can modify the screening performance has, to our knowledge, never been previously assessed. A single cut-off score combining both the Memory Impairment Screen (MIS) and Isaacs Set Test (IST) was evaluated for the detection of questionable dementia.⁹

The aim of this study was to assess the magnitude of spectrum effect and spectrum bias for the MIS-IST pairing used to detect aMCI and determine whether adjusted cut-off scores can improve their clinical efficacy.

2. Methods

2.1. Study population

Data were retrospectively obtained from the database of the Regional Network for Diagnostic Aid and Management of Patients with Cognitive Impairment (Rapid-Fr network).¹⁰ The aim of this health network, financed by the French government is to coordinate memory consultations in selected geographical areas of France (Alsace, Bretagne, Franche-Comté and Lorraine). The cognitive evaluation was made by trained neuropsychologists during memory consultations and based on the RAPID neuropsychological battery of tests which has been standardized.¹¹

All subjects were aged between 60 and 89 years old and attended these 32 local memory impairment consultations centers between January 2007 and December 2011. They were referred because of their reported subjective complaints of memory loss or cognitive alteration. All subjects with objective memory impairment, based on neuropsychological examination, underwent a standard medical and neurological history assessment, neuroimaging (computed tomography or magnetic resonance imaging) and blood analysis tests. Exclusion criteria were acute or chronic neurological pathologies (traumatic brain injury, brain tumors, stroke and dementia) or progressive psychiatric illness (schizophrenia, bipolar disorders) except for major depression because this is a factor that may be associated with early dementia.¹²

2.2. Standard protocol approval and patient consents.

The Ethics Committee of Besançon University Hospital gave its official approval to conduct the study, and all participants gave their informed written consent for study inclusion.

2.3. Neuropsychological testing

Each participant first underwent the screening tests in the following order: MIS, then IST.

Memory Impairment Screen.¹³ The MIS is a 4-minute, four item delayed free and cued recall memory test. Subjects are presented the names of items in four different categories. After an interfering task (counting backwards from 100) lasting approximately 2 minutes, the subjects are asked to perform the free recall of the four items in any order. A cued recall is proposed for any words not mentioned using free recall. The total MIS score ranged from 0 to 8.

Isaacs Set Test.¹⁴ The IST is a 1-minute verbal category fluency task. The subject has to orally produce as many words as possible for the following categories within 15 seconds: colors, animals, fruit and cities. The number of items named produces the various scores.

Then, the Mini-mental State Examination (MMSE)¹⁵ and the RAPID battery tests were administered to each subject during the same evaluation session. This battery of tests assesses

five cognitive domains: memory, psychomotor speed, executive function, language and constructional praxis. It included the cognitive measures listed below.

Free and Cued Recall Test (FCSRT).¹⁶ This test consists of a 16 item word list belong to 16 different semantic categories. The FCSRT first employs a study phase to ensure semantic encoding of all 16 items, directly followed by a recall phase that included three trials. Each trial consisted of a free recall immediately followed by a cued recall for those items not retrieved on free recall. Two measures were included: free recall (the cumulative sum of free recall from the three trials; range from 0 to 48) and total recall (the cumulative sum of free recall plus cued recall from the three trials; range from 0 to 48).

Trail Making Test (TMT).¹⁷ In part A, subjects were required to connect randomly ordered numbers (1,2,3,etc.). In part B, they were asked to alternate between numbers and letters (1,A,2,B,etc.). The time taken to complete each of the tasks was used as the two primary performances.

DENO 30.¹¹ This test measures visual naming and presents 30 pictures. These items are stratified into three groups of 10 items each representing objects, animals and actions. The maximum score is 30.

Constructional Praxis (Copy).¹⁸ This task is part of the Signoret's Battery of Cognitive Efficacy (BEC 96) in which subjects were required to copy overlapping triangles. The maximum score is 6.

2.4. Other clinical measures

Functional autonomy was assessed by the four instrumental Activities of Daily Living (IADL)¹⁹ (i.e. ability to use the telephone, independence in transportation, self administration of medication, ability to handle financial affairs). Major depression was diagnosed using the DSM IV depression scale.²⁰

2.5. Diagnostic categories

Impairment ratings were assigned based on age-educational norms of the RAPID battery tests.¹¹ The MIS, IST and MMSE were not used in the diagnostic procedure. Subjects were classified as follows:

Subjects with subjective complaints but no cognitive impairment (NCI) were characterized by a score at or above the 16th percentile on all the RAPID battery tests compared to normative data by age group and education level. They had the intact ability to perform daily living activities.

Amnesic mild cognitive impairment (aMCI) was defined according to the following criteria:²¹ (1) presence of a subjective complaint by a subject or informant; 2) objective memory impairment (from the FCRST); 3) preserved activities of daily life (from the IADL); and 4) absence of DSM-IV criteria for dementia.²⁰

Four aMCI subtypes were considered in this study:

Single domain amnesic MCI (aMCI-s): subjects with memory deficits defined as a score below the 10th percentile compared with the age and specific educational norms on at least 1 of 2 FCSRT (free recall or total recall); performance on tests of the RAPID battery assessing all other areas of cognitive functioning (language, psychomotor speed, executive functions and visuo-spatial processing skills) was required to be above the 16th percentile.

Multi-domain amnesic MCI (aMCI-m): subjects with a memory deficit and at least in one other cognitive test (i.e. <10th percentile compared with the age and education of specific norms of the RAPID battery test).

In addition to the 10th percentile cut-off score to define cognitive impairment a more moderate threshold between the 10th and the 16th percentile was used to designate subjects as having a very mild cognitive impairment: very aMCI-s and very aMCI-m, respectively.

2.6. Statistical analysis

Optimal single cut-off score that maximizes sensitivity and specificity was first determined using two diagnostic categories. A sample that included patients with aMCI (n=1163) and a sample that included patients with NCI (n=960) were selected, at random, from the participants groups so as to match with age ($t = -1.71$; $p = 0.09$) and level of education ($\chi^2 = 4.57$; $p = 0.10$).

To produce age-education adjusted cut-off scores, all subjects with aMCI and NCI were stratified into nine age-educational categories. The age groups were 60 to 69, 70 to 79 and 80 to 89. Age groups were subdivided by educational level as follows: no formal education, French primary education certificate (low educational level); Certificate of Professional Aptitude or the school leaving certificate (middle educational level); Baccalaureate or higher (high educational level).

Once optimal single and adjusted cut-off scores were determined, they were applied to the four subtypes of aMCI according to nine age-education subgroups. We next examined the ability of the MIS-IST pairing to distinguish between subgroups with aMCI and those with NCI. The same procedure was also applied on the MMSE.

Screening accuracy of the MIS-IST pairing was assessed by calculating sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR-). Sensitivity for the MIS-IST pairing was calculated using a positive result in either MIS or IST, or both. Specificity for the MIS-IST pairing was calculated using a negative result in both MIS and IST.

The spectrum effect was measured as the variation of the sensitivity and the specificity in the four subtypes of aMCI by subgroups according to age and educational level. To determine spectrum bias binding to single cut-off score, stratified analyses of LR were performed for the three categories of age and also for the three categories of educational level. To evaluate spectrum bias, a ratio expressed as the subgroup LR divided by the LR in the overall population was calculated.⁶ If the confidence interval (CI) did not include unity, the bias was considered statistically significant.⁶

Potential advantage of adjusted cut-off scores was evaluated by comparing sensitivity, specificity, LR+ and LR- with those of single cut-off score when applied to the nine age-educational categories.

Continuous variables were compared using the Student's-t-test, and categorical data were analyzed using the chi-square.

All the computations were performed with Strata Software (release 8.0, Statacorp, College Station, TX).

3. Results

A total of 2490 subjects with subjective complaints were included in this study: geographical areas of Alsace (n=630), Bretagne (n=491), Franche-Comté (n=973), Lorraine (n=396). Characteristics of the participant groups are presented in Table 1. A flowchart diagram of the participants studied was presented online.²² (*voir annexe 6*)

The optimal single cut-off score was MIS less than 7 or IST less than 28. When applying this single cut-off score to the four subtypes of aMCI in each age and educational level subgroups, sensitivity varied from 0 to 94% and specificity from 52 to 95%. (*voir annexe 7*)

Given that the highest detection rate was for the aMCI-m subjects, the analysis of spectrum bias was only made in this population. For this purpose, optimal single cut-off (MIS <7 or IST <28) was first applied in a sample group of subjects with aMCI-m (n=554) and NCI (n=772) matched with age ($t = -1.43$; $p = 0.15$) and level of education ($\chi^2 = 4,68$; $p = 0.10$). At this optimal cut-off point, the MIS-IST pairing had an overall sensitivity of 76% (95% CI: 0.72 to 0.79) and a specificity of 78% (95% CI: 0.75 to 0.81) with a LR+ of 3.45 (95% CI: 3 to 3.97) and a LR- of 0.31 (95% CI: 0.26 to 0.36). Second, the same single cut-off value was applied in the three categories of age and the free categories of educational level of the aMCI-m population: LR+ ranged from 2.24 (95% CI: 1.88 to 2.67) to 7.30 (95% CI: 4.51 to 10.58) and LR- from 0.16 (95% CI: 0.10 to 0.31) to 0.40 (95% CI: 0.33 to 0.51) (table 2). The spectrum bias was evaluated by the LR ratio for the MIS-IST result in the subgroup of aMCI-

m divided by the LR for the MIS-IST pairing for the overall population of aMCI-m. A statistically significant bias was found in 5 subgroups (Table 2).

Adjusted cut-off scores for the MIS-IST pairing are shown in table 3. These values were quite different from the single cut-off point. When applying adjusted cut-offs in the nine subgroups of aMCI-m stratified by age and education, sensitivity varied from 67% (95% CI: 0.51 to 0.79) to 79% (95% CI: 0.64 to 0.89) and was lower for the other groups. Specificity ranged from 85% (95% CI: 0.77 to 0.90) to 94% (95% CI: 0.89 to 0.96). (*voir annexe 8*). The detection rate also being higher for the aMCI-m subjects, analysis of a potential benefit of adjusted cut-off scores was made only for this group. In the aMCI-m subgroups, LR+ ranged from 4.73 (95% CI: 2.85 to 7.46) to 12.33 (95% CI: 6.27 to 21.95), and LR- from 0.23 (95% CI: 0.13 to 0.46) to 0.37 (95% CI: 0.24 to 0.60).

When comparing these performances with those obtained with single cut-off applied on the aMCI-m (Table 4), the discrepancies were observed in all subgroups except for middle education, 60-69 years.

For the MMSE, screening performance also varied in the four subtypes of aMCI (*voir annexe 9*) and a spectrum bias occurred when using a single cut-off score (i.e. MMSE <27) in the aMCI-m subgroups. (*voir annexe 11*)

4. Discussion

This study showed that MIS-IST screening performance varies across aMCI subtypes and according to age and education level, causing a spectrum effect. In subjects with NCI, specificity was also affected by age and educational level. This screening test was found to be more sensitive in detecting patients with aMCI-m, population at high risk to convert to dementia.²³ The sensitivities were more balanced in this group than in the other aMCI populations. If spectrum effect reflects an inherent subgroup variation in the performance of a test, it is important to specify whether this variation also affects the post-test probability and thus may potentially modify the clinical decision, due to spectrum bias.^{5,6}

A spectrum bias was found in all education categories and in two out of three age categories. Nevertheless, a significant spectrum bias may not routinely lead to an impact upon

clinical decision. It may depend on the pretest probability of the disorder.⁶ Moreover, the LR is a global measurement thus, an identical LR may be the result of different sensitivity and specificity. In subjects aged 80-89 years, the LR- was 0.16 with a corresponding sensitivity of 90% and a specificity of 60%. This means that after a negative MIS-IST result, the risk of having aMCI-m would decrease. For example, with an estimated pretest probability of 40%, a post-test probability would be 9.7%. Nevertheless, if a negative test tended to rule out aMCI-m in this age group, only few subjects with NCI were correctly classified (i.e. specificity of 60%). With a similar LR- but obtained with more balanced sensitivity and specificity (e.g., 87% and 83%, respectively), a greater proportion of NCI patients would have been correctly classified (i.e. specificity of 83%).

It is therefore important to obtain the most accurate balance between sensitivity and specificity. This is what we observed in our study based on age-education adjusted cut-offs in the aMCI-m population. Therefore, the spectrum effect is limited and less pronounced compared to only a single cut-off (Table 4).

In the older people, LR+s are much greater rendering positive MIS-IST results more reliable among patients that are considered at high risk for dementia. For example, in the older subjects with low education, if the pretest probability is estimated at 20%, when the MIS-IST result is positive, the post-test probability would be 36% with a single cut-off and increases to 71% with the adjusted cut-off. Though less good LRs were associated with the use of adjusted cut-offs, it yielded a lower number of false positive rates that may cause distress for people and their families.²⁴

A limitation of the present study was the small sample size in the very aMCI-s subgroups. The statistical power was thus very low to demonstrate a decrease of sensitivity in these subgroups. However, our results tended to show that the sensitivity of the MIS-IST pairing decreased with milder cognitive performance. Moreover, the prevalence of aMCI cases was different from one subgroup to another and probably had an effect on accuracy indices.²⁵ Another limitation was that non-amnesic MCI patients were not included in this study though they are likely to progress to a non-AD dementia.²⁶ Lastly, reproducibility of our results was not evaluated in an independent sample.

Although this study is not prospective, retrospective data were available for 349 participants who had been followed-up for at least 12 months after their baseline assessment, during the same study period. It is interesting to note that among patients with NCI, 27% progressed to aMCI or very aMCI whereas only 1.9% converted to dementia during a mean follow-up of 29 months. In the aMCI-m group, 24% converted to dementia at a mean follow-up of 23 months and 9% for the very aMCI-m group (mean follow-up of 26 months). By contrast, in aMCI-s and very aMCI-s groups, only 1.4 % progressed to dementia during a mean follow-up of 23 months and 25.7 months, respectively. These results indicate that, in our study population, patients presenting with amnesic multi-domain MCI, even in a very mild form, could be considered at a higher risk of developing dementia than those with a pure amnesic MCI.

If we had applied the adjusted cut-off scores at the follow-up screening, the MIS-IST pairing would have classified 82% of NCI subjects who remained stable and aMCI-m that reverted to normal. It would also have classified 86% of aMCI-m subjects who developed dementia and NCI that progressed to a-MCI-m.

The spectrum effect and spectrum bias were found not only with the MIS-IST pairing but also for MMSE. We do not think that a spectrum effect or spectrum bias depends on the type of the test used. Therefore, one should be cautious and routinely assess these phenomena before suggesting the overall performance of a screening test. The use of adjusted cut-offs, by reducing the false positive cases has a clinical impact. Although less sensitive in detecting cognitive impairment than a single cut-off in the older populations, the highest LR+ in almost all aMCI-m subgroups will help clinicians to increase their level of certainty regarding whether a cognitive impairment is present. It subsequently avoids unnecessary testing for further evaluation on patients without cognitive impairment, particularly those with a lower educational level that are expected to obtain poor cognitive scores.²⁷

In our study we attempted to value the specificity based on a short screening test (5 min.), which is simple to use and to score, which would meet the practical requirements of the clinician.²⁴ We are aware that spectrum effect in this field is probably not simply related to age, education or type of the aMCI alone. Other factors such as previous abilities,

psychological variables, social and cultural contexts should also be considered and analyzed as potential sources of variation and bias on estimates of screening test performance.

ACKNOWLEDGEMENT

The authors are grateful to Richard Medeiros, Medical Editor, Medical Editing International, for his valuable editing of the manuscript.

Disclosure

G. Chopard reports no disclosures.

F. Mauny reports no disclosures.

M. Puyraveau reports no disclosures.

A. Meyer reports no disclosures.

M. Binetruy reports no disclosures.

Dr Vandel has received research support and speaker honoraria from BMS, GSK, Wyeth, Servier, Janssen, Sanofi, Biocodex, Pfizer, Lundbeck, Lilly, Otsuka, Novartis, Astra-Zeneca.

Dr Rumbach has received research support and speaker honoraria from Biogen Idec, Sanofi, Teva, Novartis, Merck Sereno.

Dr Galmiche reports no disclosures.

REFERENCES

1. Ganguli M, Dodge HH, Shen C, DeKosky ST. Mild cognitive impairment, amnesic type: an epidemiologic study. *Neurology* 2004;63(1):115-121.
2. Van der Schouw YT, Verbeek AL, Ruijs SH. Guidelines for the assessment of new diagnostic tests. *Invest Radiol* 1995;30:334-340.
3. Ganguli M, Snitz BE, Lee CW, Vanderbilt J, Saxton JA, Chang CC. Age and education effects and norms on a cognitive test battery from a population-based cohort: the Monongahela-Youghiogheny Healthy Aging Team. *Aging Ment Health*. 2010;14:100-107.
4. Lorentz WJ, Scanlan BA, Borson S. Brief Screening Tests for Dementia. *Can J Psychiatry* 2002;47:723-733.
5. Mulherin SA, Miller WC. Spectrum bias and spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002;137:598-602.
6. Goerhing C, Perrier A, Morabia A. Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Statist Med* 2004;23:125-135.
7. Willis BH. Spectrum bias-why clinicians need to be cautious when applying diagnostic test studies. *Family practice* 2008;25:390-396.
8. Petersen RC. MCI treatment trials: failure or not? *Lancet Neurol* 2007;6:473-475.
9. Chopard G, Vanholsbeeck G, Tio G, Pitard A, Binetruy M, Rumbach L et al. Rapid screening of cognitive change in patients with questionable dementia using the Memory Impairment Screen and the Isaacs Set Test. *J Am Geriatr Soc* 2009;57:703-708.
10. Vanholsbeeck G, Ferreira S, Pitard A, Bouiller C, Chopard G, Rumbach L et al. Intérêt d'un dossier national de réseau de santé dans le cadre de la veille sanitaire des maladies d'Azheimer et syndromes apparentés: le logiciel « rapid ». *Filière et Réseaux de Santé* 2007;13:37-39.
11. Ferreira S, Vanholsbeeck G, Chopard G et al. Comparative norms of RAPID neuropsychological battery tests for subjects aged between 50 and 89 years. *Revue neurol* 2010;166:606-614.

12. Modrego PJ, Ferrandez J. Depression in patients with mild cognitive impairment increases the risk of developing dementia of Alzheimer type: a prospective cohort study. *Arch Neurol* 2004;61:1290-1293.
13. Buschke H, Kuslansky G, Katz M, Stewart WF, Sliwinski MJ, Eckholdt HM et al. Screening for dementia with the Memory Impairment Screen. *Neurology* 1999;52:231-238.
14. Isaacs B, Kennie AT. The Set Test as an aid to the detection of dementia in old people. *Br J Psychiatry* 1973;123:467-470.
15. 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;12:189-198.
16. Grober E, Lipton RB, Hall C, Crystal H. Memory impairment on free and cued selective reminding predicts dementia. *Neurology* 2000;54:827-832.
17. Reitan RM, Wolfson D. The Halstead-Reitan Neuropsychological Test Battery: Therapy and clinical interpretation. Tucson, AZ: Neuropsychological Press; 1985.
18. Signoret JL, Allard M, Benoit N, Bolgert F, Bonvarlet M, Eustache F. Evaluation des troubles de mémoire et des désordres cognitifs associés. BEC 96. Paris: Ipsen, 1998.
19. Lawton MP, Brody EM. Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist* 1969;9:179-186.
20. American Psychiatric Association. Diagnostic and Statistical Manual of Mental disorders. 4th ed. Washington DC: 1994.
21. R. C. Petersen, J.C. Stevens, M. Ganguli, E. G. Tangalos, J.L. Cummings, S. T. DeKosky, Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review) Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2001;56:1133-1142.
22. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Inwig LM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003;138:40-44.
23. Tabert MH, Manly JJ, Liu X, Pelton GH, Rosenblum S, Jacobs M et al. Neuropsychological prediction of conversion to AD in patients with mild cognitive impairment. *Arch Gen Psychiatry* 2006;63:916-24.

24. Holsinger T, Deveau J, Boustani M, Williams JW. Does this patient have dementia? *Jama* 2007;297:2391-2404.
25. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med* 1997;16:981-991.
26. Busse A, Hensel A, Gühne U, Angermeyer MC, Riedel-Heller SG. Mild cognitive impairment. Long-term course of four clinical subtypes. *Neurology* 2006;67:2176-85.
27. Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the Mini-Mental State Examination by age and educational level. *Jama* 1993;269:2386-91.

Table 1 Demographic and neuropsychological characteristics of participants by cognitive status

	NCI	very aMCI-s	very aMCI-m	aMCI-s	aMCI-m
	(n=990)	(n = 133)	(n = 407)	(n = 366)	(n = 594)
Age, y	71.7 ± 7.7	72.2 ± 7.9	73.9 ± 6.9	74.7 ± 7.6	74.0 ± 6.4
Sex, % male	37.8	53.4	44.9	58.7	49.1
Education, %					
High	20.6	16.5	23.6	30.3	29.2
Middle	34.8	39.1	30.0	37.7	41.0
Low	44.6	44.4	46.4	32.0	29.8
Depression, %	6.5	7.5	7.9	9.3	8.1
Memory Impairment Screen	7.5 ± 0.9	7.0 ± 1.4	6.4 ± 1.7	5.8 ± 2.0	6.0 ± 1.9
Isaacs Set Test	35.9 ± 6.6	33.6 ± 5.8	29.6 ± 6.6	32.3 ± 6.0	26.7 ± 6.4
MMSE	27.7 ± 1.9	27.1 ± 1.9	26.0 ± 2.4	26.4 ± 2.2	25.2 ± 2.7
FCSRT Free recall	26.8 ± 5.7	18.9 ± 4.8	17.1 ± 6.2	12.6 ± 6.0	12.4 ± 5.5
FCSRT Total recall	45.7 ± 2.4	41.4 ± 3.9	39.2 ± 6.1	35.2 ± 7.0	35.0 ± 7.1
Trail Making Test A	39.2 ± 11.6	41.6 ± 12.1	64.5 ± 30.4	41.5 ± 12.0	75.9 ± 39.1
Trail Making Test B	113.7 ± 42.4	118.0 ± 41.0	207.4 ± 104.5	123.1 ± 42.3	257.3 ± 122.1
DENO 30	29.8 ± 0.6	29.6 ± 0.8	28.7 ± 1.8	29.6 ± 0.8	28.1 ± 2.4
Copy	5.9 ± 0.2	5.9 ± 0.1	5.8 ± 0.7	5.9 ± 0.2	5.6 ± 1.0

Abbreviations: NCI = No Cognitive Impairment; very aMCI-s = very amnesic Mild Cognitive Impairment single domain; very aMCI-m = very amnesic Mild Cognitive Impairment multi-domain; aMCI-s = amnesic Mild Cognitive Impairment single domain; aMCI-m = amnesic Mild Cognitive Impairment multi-domain; MMSE = Mini-Mental State Examination; FCSRT = Free and Cued Selective Recall Reminding Test.

Table 2 Ratio of likelihood ratio for MIS-IST pairing in subgroups of aMCI-m

Groups	N	LR +	Ratio of LR+	95% CI	LR -	Ratio of LR -	95% CI
Overall		3.45	1	Referent	0.31	1	Referent
Age							
60 - 69 y	401	7.30	2.09*	1.45-3.00	0.40	1.28*	1.10-1.50
70 - 79 y	597	4.11	1.17	0.97-1.43	0.29	0.93	0.82-1.04
80 - 89 y	328	2.24	0.64*	0.55-0.75	0.16	0.51*	0.41-0.64
Education							
High	378	7.09	2.03*	1.37-3.01	0.37	1.20*	1.04-1.39
Middle	490	4.46	1.27*	1.03-1.58	0.27	0.86	0.74-1.00
Low	458	2.36	0.68*	0.58-0.78	0.27	0.86	0.73-1.02

Abbreviations: MIS = Memory Impairment Screen; IST = Isaacs Set Test; aMCI-m = amnesic Mild Cognitive Impairment multi-domain; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; CI = Confidence Interval.

* = statistically significant spectrum bias

Table 3 Optimal cut-off of the MIS-IST pairing according to age and level of education.

Variables	MIS	IST
Age, education		
60 - 69 years		
High	< 7	< 33
Middle	< 7	< 28
Low	< 7	< 27
70 - 79 years		
High	< 6	< 30
Middle	< 6	< 29
Low	< 6	< 27
80 - 89 years		
High	< 6	< 26
Middle	< 6	< 26
Low	< 6	< 23

Abbreviations: MIS = Memory Impairment Screen; IST = Isaacs Set Test.

Table 4 Sensitivity, specificity, positive and negative likelihood ratios for aMCI-m by cut-off scores of the MIS-IST combination

Demographic variables	Single cut-off score				Adjusted cut-off score			
	Sens %	Spec %	LR+	LR-	Sens %	Spec %	LR+	LR-
Age, education								
60 - 69 years								
High	54	95	10.80	0.48	72	93	10.28	0.30
Middle	74	94	12.33	0.28	74	94	12.33	0.28
Low	69	84	4.31	0.37	67	87	5.15	0.37
70 - 79 years								
High	68	90	6.80	0.35	76	90	7.60	0.27
Middle	78	85	5.20	0.26	71	89	6.45	0.32
Low	84	76	3.50	0.21	69	87	5.31	0.36
80 - 89 years								
High	85	81	4.47	0.18	77	93	11	0.24
Middle	94	59	2.29	0.10	79	92	9.87	0.23
Low	93	52	1.93	0.13	71	85	4.73	0.34

Abbreviations: IST = Isaacs Set Test, Sens = Sensitivity; Spec = Specificity; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; a MCI-m = amnesic Mild Cognitive Impairment multi-domain, MIS = Memory Impairment Screen

IV. DISCUSSION GENERALE

Le premier objectif de ce travail visait d'abord à élaborer un outil permettant de détecter une altération dans plusieurs domaines cognitifs compte tenu de la diversité des maladies neurodégénératives. En effet, si la MA typique débute le plus souvent par des troubles de la mémoire, ces derniers ne sont pas au premier plan dans les autres formes de pathologie dites non-Alzheimer qui peuvent d'avantage se manifester en début d'évolution par une altération des fonctions exécutives, visuo-spatiales, du langage ou par un ralentissement (Knopman *et al.*, 2003). C'est pourquoi nous avons conçu un outil en associant deux tests : le MIS élaboré par Buschke *et al.*, (1999) qui est un test d'exploration simple et rapide (4 min.) des troubles de la mémoire et le STI (Isaacs et Kennie, 1972), épreuve de fluence verbale catégorielle qui permet notamment d'explorer en 1 minute l'accès lexical sémantique, les fonctions exécutives et langagières. Le choix du STI repose notamment sur son évaluation dans l'étude française PAQUID (Dartigues *et al.*, 1991) qui a montré que ce test était sensible à un déclin cognitif futur (Jacquin-Gadda *et al.*, 2000, Amieva *et al.*, 2005).

L'objectif de notre premier article était d'évaluer la validité de l'association MIS/STI dans la détection des troubles cognitifs liés aux maladies neurodégénératives, et ce, dès le stade léger. Les résultats de l'étude cas-contrôles ont montré de bonnes performances globales en termes de sensibilité (92%) et de spécificité (90%) pour distinguer des patients atteints de troubles cognitifs liés à une pathologie neurodégénérative et des sujets sains non plaintifs. La comparaison des aires sous la courbe ROC a montré que la performance du couple MIS/STI était supérieure à celle du MMSE tout comme à celle du MIS ou du STI utilisés seuls. Toutefois, on a pu constater une variation de la performance de cette combinaison de tests selon le type de pathologie (de type MA vs non-Alzheimer) et son niveau de sévérité (stade léger, modéré, sévère). Ainsi, pour les mêmes valeurs seuils, la sensibilité passait de 98% pour les formes sévères à 89% pour les formes légères, quelle que soit l'origine de la pathologie. La sensibilité était meilleure pour les formes légères de la MA (96%) que pour les formes légères des autres pathologies non-Alzheimer (73%). Cette différence de performance pourrait s'expliquer par la prédominance des manifestations comportementales (apathie, irritabilité, délire, hallucinations) dans les formes légères de pathologies non-Alzheimer alors que les troubles cognitifs, plus discrets à ce stade, seraient donc plus difficiles à détecter par les tests que nous avons utilisés. Compte tenu de ces résultats, on devine que la performance

globale d'un test doit être considérée avec précaution si l'on ne tient pas compte de l'influence de l'effet de spectre qui nécessite la constitution d'échantillons de taille suffisamment grande pour pouvoir être stratifiés. Par ailleurs, cette performance globale dépend aussi de la composition des groupes. Elle peut être largement surestimée si les échantillons étudiés sont constitués d'une majorité de patients atteints d'une démence à un stade modéré ou sévère et de sujets volontaires sains qui ont, en général, des performances cognitives supérieures à celles de la population générale (Huppert, 2002). Ainsi, dans son étude originale de validation du MIS, Buschke *et al.*, (1999) mettent en évidence pour un seuil optimal du MIS à 4 (i.e. < 5), une sensibilité de 87% et une spécificité de 96% pour la détection des troubles cognitifs dans la MA. La VPP (probabilité d'avoir une MA pour un sujet ayant un score < 5) était de 85% pour une prévalence de la MA estimée à 20%. Cependant, la sévérité des troubles cognitifs des patients inclus dans cette étude (score moyen au Blessed dementia information memory concentration test à 14,7), en abaissant le seuil déficitaire du MIS, a probablement limité le nombre de cas faux positifs (sujets contrôles avec un score du MIS < 5) entraînant ainsi une surestimation de la spécificité et de la VPP. De plus, l'âge moyen, relativement élevé (81,1 ans) des patients inclus, la faiblesse de l'échantillon ($n=50$) et la forte proportion de patients atteints par la MA (78%) ne permettent pas la généralisation de ces résultats à l'ensemble du spectre clinique d'une population de consultants pour troubles cognitifs.

Dans notre premier article, nous avons tenté de limiter ce biais de sélection en incluant dans notre échantillon de patients 41% de cas atteints d'une forme légère de démence dont les performances aux tests étudiés ont été comparées à celles de sujets contrôles, sélectionnés par tirage au sort, parmi la clientèle de cabinets de médecine générale.

Ce travail nous permet de confirmer également l'intérêt d'associer ces deux tests pour une meilleure efficacité dans la détection des troubles cognitifs de la démence légère. En effet, nos résultats montrent que la combinaison MIS/STI avait une meilleure sensibilité que le MIS ou le STI utilisés de façon isolée, dans la détection des troubles cognitifs légers liés aux démences non-Alzheimer (73% vs 59%) et à la MA (96% vs 76%), respectivement. Si la sensibilité retrouvée du MIS pour détecter les troubles cognitifs légers de la MA est très bonne (94%), elle est en revanche très faible (59%) pour ceux rencontrés dans les autres démences. L'efficacité du MIS semble donc être limitée à la détection des troubles de la mémoire chez les patients atteints de MA. De ce fait, valider la pertinence d'un test par

rapport à un seul type de pathologie peut contribuer à entretenir la confusion entre le résultat positif à ce test et le diagnostic précoce de la maladie recherchée comme par exemple confondre un score déficitaire de mémoire et le diagnostic précoce de MA. En couvrant un plus large éventail de fonctions cognitives, l'association MIS/STI permet de compenser les limites de la réalisation isolée d'un test sans préjuger de l'étiologie du déficit cognitif objectif.

La sensibilité de cette combinaison de tests était également significativement supérieure à celle du MMSE (96% vs 86%) pour repérer les troubles cognitifs de la MA légère. Pourtant, le MMSE est un test composite qui comprend 30 items regroupés en 7 subtests (et autant de sous-scores) explorant un grand nombre de fonctions cognitives comme l'attention, le langage oral et écrit, le calcul, la mémoire, l'orientation temporelle et spatiale ainsi que les praxies constructives. Sa moins bonne sensibilité pourrait être due, non pas à une moindre efficacité du MMSE, mais au fait que seul son score global ait été analysé dans cette étude. Il aurait probablement été intéressant de préciser si certains sous-scores du MMSE permettaient de mieux discriminer les patients des sujets contrôles.

Le couple de tests MIS/STI présente des qualités métrologiques intéressantes obéissant aux critères d'un outil dit de dépistage selon l'OMS. C'est un outil simple, rapide (5 min.) et facile à coter. Une bonne reproductibilité (fidélité test-retest) a été observée pour le MIS ($r=0,78$) et pour le STI ($r=0,91$). Il présente de bonnes corrélations avec des épreuves administrées couramment en pratique neuropsychologique : la corrélation entre le MIS et le score total de l'épreuve de rappel libre/rappel indicé à 16 items (RL/RI-16) (Van der Linden, 2004), adaptation française du test d'apprentissage verbal de Grober et Buschke (1987) est de 0,70 et celle entre le STI et la forme B du Trail Making Test (Reitan, 1958) est de -0,76. Par contre, le MIS ($r=-0,20$) et le STI ($r=-0,22$) étaient sensibles à l'âge. Seul le STI était influencé par le niveau de scolarité ($r=0,30$). Le sexe des patients n'avait aucune influence sur la performance à ces deux tests. Cette corrélation du MIS et du STI avec des variables démographiques pose en tout cas la question de l'utilisation d'une valeur fixe comme seuil déficitaire (ou d'anormalité) et de son retentissement sur les indices de précision qui ont été présentés dans ce premier article.

La principale limite de cette étude cas-contrôles réside dans le choix de la population qui ne reflète pas suffisamment la diversité des cas cliniques rencontrés au sein d'une consultation

mémoire. La classification des sujets en deux états très distincts, avec démence (les cas) et sans troubles cognitifs (les contrôles), et la non-inclusion de cas MCI ont très certainement entraîné une surestimation de la performance globale de l'outil MIS/STI.

L'objectif du second article visait non pas à distinguer des patients MCI de sujets contrôles mais de pouvoir repérer au sein d'une cohorte de sujets plaintifs présentant un état MCI (ou Questionable Dementia) ceux qui vont évoluer vers une démence (les convertisseurs) et ceux qui vont rester stables (les non convertisseurs).

La sensibilité et la spécificité du couple MIS/STI étaient de 74% et 81%, respectivement. La performance de cette combinaison de tests était supérieure à celle du MMSE qui donnait une sensibilité similaire (76%) mais une plus faible spécificité (40%).

Tous les patients MCI inclus dans cette étude ont été suivis sur une période moyenne de 14,9 mois (de 6 à 24 mois). Les patients qui avaient bénéficié d'un suivi au-delà de 2 ans ont été exclus de cette étude. La combinaison MIS/STI de par sa simplicité et sa rapidité pourrait s'avérer effectivement être un outil utile dans la prédiction à court terme mais certainement pas à long terme de l'évolution de patients MCI.

Les résultats obtenus dans cette étude soulèvent un certain nombre de questions d'ordre méthodologique. La conversion rapide (inférieure à 12 mois) de certains patients pose le problème du degré de leur atteinte cognitive objectivé lors du bilan initial. Dans cette étude, compte tenu de l'absence d'utilisation de critères psychométriques stricts pour poser le diagnostic de MCI (d'où l'utilisation du terme « Questionable Dementia »), il y avait un risque, non négligeable, d'inclure dans la cohorte étudiée, des sujets de haut niveau scolaire qui, du fait de leur réserve cognitive, auraient pu être étiquetés à tort comme MCI alors qu'ils étaient déjà atteints d'une démence à un stade précoce ; l'inclusion de tels sujets étant susceptible de surestimer la sensibilité de notre outil. Afin de minimiser ce biais, les patients qui avaient développé une démence en moins de 6 mois n'ont donc pas été inclus. De plus, la comparaison des groupes de sujets convertisseurs selon la durée du suivi (< 12 mois vs \geq 12 mois) a montré des performances cognitives initiales similaires pour l'ensemble des tests de la batterie neuropsychologique RAPID (excepté pour l'épreuve de dénomination orale).

Un autre argument qui va dans le sens d'une possible surestimation de la sensibilité de l'outil MIS/STI est la faible proportion de patients MCI suivis (29%). On peut supposer que ces patients n'ont pas bénéficié d'un bilan de contrôle, soit parce qu'ils n'avaient plus de

plaintes cognitives, soit parce qu'ils avaient été rassurés par les conclusions du praticien. A l'inverse, on peut légitimement penser que des patients MCI qui ont bénéficié d'un suivi l'ont été en raison de la persistance des plaintes et/ou de la suspicion du clinicien envers une possible conversion future. Pour des questions d'ordre éthique, il nous a été impossible de contacter les patients afin de vérifier ces informations. Toutefois, les motifs invoqués plus haut nous ont conduits à penser que l'atteinte cognitive initiale des patients suivis serait en moyenne plus sévère que celle des patients qui n'avaient pas été revus en consultation. Cependant, la comparaison des performances cognitives initiales n'a montré aucune différence significative entre ces deux groupes. Il n'y avait pas non plus de différence au niveau de la moyenne d'âge qui est un facteur de risque accru de conversion vers la démence (McCullagh *et al.*, 2001). Enfin, afin de contrôler que nos patients MCI se situaient bien sur un continuum cognitif, nous nous sommes assurés que l'ensemble des scores aux tests de la batterie neuropsychologique RAPID était significativement supérieur par comparaison avec ceux de patients atteints de démence légère et inférieures avec ceux de sujets plaintifs sans troubles.

Il est important en revanche de souligner que les sujets MCI de cette étude ont été recrutés lors de consultations mémoire, dont la prévalence de la démence est bien plus élevée qu'en population générale. Ceci pourrait expliquer le taux élevé de conversion vers la démence de nos patients MCI (36%) par rapport à celui habituellement indiqué dans la population générale (de 10 à 15% par an). Il serait donc intéressant de déterminer dans quelle mesure la diversité du contexte de recrutement des patients pourrait faire varier la sensibilité et la spécificité de notre outil.

Il est curieux de constater qu'on retrouve les mêmes valeurs seuils optimales pour notre cohorte de sujets MCI que dans l'étude cas-contrôles présenté dans le premier article (Chopard *et al.*, 2007). Pour ces valeurs (MIS<6 et/ou STI<25), la sensibilité (74%) et la spécificité (81%) sont plutôt modestes. La spécificité est importante cliniquement, car un résultat faussement positif peut causer des inquiétudes chez le patient et entraîner parfois des examens approfondis inutilement coûteux.

Pour tenter d'améliorer la précision de l'outil MIS/STI, nous avons proposé l'utilisation de « valeurs limites » qui sont des seuils en dessous ou au-delà desquels il n'y a plus de faux positifs ou de faux négatifs, respectivement. Ainsi, seuls des patients

convertisseurs (n=5) présentaient un résultat déficitaire aux deux tests (MIS<6 et STI<25) et seuls des patients non-convertisseurs (n=11) obtenaient un score supérieur aux valeurs suivantes : MIS>7 et STI>29. Si ces valeurs limites ont permis d'atteindre une précision de 100% (16/16), elles n'ont classé qu'un très petit nombre de patients (16/106). Une autre technique, basée sur le « retest » a été proposée par Fujji *et al.* (2003) pour limiter le nombre de cas faux positifs. Ces auteurs ont constaté qu'en administrant à nouveau le même test (Cognitive Assessment Screening Instrument) aux mêmes sujets après un intervalle de 2 à 3 mois permettait une réduction de 20,2% des cas faux positifs (i.e. des patients suspectés à tort d'avoir des troubles cognitifs). Cette condition de retest est pourtant très souvent considérée comme une source d'erreur de la mesure car l'augmentation de la performance cognitive après répétition du même test (ou effet d'apprentissage), peut biaiser l'interprétation du score de retest (MacCaffrey *et al.*, 2000). Néanmoins, de nombreux auteurs se sont intéressés aux liens qui peuvent exister entre l'effet d'apprentissage et le statut cognitif de patients (Darby *et al.*, 2002 ; Cooper *et al.*, 2004 ; Duff *et al.*, 2008,). Bien que certaines études s'accordent pour affirmer la présence d'un effet d'apprentissage chez les sujets sains (MacCaffrey *et al.*, 2000) et sa diminution voire son absence chez les sujets atteints de démence (Cooper *et al.*, 2001), très peu se sont intéressées aux capacités d'apprentissage chez les sujets MCI (Cooper *et al.*, 2004 ; Duff *et al.*, 2008). Une étude récente menée par Duff *et al.* (2011) a montré qu'en répétant la même version d'une batterie de tests cognitifs à 1 semaine, les sujets MCI qui ne bénéficiaient pas d'un effet d'apprentissage présentaient un déclin cognitif après un 1 an de suivi ; ceux qui avaient bénéficié d'un effet d'apprentissage avaient obtenu par contre de bien meilleures performances à 1 an. Les mesures du niveau d'effet d'apprentissage pourraient donc constituer un marqueur intéressant du devenir cognitif des sujets MCI.

Dans notre premier article, nos résultats avaient montré que les performances du MIS et de du STI étaient sensibles aux effets de l'âge, mais aussi à celui du niveau de scolarité pour le STI seulement. Une des limites de cette seconde étude est de ne pas avoir pu mesurer, par manque d'effectif, l'effet de spectre sur la performance de notre outil en tenant compte des caractéristiques démographiques des patients MCI.

Dans le troisième article, nous avons pris soins de définir dans un premier temps les différents types de MCIa selon des critères cliniques mais aussi psychométriques. Ainsi le statut cognitif de chaque participant a été établi par rapport aux normes de la batterie de tests neuropsychologiques RAPID (Ferreira *et al.*, 2010). Tous les tests de cette batterie ont été

étalonnés selon l'âge et le niveau scolaire, à partir d'un même échantillon de sujets témoins, dans des conditions les plus proches possibles de celles des patients qui ont été sélectionnés dans cette étude (tests effectués en milieu médical, selon le même ordre de passation). Par ailleurs, nous avons souhaité également évaluer les performances de l'outil MIS/STI, non pas dans sa capacité à séparer deux groupes bien distincts (MCIa vs sujets volontaires sains) mais à repérer des troubles légers et très légers de la mémoire associés ou non à d'autres déficits cognitifs, parmi un large échantillon de sujets (n=2490) exprimant une plainte cognitive et vus au sein d'une consultation mémoire.

Cette étude a souligné dans un premier temps un effet de spectre sur les performances de l'outil MIS/STI dans la population étudiée. Cet effet de spectre s'est traduit par une variation très importante de la sensibilité (0% à 94%) et de la spécificité (52% à 95%) selon l'âge et le niveau scolaire en fonction du type de MCIa et du degré de l'atteinte cognitive. Par ailleurs, les sensibilités et spécificités les plus élevées ayant été retrouvées dans le groupe de patients MCI amnésique multidomaines (MCIa-MD ; aMCI-m en anglais dans le 3^{ème} article), l'impact du biais de spectre a donc été évalué sur cette seule population de patients.

Nous avons pour cela étudié les conséquences de cet effet de spectre sur la probabilité post-test d'avoir ou pas des troubles cognitifs liés à cet état MCIa-MD. Pour des seuils constants (i.e. seuils uniques déterminés à partir de la population globale MCIa-MD vs population globale de sujets contrôles), les RVP et les RVN ont varié de façon importante en fonction de l'âge et du niveau scolaire. La présence d'un biais spectral suggère donc que la performance globale d'un test, déterminée sur un échantillon unique non stratifié, ne serait qu'une approximation et exposerait à un risque de sous-estimer ou de surestimer la présence de troubles cognitifs dans certains sous-groupes de cet échantillon, et d'affecter ainsi l'interprétation clinique de son résultat. Par exemple, pour des sujets âgés de 80 à 89 ans, l'application des seuils constants donnait un RVN de 0,16 correspondant à une sensibilité et une spécificité de 90% et 61%, respectivement. Si la probabilité pré-test d'avoir des troubles cognitifs pour un patient de cette tranche d'âge était estimée à 40%, la probabilité post-test diminuerait à 9,7% en cas de résultat négatif au couple de tests MIS/STI. Néanmoins, avec un même RVN (i.e. égal à 0,16) mais obtenu à partir d'une sensibilité (87%) et d'une spécificité (83%) nettement plus équilibrées, la probabilité post-test resterait, certes, inchangée, mais la proportion de sujets sans troubles cognitifs correctement identifiés comme tels aurait été bien plus élevée (83% vs 61%) limitant ainsi le taux de faux positifs (17% vs 39%). Ce constat met

en lumière les problèmes d'articulations entre l'interprétation des performances d'un test au niveau psychométrique et son application au niveau clinique.

Nos résultats montrent notamment que l'usage d'une valeur seuil fixe sur des populations MCIa-MD, d'âge et de niveaux scolaires différents, augmente sensiblement la proportion de faux négatifs chez les sujets de haut niveau scolaire, âgés de 60 à 69 ans, et de faux positifs chez les sujets plus âgés de 80 à 89 ans, quel que soit leur niveau d'éducation. L'application de seuils ajustés à ces deux variables démographiques permet de limiter l'effet de spectre observé dans ces sous-groupes. Ces seuils ajustés diminuent notamment de façon considérable le taux de faux positifs chez les sujets les plus âgés. Cette limitation entraîne des conséquences sur le plan clinique. A titre d'exemple, pour un patient âgé de 81 ans, titulaire d'un certificat d'aptitudes professionnel (niveau scolaire moyen), chez qui la probabilité pré-test d'avoir des troubles cognitifs serait estimée à 20%, la probabilité post-test en cas de résultat positif au MIS ou au STI, serait de 36% avec une valeur-seuil fixe ($RVP=2,29$) et de 71% avec une valeur seuil ajustée ($RVP=9,87$). On peut voir ici l'intérêt d'utiliser des seuils ajustés, qui en augmentant le gain de certitude, peuvent faciliter la prise de décision clinique, d'exams complémentaires en particulier, chez des personnes âgées.

Parmi les limites de cette étude, on peut citer la faiblesse de l'échantillon ($n=133$) des sujets présentant un trouble mnésique isolé très léger même si les résultats obtenus semblent indiquer clairement que l'outil étudié a une plus grande efficacité pour le seul groupe MCIa-MD. Par ailleurs, nous n'avons pas inclus dans ce travail, des patients MCI non mnésique qui présentent pourtant aussi un risque d'évoluer vers une démence (Busse *et al.*, 2006). Enfin, il faut préciser le caractère non prospectif de cette étude.

Cependant, une analyse rétrospective des patients ayant bénéficié d'un suivi ($n=321$) a montré une variation du taux de conversion vers la démence selon le type de MCIa. Pour une période de suivi moyen de 24 mois (12-54 mois), le taux de conversion vers la démence était de 24% pour les MCIa-MD, 9% pour les MCIa-MD avec atteinte très légère et seulement 1,4% pour les MCIa présentant un trouble mnésique isolé léger ou très léger. Le caractère diffus de l'atteinte cognitive, qu'elle soit légère ou très légère, semble donc constituer un facteur de risque accru de conversion vers la démence chez les cas MCIa. Il est intéressant de noter également qu'après un suivi moyen de 32 mois (12-89 mois), plus d'un quart (27%) des sujets avec plainte sans troubles cognitifs avaient évolué vers un des quatre états MCIa définis

dans cette étude, tandis que 1,9% avaient progressé vers une démence. Ces résultats signifient que pour une petite proportion de ces sujets sans troubles cognitifs, la plainte subjective pourrait être le témoignage d'un déclin cognitif non encore décelable par les tests en l'absence d'une évaluation antérieure.

L'analyse rétrospective des scores du MIS et du STI relevés au cours du suivi de ces patients a montré qu'en utilisant des scores-seuils ajustés, 82% des patients sans troubles cognitifs ou MCIa-MD restés stables ou redevenus normaux, auraient été correctement classés (vs 73% pour un score seuil fixe) et 86% (vs 94% pour un seuil fixe) de ceux qui avaient évolué vers un état MCIa-MD ou une démence, respectivement. Ce résultat souligne l'importance de la surveillance (Wilson et Jungner, 1970) consistant à pratiquer des tests par intervalles comme c'est le cas par exemple des examens qui sont faits régulièrement pour dépister le cancer du sein ou du côlon. Le couple de tests étudié dans ce travail pourrait donc aussi constituer, au même titre que ces examens de dépistage, un outil permettant au praticien le suivi de ces patients sur le plan cognitif.

Une étude prospective menée par Saxton *et al.* (2004) a retrouvé, à l'inverse des tests d'évocation phonologique (donner en 1 min. des mots communs commençant par une lettre donnée), une altération de la fluence verbale catégorielle (donner des noms d'animaux en 1 min) chez des patients, 3 à 5 ans avant l'apparition d'une MA. Au vu de ce résultat, on peut se demander si l'utilisation de cette tâche d'évocation catégorielle plutôt que celle du STI eut été plus pertinente pour la détection de nos patients MCIa qui ont évolué vers une démence. On peut supposer en effet, que les patients convertisseurs épuiseraient plus facilement leur stock sur un temps plus long (1 ou 2 min. vs 15 sec. pour le STI). Afin de limiter les effets de la contrainte temporelle du STI sur la performance des patients, nous avons modifié les consignes initiales de cette épreuve pouvant induire un effet plafond (10 items max. par catégorie soit un score global max. de 40) en acceptant toutes les réponses produites durant les 15 sec (Thomas Antérion *et al.*, 2001).

Par ailleurs, la contrainte temporelle du STI pourrait également perturber et pénaliser les sujets contrôles anxio-dépressifs (par un effet inhibiteur ?) et générer ainsi des cas faux positifs. Toutefois, si l'on compare le score total moyen du STI, les sujets contrôles dépressifs (n=64) de notre étude (article 3) ont produit autant de mots (moy.=35,67 ; écart-type=6,09 ; min.=25, max.=57) que les sujets contrôles non dépressifs (n=926 ; moy.=35,90 ; écart-

type=6,58 ; min.=15, max.=63) avec une proportion moins importante de cas faux positifs (15,6% vs 19%).

Ces trois articles nous ont donc permis de préciser quels étaient les qualités métrologiques nécessaires pour que, dans le domaine cognitif, un test de repérage puisse être considéré comme ayant les mêmes critères qu'un outil de dépistage. Si ces critères restent d'actualité (Gifford et Cummings, 1999) : rapide, simple d'utilisation, facile à coter, reproductible et valide, sensible et spécifique ; la preuve de la qualité et de l'utilité d'un test d'évaluation cognitive doit être vérifiée méthodologiquement. Ainsi, les niveaux de sensibilité et de spécificité, parfois élevés de certains tests, tels qu'ils sont présentés dans la littérature, doivent être considérés avec précaution, car ils vont dépendre de la taille des échantillons étudiés, de la sévérité des troubles cognitifs des patients inclus mais aussi de la population des sujets contrôles utilisée. Dans nos deux premiers articles, le schéma cas/contrôle utilisé (Chopard *et al.*, 2007) et le petit nombre de patients MCI suivis (Chopard *et al.*, 2009) constituent probablement des biais qui ont pu influencer nos résultats. De la même manière, nous avons constaté qu'il était important de garder en tête que les valeurs de prédiction (VPP ou VPN) donnent une estimation quantitative de la probabilité d'avoir la condition (maladie ou anomalie) pour une prévalence donnée mais qu'elles ne donnent aucune certitude sur le plan clinique. Par exemple, les bonnes VPP retrouvées par Buschke *et al.*, (1999) dans son étude de validation du MIS étaient liées à l'inclusion de patients atteints de démence à un stade avancé. De plus, l'âge moyen (81,4 ans), relativement élevé de ces patients fait que sa population d'étude se rapproche plus des caractéristiques de personnes hospitalisées en gériatrie et ne reflètent donc pas le spectre beaucoup plus vaste des patients vus dans la pratique ambulatoire. De ce fait, les valeurs de prédiction devraient n'être utilisées que pour cibler les populations à partir desquelles elles ont été déterminées et dans ce sens, il importe d'être attentif à ne pas faire d'extrapolations excessives.

Un autre point développé dans ce 3^{ème} article a permis de souligner les limites de l'utilisation d'une valeur fixe comme seuil de détection des patients MCIa-MD. Nos résultats ont permis de constater en effet que la détermination d'une valeur seuil optimale, dérivée d'une population « moyenne », non stratifiée, pouvait faire croire, à tort, que l'apport d'un test en terme de décision clinique (apport fonction de la valeur des RV) était identique pour tous les sujets ou groupes de sujets appartenant à cette population (Goerhing *et al.*, 2004, Mulherin et Miller, 2002). Par conséquent, le choix d'une valeur seuil devrait être basée non pas sur son

efficacité globale dans un échantillon, quelle que soit sa taille, mais sur les conséquences relatives d'un nombre trop élevé de faux positifs ou de faux négatifs sur un échantillon suffisamment large de patients et de sujets contrôles.

A titre d'exemple, le MOCA (Nasreddine *et al.*, 2005) est considéré comme un instrument particulièrement intéressant pour distinguer des patients MCI de sujets contrôles. Nasreddine *et al.* (2005) dans son travail initial de validation ont mis en évidence une sensibilité de 90% (94 patients MCI) et une spécificité de 87% (90 sujets volontaires sains sans plainte cognitive) pour un score seuil fixe de 26, avec des RVP et RVN respectifs de 6,92 et 0,11. Le MOCA a fait l'objet de très nombreuses traductions et est désormais largement utilisé pour la détection des troubles cognitifs dans le cadre de nombreuses affections comme la MA (Nasreddine *et al.*, 2005), la maladie de Parkinson (Dalrymple-Alford *et al.*, 2010) ou la Chorée de Huntington (Videnovic *et al.*, 2010). Toutefois, l'influence de facteurs démographiques sur les résultats du MOCA constitue une limite de cet outil. En effet, une étude très récente (Freitas *et al.*, 2012) a montré une influence importante des variables d'âge mais également du niveau d'éducation sur le score total de ce test. Ce résultat encourage donc l'étude de la variabilité des paramètres du MOCA en fonction de ces facteurs démographiques voire l'utilisation d'un ajustement de la valeur seuil afin d'éviter de classer par exemple des personnes âgées dont le niveau d'éducation est faible, comme de faux positifs.

Compte tenu de ces constatations, il nous paraît donc important, avant de recommander l'utilisation d'un test, de le valider sur de plus grands échantillons qui permettent d'évaluer sa capacité à discriminer des sous-groupes d'individus les plus proches possibles afin de se placer au plus près de la position du clinicien dans sa pratique courante. En effet, les sujets contrôles volontaires sains ont moins de risque d'avoir un test positif que des sujets contrôles qui présenteraient des symptômes similaires à ceux des patients étudiés (Rutjes *et al.*, 2005) comme l'expression d'une plainte mnésique. De la même manière, des patients atteints d'une démence avérée ont moins de risque d'avoir un test négatif que des patients MCI dont l'atteinte cognitive est beaucoup moins sévère. Dans cette optique, l'étude de l'effet de spectre et du biais spectral sur les performances d'un test devrait faire partie des étapes obligatoires de sa validation.

Dans notre 3^{ème} article, nous sommes bien conscients que les variables démographiques (âge et niveau scolaire) et le type de MCIa étudiés ne sont peut-être pas les seuls éléments

susceptibles de faire varier la performance de notre outil et son interprétation. Il existe, en effet, une grande hétérogénéité dans l'évolution des troubles cognitifs des patients MCI. Certains vont voir leurs performances se dégrader au fil des mois ou des années (Busse *et al.*, 2006) alors que d'autres vont rester stables voire même se normaliser (De Rotrou *et al.*, 2005). Ces trajectoires différentes sont probablement liées à de nombreux autres facteurs pouvant intervenir dans l'effet de spectre. Il conviendrait notamment de s'interroger sur la qualité du sommeil du patient (Beaulieu-Bonnon et Hudon, 2009), ses traitements notamment par psychotropes (Moore et Keefe, 1999 ; Barker *et al.*, 2004) en cours ou passés, son degré d'anxiété au moment de l'examen (Devier *et al.*, 2009), sa co-morbidité (Stephan *et al.*, 2011), son parcours professionnel (qui n'est pas forcément lié au niveau scolaire) (Marquier *et al.*, 2010), ses loisirs (Verghese *et al.*, 2003 ; Verghese *et al.*, 2006) ainsi que le contexte de soins dans lequel il est évalué (cabinet de médecine générale, centre de consultation mémoire) (Knottnerus et Leffers, 1992). Ces facteurs pourraient contribuer également aux variations des performances d'un test de détection des troubles cognitifs.

Ainsi, l'analyse de l'effet de spectre et du biais spectral sur la performance cognitive, permet de prendre en compte la singularité des individus évalués, la diversité étiologique des troubles cognitifs et la complexité de l'interprétation d'un test. Or, le dépistage, en se focalisant sur la recherche systématique d'une maladie spécifique comme la MA, ne permet sûrement pas de faire face aux différentes situations cliniques des patients concernés par les troubles cognitifs. Si la MA constitue la cause la plus fréquente des maladies neurodégénératives, il serait dommage de ne s'intéresser qu'à ce type de pathologie alors qu'il existe de nombreuses autres étiologies potentiellement curables (Weytingh *et al.*, 1995) (perturbations du sommeil, iatrogénie médicamenteuse, facteurs cardio-vasculaires, état anxio-dépressif, carences vitaminiques, troubles endocriniens) pour lesquelles nous disposons d'un traitement efficace ou de moyens de prévention (Duron et Hanon, 2008, Panza *et al.*, 2008, Vercambre *et al.*, 2012, Strange et Backman, 2012). Par exemple, les troubles du sommeil sont très fréquents dans la population générale (environ 30%) et augmentent sensiblement après 65 ans (50%) (Duran *et al.*, 2001) mais ils sont mal repérés puisque seulement 20% des individus ayant des troubles du sommeil sont correctement diagnostiqués (Ohayon, 2007). Il est par ailleurs important de préciser que des troubles du sommeil non décelés chez la personne âgée, comme le syndrome d'apnées du sommeil, peuvent parfois engendrer d'authentiques troubles de la concentration, de la mémoire, du raisonnement et des

fonctions exécutives avec augmentation du risque d'accidents routiers ou domestiques (Fulda et Schulz, 2001, Ohayon et Vecchierini, 2002, Ohayon, 2007, Lal *et al.*, 2012). Ces troubles cognitifs sont totalement ou partiellement réversibles après un traitement adapté (Michel et Sellal, 2011).

Dans le chapitre introductif de ce travail, nous avons rappelé la définition du dépistage selon l'OMS comme un moyen visant à identifier la présence d'une maladie ou d'une anomalie dans une population asymptomatique, afin d'accroître le taux de diagnostic précoce et le pronostic de la maladie ou de l'anomalie recherchée. Nous avons déjà évoqué au cours de ce travail l'ambiguïté du terme dépistage, et la confusion possible entre le dépistage d'un trouble cognitif et le diagnostic précoce de la MA. Nous avons essayé de montrer combien la volonté d'établir un diagnostic précoce de la MA en considérant le MCI comme une phase prodromique de cette maladie, avait pu générer une telle ambiguïté.

La population étudiée dans le 3^{ème} article de cette thèse incluait exclusivement des patients et des sujets contrôles qui présentaient une plainte cognitive. Or, à partir du moment où des symptômes sont exprimés, ce n'est plus du dépistage proprement dit. Une définition stricte du concept de dépistage n'est probablement pas appropriée dans le domaine des troubles cognitifs, et en particulier dans le vaste champ des démences (Gallez, 2005). Par ailleurs, l'effet de spectre objectivé dans ce travail, même s'il n'a pas pris en compte la complexité et la dynamique de nombreux facteurs, confirme bien en tout cas la nécessité d'aborder les troubles cognitifs en termes de continuum (Hachinski, 2008) et non pas sur la base de catégories malades/non malades. Il n'est donc pas légitime de proposer un dépistage tel qu'il est décrit par l'OMS et dans lequel des personnes asymptomatiques de la population générale âgées de 75 ans ou plus seraient invitées à passer un test de dépistage de la MA à titre systématique (Ashford *et al.*, 2006). Prenons pour exemple, l'examen du dépistage du cancer colorectal, où toute personne est conviée par courrier à partir de 50 ans à retirer un test Hemocult[®] qui permet de détecter des traces de sang dans les selles. La personne dépistée reçoit ensuite le résultat par courrier et est invitée à consulter son médecin traitant qui lui proposera, en cas de test positif, la réalisation d'un examen diagnostique par coloscopie. Dans cet exemple, on voit bien que le dépistage est une procédure de tri avant l'expertise diagnostique d'une maladie. Il est caractérisé par une approche dichotomique entre résultat positif et résultat négatif d'un test. Un dépistage positif oriente le patient vers des examens complémentaires de diagnostic de la maladie recherchée. De ce fait, le dépistage des troubles

cognitifs est souvent considéré à tort comme la première étape vers un diagnostic précoce de la MA (Dubois *et al.*, 2002) ou d'une démence (Chopard *et al.*, 2003). Dans ces conditions, il apparaît pour le moins difficile de vouloir pratiquer un dépistage des troubles cognitifs de la MA dans la population générale alors que plusieurs études (Patterson *et al.*, 1999 ; White *et al.*, 2002) ont démontré qu'il pouvait donner un taux très élevé de cas faux positifs (86 à 93%) susceptibles d'être orienté vers une procédure diagnostique inutile. C'est la raison pour laquelle nous proposons plutôt d'utiliser le terme de « repérage » (Gallez, 2005) qui, à la différence du dépistage, ne préjuge pas d'une étiologie particulière. Ce terme, rappelons-le, en anglais, se traduit de la même manière que dépistage (i.e. « screening »), mais renvoie tout de même à une pratique différente. Le repérage des troubles cognitifs s'adresse à un patient qui présente une plainte cognitive et/ou comportementale (émanant du patient lui-même ou de son entourage) et peut se concevoir comme un processus à plusieurs étapes. A titre d'exemple, en cas de résultats strictement normaux aux deux tests (MIS et STI), une évaluation comparative pourrait être proposée au patient dans le cadre d'un suivi 6 mois plus tard à la recherche d'une éventuelle modification cognitive après ce laps de temps. En cas de résultat positif (au moins 1 des 2 tests déficitaires), il nous paraît plus judicieux pour le médecin (qui connaît bien les antécédents de son patient) de rechercher et de traiter dans un premier temps les causes potentiellement curables suspectées d'être à l'origine des troubles cognitifs plutôt que d'orienter systématiquement son patient vers un centre de consultation spécialisée pour y réaliser des examens coûteux voire stressant pour certaines personnes (Grimes et Schultz, 2002 ; Parker et Philp, 2004). A la suite de cette (éventuelle) première prise en charge (traitement de la cause curable), le médecin pourrait réaliser une seconde évaluation à 1 ou 2 mois : si le résultat du test s'avère de nouveau positif, le patient pourrait alors bénéficier d'un bilan neuropsychologique complet voire d'autres examens complémentaires au sein d'une consultation mémoire. En cas de résultat négatif aux 2 tests, un contrôle à 6 mois pourrait être proposé (voir algorithme annexe 12).

Dans le troisième article de cette étude, l'utilisation de valeurs seuils ajustées pour le couple MIS-STI a montré des spécificités beaucoup plus élevées que celles retrouvées pour des valeurs seuils fixes. Si la sensibilité d'un test est souvent privilégiée dans le cadre de la recherche d'une maladie grave mais potentiellement curable comme pour certains cancers, les seuils ajustés semblent plus appropriés pour du repérage des troubles cognitifs. En effet, il semble important de rassurer avant tout un patient sain qui se plaint de sa mémoire avec le

moins de risque d'erreur possible. Par ailleurs, nous avons déjà souligné dans ce travail la diversité d'évolution du MCI qui peut varier dans le temps, s'aggraver ou s'améliorer (Busse *et al.*, 2006) et que cette évolution ne préjuge en aucun cas de son étiologie (Petersen, 2007).

En conclusion, nous proposons de procéder à un repérage des troubles cognitifs chez tous les sujets exprimant une plainte cognitive (plainte spontanée du patient ou rapportée par son entourage) comme cela a été évoqué dans certaines publications (Ashford *et al.*, 2006 ; Ashford *et al.*, 2007). Il pourrait aussi être envisagé chez des personnes certes non plaintives mais qui présentent des facteurs cumulés de risque de déclin cognitif (hypertension artérielle, diabète, accident vasculaire) (Li *et al.*, 2011) ou des signes d'alerte comme la répétition de chutes (Muir *et al.*, 2012), d'accidents de la route (Drachman et Swearer, 1993) ou encore une modification récente et inexplicée de l'état thymique (dépression, perte d'intérêt) (Stepaniuk *et al.*, 2008). Cette approche, réalisée au cas par cas, à l'initiative du médecin traitant, est à l'opposé d'une véritable stratégie épidémiologique de dépistage systématique d'une MA. Elle permettrait de hiérarchiser la prise en charge des patients afin de mieux cibler ceux pouvant être traités dans le cadre de la médecine générale et le cas échéant ceux devant bénéficier d'une consultation mémoire en milieu spécialisé.

Enfin, il convient pour le clinicien non pas de s'interroger d'abord sur le choix de l'utilisation d'un outil plutôt qu'un autre, mais de bien connaître ses limites, savoir comment il a été validé et s'il peut l'utiliser à bon escient pour la population qui le consulte au cours de sa pratique quotidienne. Dans ce cadre, il nous semble intéressant de privilégier l'utilisation d'un outil bref et validé qui permet d'assurer le suivi régulier d'un patient, en tenant compte de sa singularité.

BIBLIOGRAPHIE

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental disorders. 4th ed. Washington DC: 1994.
- Amieva, H., Jacquin-Gadda, H., Orgogozo, J.M., Le Carret, N., Helmer, C., Letenneur, L. et al. (2005). The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain* 128, 1093-1101.
- Amieva, H., Le Goff, M., Millet, X., Orgogozo, J.M., Pérès, K., Barberger-Gateau, P. et al. (2008). Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. *Ann Neurol* 64, 492-498.
- Anaes. Recommandations pratiques pour le diagnostic de la maladie d'Alzheimer. Février 2000. www.has.fr.
- Anthony J.C., Le Resche L., Niaz U., Von Korff M.R., Folstein M.F. (1982). Limits of the "Mini Mental State" as a screening test for dementia and delirium among hospital patients. *Psychological Medicine* 12, 397-408.
- Ankri J. (2006). Epidémiologie des démences et de la maladie d'Alzheimer. *BEH*, 5-6, 42-44.
- Ashford, J.W., Borson, S., O'Hara, R., Dash, P., Frank, L., Robert, P., et al. (2006). Should older adults be screened for dementia? *Alzheimers Dement* 2, 76-85.
- Ashford, J.W., Borson, S., O'Hara, R., Dash, P., Frank, L., Robert, P., et al. (2007). Should older adults be screened for dementia? It is important to screen for evidence of dementia! *Alzheimers Dement* 3, 75-80.
- Ayala, A.R., Ilias, I., Nieman L.K. (2003). The spectrum effect in the evaluation of Cushing syndrome. *Curr Opin Endocrinol Diabetes* 10, 272-276.
- Barker, M.J., Greenwood, K.M., Jackson, M., Crowe, S.F. (2004) Cognitive effects of long-term benzodiazepine use: a meta-analysis. *CNS Drugs* 18, 37-48.
- Barnes, D.E., Covinsky, K.E. Whitner, R.A., Kuller, L.H., Lopez, O.L., Yaffe, M.D. (2009). Predicting risk of dementia in older adults. The late-life dementia risk index. *Neurology* 73, 173-179.
- Bayes T. (1763). An Essay Toward Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London* 53, 370-418.
- Beaulieu-Bonneau, S., Hudonc, C. (2009). Sleep disturbances in older adults with mild cognitive impairment. *Int Psychogeriatr* 21, 654-666.
- Begg, C.B. (1987). Biases in the assessment of diagnostic tests. *Statist Med* 6, 411-423.

- Belmin, J., Pariel-Madjilessi, S., Surun, P., Bentot, C., Feteanu, D., Lefebvre des Noettes, V., Onen, F. (2007). The cognitive disorders examination (Codex) is a reliable 3-minute test for detection of dementia in the elderly (validation study on 323 subjects). *Presse Med* 36, 1183-1190.
- Benoît, M., Brocher, P., Clement, J.P., Cnockaert, X., Hinault, P., Nourashemi, M.P. et al. (2005). Les symptômes psychologiques et comportementaux de la démence : description et prise en charge. *Rev Neurol* 161, 357-366.
- Blessed, G., Tomlinson, B.E., Roth, M. (1968). The Association Between Quantitative Measures of Dementia and of Senile Change in the Cerebral Grey Matter of Elderly Subjects. *BJP* 114, 797-811.
- Borson, S., Scanlan, J.M., Chen, P., Ganguli, M. (2003). The Mini-Cog as a screen for dementia: validation in a population-based sample. *J Am Geriatr Soc* 51, 1451-1454.
- Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., et al. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Standards for Reporting of Diagnostic Accuracy. Clin Chem* 49, 1-6.
- Boustani, M., Peterson, B., Hanson, L., Harris, R., Lohr, K.N. (2003). Screening for dementia in primary care: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 138, 927-941.
- Braak, H., Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol Aging* 16, 271-278.
- Brayne, C., Fox C., Boustani M. (2007). Dementia Screening in Primary: is it time? *JAMA* 298, 2409-2411.
- Brenner, H., Gefeller, O. (1997). Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med* 16, 981-991.
- Brodaty, H., Moore, C.M. (1997). The clock drawing test for dementia of the Alzheimer's type: a comparison of three scoring methods in memory disorders clinic. *Int J Geriatr Psychiatry* 12, 619-627.
- Brodaty, H., Clarke, J., Ganguli, M., Grek, A., Jorm, A.F., Khachaturian, Z., et al. (1998). Screening for cognitive impairment in general practice: toward a consensus. *Alzheimer Dis Assoc Disord* 12, 1-13.
- Brodaty, H., Clarke, J., Ganguli, M., Grek, A., Jorm, A.F., Khachaturian, Z., et al. (2002). The GPCOG: a new test for dementia designed for general practice. *J Am Geriatr Soc* 50, 530-534.

- Brodaty, H., Low, L.F., Gibson, L., Burns, K. (2006). What is the best dementia screening instrument for general practitioners to use? *Am J Geriatr Psychiatry* 14, 391-400.
- Brooks, B.L., Iverson, G.L., White, T. (2007). Substantial risk of "Accidental MCI" in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *J Int Neuropsychol Soc* 13, 490-500.
- Bulpitt, C.J. (1987). Confidence Intervals. *Lancet* 329, 494-497.
- Buschke, H., Kuslansky, G., Katz, M., Stewart, W.F., Sliwinski, M.J., Eckholdt, H.M., et al. (1999). Screening for dementia with the Memory Impairment Screen. *Neurology* 52, 231-231.
- Busse, A., Hensel, A., Gühne, U., Angermeyer, M.C., Riedel-Heller, S.G. (2006). Mild cognitive impairment: long-term course of four clinical subtypes. *Neurology* 67, 2176-2185.
- Calahan, C.M., Boustani, M.A., Inverzagt, F.M. (2006). Effectiveness of collaborative care for older adults with Alzheimer disease in primary care: a randomized controlled trial. *JAMA* 295, 2148-2157.
- Canguilhem, G. (1952). Le normal et le pathologique. *In: Canguilhem G. (2003). La connaissance de la vie. J Vrin.*
- Carcaillon, L., Amieva H., Auriacombe, S., Helmer, C., Dartigues, J.F. (2009). A substest of the MMSE as a valid test of episodic memory? Comparison with the Free and Cued Reminding Test. *Dement Geriatr Cogn Disord* 27, 429-438.
- Centor, R.M., Witherspoon, J.M., Dalton, H.P., Brody, C.E., Link, K. (1981). The diagnosis of strep throat in adults in the emergency room. *Med Decis Making* 1, 239-246.
- Chehouilli, P. (2001). La plainte mnésique du sujet âgé. *La revue du généraliste et de la gerontologie* 8, 413-419.
- Chevalier, P. (2011). Précision d'un test diagnostique et prévalence. *Minerva* 10, 51.
- Choi, B.C.K., Pak, A.W.P. (2004). Understanding and Minimizing Epidemiologic Bias in Public Health Research. *Can J public Health* 96, 284-286.
- Chopard, G., Morizot, C., De Murcia, D., North, C., Perrot, P., Sevrin, C. et al. (2003). Comparaison de 4 outils de détection des démences : première étape du diagnostic précoce. *Rev Fr de Ger et Gerontol* 10, 6-11.
- Chopard, G., Pitard, A., Ferreira, S., Vanholsbeeck, G., Rumbach, L., Galmiche, J. (2007). Combining the Memory Impairment Screen and the Isaacs Set Test: a practical tool for screening dementias. *J Am Geriatr Soc* 55, 1426-1430.

- Chopard, G., Vanholsbeeck, G., Tio, G., Pitard, A., Binetruy, M., Rumbach, L., et al. (2009). Rapid Screening of Cognitive Change in Patients with Questionable Dementia using the Memory Impairment Screen and the Isaacs Set Test. *J Am Geriatr Soc* 57, 703-708.
- Cooper, D.B., Lacritz, L.H., Weiner, M.F., Rosenberg, R.N., Cullum, C.M. (2004). Category Fluency in Mild Cognitive Impairment. Reduced Effect of Practice in Test-Retest Conditions. *Alzheimer Dis Assoc Disord* 18,120-122.
- Cooper, D.B., Epker, M., Lacritz, L. Weine, M., Rosenberg, R.N. Honig, et al. (2001). Effects of practice on category Fluency in Alzheimer's Disease. *The Clin Neuropsychol* 15,125-128.
- Courtney, C., Farrel, D., Gray, R., Hills, R., Lynch, L., Sellwood, E. et al. (2004). Long-term donepezil treatment in 565 patients with Alzheimer's disease. *Lancet* 363, 2105-2115.
- Crum, R.M., Anthony, J.C., Bassett, S.S., Folstein, M.F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA* 269, 2386-2391.
- Cullen, B., O'Neill, B., Evans, J.J., Coen, R.F., Lawlor, B.A. (2007). A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatr* 78, 790–799.
- Dalrymple-Alford, J.C, MacAskill, M.R., Nakas, C.T., Livingston, L., Graham, C., Crucian, G.P. et al. (2010). The MoCA: well-suited screen for cognitive impairment in Parkinson disease. *Neurology* 75, 1717-1725.
- Dalziel, W.B. (2009). Dépistage ciblé de la démence. *Revue canadienne de la maladie d'Alzheimer et autres démences*, 12, 8-12.
- Darby, D., Maruff, P., Collie, A., McStephen, M. (2002). Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology* 59, 1042-1046.
- Dartigues, J.F., Gagnon, M., Michel, P., Letenneur, L., Commenges, D., Barberger-Gateau, P. et al.(1991). The Paquid research program on the epidemiology of dementia. Methods and initial results. *Rev Neurol* 147, 225-230.
- Daubs, J., Crick, R.P. (1980). Epidemiological analysis of the King's College Hospital glaucoma data. *Res Clin Forums* 2, 41-59.
- Davis, K.L., Thal, L.J., Gamzu, E.R., Davis, C.S., Woolson, R.F., Gracon, S.I., et al. (1992). A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease. The Tacrine Collaborative Study Group. *N Eng J Med*, 327, 1253-1259.
- de la Torre, J.C. Alzheimer's Disease is Incurable but Preventable (2010). *J Alzheimers Dis* 20, 861-870.

- de Rotrou, J., Wenisch, E., Chausson, C., Dray, F., Faucounau, V., Rigaud, A.-S. (2005). Accidental MCI in healthy subjects: a prospective longitudinal study. *Eur J Neurol* 12, 879-885.
- Delacour, H., Servonnet, A., Perrot, A., Vigezzi, J.F., Ramirez, J.M. (2005). La courbe ROC (receiver operating characteristic): principes et principales applications en biologie clinique. *Ann Biol Clin*, 63, 145-154.
- Delacour, H., Servonnet, A., Roche, C. (2009a). Critères d'évaluation de la validité d'un test biologique. *Revue francophone des laboratoires* 412, 41-48.
- Delacour, H., François, N., Servonnet, A., Gentile, A., Roche, B. (2009b). Les rapports de vraisemblance : un outil de choix pour l'interprétation des tests biologiques. *Immuno-analyse & Biologie Spécialisée* 24, 92-99.
- Derouesné, C., Poitreneau, J., Hugonot, L., Kalafat, M., Dubois B., Laurent B. Au nom du groupe de recherche sur l'évaluation cognitive (GRECO). (1999). Le Mental-State Examination (MMSE) : un outil pratique pour l'évaluation de l'état cognitif des patients par le clinicien. Version française consensuelle. *Presse Med* 28, 1141-1148.
- Devanand, D.P., Folz, M., Gorlyn, M., Moeller, J.R., Stern, Y. (1997). Questionable dementia: clinical course and predictors of outcome. *J Am Geriatr Soc* 45, 321-328.
- Devier, D.J., Pelton, G.H., Tabert, M.H., Liu, X., Eisenstadt, R. et al. (2009). The impact of anxiety on conversion from mild cognitive impairment to Alzheimer's disease. *Int J Geriatr Psychiatry* 24, 1335-1342.
- Doody, R.S., Stevens, J.C., Beck, C., Dubinsky, R.M., Kaye, J.A., Gwyther, L., et al. (2001). Practice parameter: management of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 56, 1154-1166.
- Drachman, D.A., Swearer, J.M. (1993). Driving and Alzheimer's disease: the risk of crashes. *Neurology* 43, 2448-2456.
- Dubois, B. (2000). "Prodromal Alzheimer's disease": a more useful concept than mild cognitive impairment? *Curr Opin Neurol* 13, 367-369.
- Dubois, B., Touchon, J., Portet, F., Ousset, P.J., Vellas, B., Michel, B. (2002). Les "5 mots" épreuve simple et sensible pour le diagnostic de la maladie d'Alzheimer. *Presse Med* 31, 1696-1699.
- Duff, K., Beglinger, L.J, Van Der Heiden, S., Moder, D.J, Arndt, S., Schultz, S.K. (2008). Short-term practice effects in amnesic mild cognitive impairment: implications for diagnosis and treatment. *Inter psychogeriatr Assoc* 20, 986-999.

- Duff, K., Lyketsos, C.G., Beglinger, L.J., Chelune, G., Moser, D.J., Arndt, S., et al. (2011). Practice effects predict cognitive outcome in amnesic mild cognitive impairment. *Am J Geriatr Psychiatry* 19, 932–939.
- Duran, J., Esnaola, S., Rubio, R., Izutueta, A. (2001). Obstructive sleep apnea-hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 years. *Am J Respir Crit Care* 163, 685-689.
- Duron, E., Hanon, O. (2008). Vascular risk factors, cognitive decline, and dementia. *Vasc Health Risk Manag* 4, 363-381.
- Ehreke, L., Luck, T., Lupp, M., König H.H., Villringer, A., Riedel-Heller, S.G. (2011). Clock Drawing Test – screening utility for mild cognitive impairment according to different scoring systems: results of the Leipzig Longitudinal Study of the Aged (LEILA 75+). *Int Psychogeriatr* 23, 1592-1609.
- Elias, M.F., Beiser, A., Wolf, P.A., Au, R., White, R.F., D’Agostino, R.B. (2000). The preclinical phase of alzheimer disease: A 22-year prospective study of the Framingham Cohort. *Arch Neurol* 57, 808-813.
- Ewing, J.A. (1984). Detecting alcoholism: the CAGE questionnaire. *JAMA* 252, 1905-1907.
- Elwood, R.W. (1993). Clinical discriminations and neuropsychological tests: An appeal to Bayes’ theorem. *Clin Neuropsychol* 7, 224–233.
- Fagan TJ. (1975). Letter: Nomogram for Bayes theorem. *N Engl J Med*, 293, 257.
- Ferreira, S., Vanholsbeeck, G., Chopard, G., Pitard, A., Tio, G., Vandell, P., et al. (2010). Normes comparatives de la batterie de tests neuropsychologiques RAPID pour les sujets âgés de 50 à 89 ans. *Rev Neurol* 166, 606-614.
- Fletcher, R.H. (1986). Carcinoembryonic antigen. *Ann Intern Med* 104, 66–73.
- Fletcher, R.H. (2000). Interpretation of diagnostic tests. *Indian J Pediatr* 67, 49-53.
- Fletcher, S.W., Black, W., Harris, R., Rimer, B.K., Shapiro, S. (1993). Report of the International Workshop on Screening for Breast Cancer. *J Natl Cancer Inst* 85, 1644-1656.
- Flicker, L., Loguidice, D., Carlin, J.B., Ames D. (1997). The predictive value of dementia screening instruments in clinical populations. *Int J Geriatr Psychiatry* 12, 203-209.
- Flint Beal, M., Lang, A.E., Ludolph, A.C. *Neurodegenerative Diseases: Neurobiology, Pathogenesis and therapeutics*. Cambridge University Press, 2005.
- Folstein, M.F., Folstein, S.E., McHugh, P.R. (1975). “Mini-mental state.” A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12, 189-198.

- Freedman, M.I., Leach, L., Kaplan, E., Winocur, G., Shulman, K.J., Delis, D.C. eds. (1994) Clock Drawing. Oxford: Oxford University.
- Freitas, S., Simões, M.R., Alves, L., Santana, I. (2012). Montreal Cognitive Assessment: influence of sociodemographic and health variables. *Arch Clin Neuropsychol* 27, 165–175.
- Fujii, D., Hishinuma, E., Masaki, K., Petrovich, H., Ross, G.W., White, L. (2003). Dementia screening: can a second administration reduce the number of false positives? *Am J Geriatr Psychiatry* 11, 462-465.
- Fulda, S., Schulz, H. (2001). Cognitive dysfunction in sleep disorders. *Sleep Medicine Reviews* 5, 423-445.
- Galasko D., Abramson I., Corey-Bloom J., Thal L.J. (1993). Repeated exposure to the Mini Mental State Examination and the Information-Memory-Concentration Test results in a practice effect in Alzheimer's Disease. *Neurology* 43, 1559-1562.
- Gallez, C. (2005). Rapport parlementaire sur la maladie d'Alzheimer et les maladies apparentées. (www.assembleenationale.fr).
- Galmiche, J., Chopard, G., Ferreira, S. et al. (2005). Evaluation cognitive des démences par la batterie Rapid: normes, valeurs seuil, seuils pondérés. *Rev Neurol* 159, 2S171.
- Galvin, J.E., Roe, C.M., Powlishta, K.K., Coats, M.A., Muich, S.J., Grant, E., et al. (2005). The AD8: a brief informant interview to detect dementia. *Neurology* 65, 559–564.
- Ganguli, M., Dodge, H.H., Shen, C., DeKosky, S.T. (2004). Mild cognitive impairment, amnesic type: an epidemiologic study. *Neurology* 63, 115-121.
- Ganguli, M., Snitz, B.E., Lee, C.W., Vanderbilt, J., Saxton, J.A., Chang, C.C. (2010). Age and education effects and norms on a cognitive test battery from a population-based cohort: The Monongahela–Youghiogheny Healthy Aging Team (MYHAT). *Aging Ment Health* 14, 100-107.
- Gauthier, S., Reisberg, B., Zaudig M., Petersen, R.C., Ritchie, K., Broich, K. et al. (2006). Mild Cognitive Impairment. *Lancet* 367, 1262-1270.
- Gifford, D.R., Cummings, J.L. (1999). Evaluating dementia screening. Methodologic standards to rate their performance. *Neurology* 52, 224-227.
- Giordani, B., Boivin, M.J., Hall, A.L., Foster, N.L., Lehtinen, S.J., Bluemlein, L.A., et al. (1990). The utility and generality of Mini-Mental State Examination scores in Alzheimer's disease. *Neurology* 40, 1894-1896.
- Girard, J.F, Canestri, A. (2000). La maladie d'Alzheimer. Rapport, ministère de l'emploi et de la solidarité. (<http://www.sante.gouv.fr/hm/actu/alzheimer/index.htm>).

- Goehring, C., Perrier, A., Morabia, A. (2004). Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat Med* 23, 125-135.
- Golafshani, N. (2003). Understanding Reliability and validity in Qualitative Research. *Qualitative Rep* 8, 597-607.
- Grimes, D.A., Schulz, K.F. (2002). Uses and abuses of screening tests. *Lancet* 359, 881-884.
- Grober, E., Buschke, H. (1987). Genuine memory deficits in dementia. *Developmental Neuropsychology* 3, 13-36.
- Grober, E., Lipton, R.B., Hall, C., Crystal, H. (2000). Memory impairment on free and cued selective reminding predicts dementia. *Neurology* 54, 827-832.
- Grober, E., Hall, C., Lipton, R.B., Teresi, J.A. (2008). Primary care screen for early dementia. *J Am Geriatr Soc* 56, 206-213.
- Guessous, A. (2010). Dépistage: principes et méthodes. *Dépistage* 6, 1390-1394.
- Guthrie, R., Susi, A. (1963). A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics* 32, 338-343.
- Hachinski, V. (2008). Shifts in Thinking About Dementia. *JAMA* 300, 2172-2173.
- Hall, M.C., Kieke, B., Gonzales, R., Belongia, E.A. (2004). Spectrum Bias of a Rapid Antigen Detection Test for Group A β -Hemolytic Streptococcal Pharyngitis in a Pediatric Population. *Pediatrics* 114, 182-186.
- Hanley, J.A., McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29-36.
- Harper, R.A., Reeves, B.C. (1999). Glaucoma screening: the importance of combining test data. *Optom Vis Sci* 76, 537-543.
- Harper, R., Henson, D., Reeves, B.C. (2000). Appraising evaluations of screening/diagnostic tests: the importance of the study populations. *Br J Ophtalmol* 84, 1198-1202.
- Haute Autorité de Santé. (2011). Maladies d'Alzheimer et maladies apparentées: diagnostic et prise en charge. Recommandations, décembre 2011. <http://www.has.sante.fr/portail/joms/c1148883/maladie-dalzheimer-et-maladies-apparentes-diagnostic-et-prise-en-charge>.
- Hlatky, M.A., Pryor, D.B., Harrell, F.E., Jr, Califf, R.M., Mark, D.B., Rosati, R.A. (1984). Factors affecting sensitivity and specificity of exercise electrocardiography. Multivariable analysis. *Am J Med* 77, 64-71.
- Hodkinson, H.M. (1972). Evaluation of mental impairment in the elderly. *Age Ageing* 1, 233-238.

- Hohenberg, M. (2004). The Epidemiological Approach. *J R Soc of Med* 97, 453.
- Holsinger, T., Deveau, J., Boustani, M., Williams, J.W., Jr. (2007). Does this patient have dementia? *JAMA* 297, 2391-2404.
- Horvath, A.R., Pewsner, D. (2004). Systematic reviews in laboratory medicine: principles, processes and practical considerations. *Clin Chim Acta*, 342, 23-39.
- Hughes, C.P., Berg, L., Danziger, W.L., Coben, L.A., Martin, R.L. (1982). A new clinical scale for the staging of dementia. *Br J Psychiatry* 140, 566-572.
- Huppert, F.A. (2002). Guidelines for assessing the predictive value of neuropsychological measures. *Cortex* 38, 414-416.
- Irwig, L., Bossuyt, P., Glasziou, P., Gatsonis, C., Lijmer, J. (2002). Designing studies to ensure that estimates of test accuracy are transferable. *BMJ* 324, 669-671.
- Isaacs, B., Kennie, A.T. (1973). The Set test as an aid to the detection of dementia in old people. *Br J Psychiatry* 123, 467-470.
- Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D., Letenneur, L., Dartigues, J.F. (2000). A cognitive screening battery for dementia in the elderly. *J Clin Epidemiol* 53, 980-987.
- Jaeschke, R., Guyatt, G.H., Sackett, D.L. (1994). III. How to use an article about a diagnostic test; B. What are the results and will they help me in caring for my patients? *User's Guides to the Medical Literature* 271, 703-707.
- Junod, A.F. (1995). « Spectrum bias » ou « biais de diversité » ou comment interpréter l'information donnée par un test en fonction de l'extension ou de la gravité de la maladie ? *Méd et hyg* 53, 2270-2273.
- Karakaya, J., Aksoy, D.Y., Harmanci, A., Karaagaoglu, E., Gurlek, A. (2007). Predictive ability of fasting plasma glucose for a diabetic 2-h postload glucose value in oral glucose tolerance test: spectrum effect. *J Diabetes Complications* 21, 300-305.
- Kiencke, P., Daniel, D., Grimm, C., Rychlik, R. (2011). Direct costs of Alzheimer's disease in Germany. *Eur J Health Econ* 12, 533-539.
- Kilada, S., Gamaldo, A., Grant, E.A., Moghekar, A., Morris, J.C., O'Brien, R.J. (2005). Brief screening tests for the diagnosis of dementia: comparison with the mini-mental state exam. *Alzheimer Dis Assoc Disord* 19, 8-16.
- Kivipelto, M., Ngandu, T., Laatikainen, T., Windblad, B., Soininen, H., Tuomilehto, J. (2006) Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol* 5, 735-741.

- Knopman, D.S., Boeve, B.F., Petersen, R.D. (2003). Essentials of the Proper Diagnoses of Mild Cognitive Impairment, Dementia, and Major Subtypes of Dementia. *Mayo Clin Proc* 78, 1290-1308.
- Kuslansky, G., Buschke, H., Katz, M., Sliwinski, M., Lipton, R.B. (2002). Screening for Alzheimer's disease: the memory impairment screen versus the conventional three-word memory test. *J Am Geriatr Soc* 50, 1086-1091.
- La Rosa, E. (1996). *Le dépistage chez l'adulte*. Presses Universitaires de France.
- Lachs, M.S., Nachmkin, I., Edelstein, P.H., Goldman, J., Feinstein, A.R., Schwaetz, J.S. (1992). Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. *Ann Intern Med* 117, 135-140.
- Lal, C. Strange, C., Bachman, D. (2012). Neurocognitive Impairment in Obstructive Sleep Apnea. *Chest* 141, 1601-1610.
- Lawton, M.P., Brody, E.M. (1969). Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 9, 179-186.
- Lebert, F., Pasquier, F., Souliez, L., Petit, H. (1998). Frontotemporal behavioral scale. *Alzheimer Dis Assoc Disord* 12, 335-339.
- Lee, H., Swanwick, G.R., Coen, R.F., Lawlor, B.A. (1996). Use of the Clock Drawing Task in the Diagnosis of Mild and very Mild Alzheimer's Disease. *Int Psychogeriatr* 8, 469-476.
- Leeflang, M.M.G., Moons, K.G.M., Reitsma, J.B., Zwinderman, A.H. (2008a). Bias in Sensitivity and Specificity Caused by Data-Driven Selection of Optimal Cutoff Values: Mechanisms, Magnitude, and Solutions. *Clinical Chemistry* 54, 729-737.
- Leeflang, M.M.G., Deeks, J.J., Gatsonis, C., Bossuyt, P.M.M. (2008b). Systematic Reviews of Diagnostic Test Accuracy. *Ann Intern Med* 149, 889-897.
- Li, J., Wang, Y.J, Zhang, M., Yu, Z.K., Gao, C.Y., Fang., C.Q. et al. (2011). Vascular risk factors promote conversion from mild cognitive impairment to Alzheimer Disease. *Neurology* 76, 1485-1491.
- Lin, K.N., Wang, P.N., Chiu, Y.H., Kuo, C.C., Chuang, Y.Y. et al. (2003). The three-item clock-drawing test: a simplified screening test for Alzheimer disease. *Eur Neurol* 49, 53-58.
- Lijmer, J.G., Mol, B.W., Heisterkamp, S., Bossel, G.J., Prins, M.H., van der Meulen, J.H., et al. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 282, 1061-1066.

- Linn, R.T., Wolf, P.A., Bachman, D.L., Knoefel, J.E., Cobb, J.L., Belanger, A.J., et al. (1995). The “preclinical phase” of probable Alzheimer’s disease. A 13-year prospective study of the Framingham cohort. *Arch Neurol* 52, 485–490.
- Lobo, J.A., Kalu, U.G., Ebmeier, K.P. (2000). Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. *Neurologic Diseases in the Elderly Research Group. Neurology* 54(suppl 11):54-59.
- Lonie, J.A., Tierney, K.M., Ebmeier, K.P. (2009). Screening for mild cognitive impairment: a systematic review. *Int J Geriatr* 24, 902-915.
- Lonie, J.A., Kalu, U.G., Ebmeier, K.P. (2010). A Systematic Review of Cognitive Screening for Mild Cognitive Impairment. *Mind&Brain* 24, 902-915.
- Lorentz, W.J., Scanlan, J.M., Borson, S. (2002). Brief screening tests for dementia. *Can J Psychiatry* 47, 723-733.
- Lyketsos, C.G., Lopez, O., Jones, B., Fitzpatrick, A.L., Breitner, J., DeKosky, S. (2002). Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: results from the cardiovascular health study. *JAMA* 288, 1475-1483.
- Mackinnon, A., Mulligan, R. (1998). Combining cognitive testing and informant report to increase accuracy in screening for dementia. *Am J Psychiatry* 155, 1529-1535.
- Marquié, J.C., Duarte, L.R., Bessières, P., Dalm, C., Gentil, C., Ruidavets, J.B. (2010). Higher mental stimulation at work is associated with improved cognitive functioning in both young and older workers. *Ergonomics* 53, 1287-1301.
- McCaffrey, RJ, Duff, K., Westervelt HJ. (2000). *Practitioner’s Guide to Evaluating Change with Neuropsychological Assessment Instruments*. New York: plenum/Kluwer.
- McCullagh, C.D., Craig, D., Mclroy, S.D., Passmore, P. (2001). Risk Factors for Dementia. *APT* 7, 24-31.
- Masur, D.M., Sliwinski, M., Lipton, R.B., Blau, A.D., Crystal, H.A. (1994). Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons. *Neurology* 44, 1427-1432.
- McKeith, I.G., Galasko, D., Kosaka, K., Perry, E.K., Dickson, D.W., Hansen, L.A., et al. (1996). Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. *Neurology* 47, 1113-1124.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M. (1984). 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* 34, 939–939.

- McKhann, G., Kopman D.S., Chertkow H., Hyman B.T., Jack C.R., Kawas C.H et al. (2011). 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. *Alzheimers dement* 7, 263-269.
- Mendez, M.F. Ala, T., Underwood, K.L. (1992). Development of scoring criteria for the clock drawing task in Alzheimer's disease. *J Am Geriatr Soc* 40, 1095-1099.
- Merril, F., Beiser, A., Wolf, P.A., Au, R., White, R.F., D'Agostino, R.B. (2000). The Preclinical Phase of Alzheimer Disease. A 22-year Prospective Study of Alzheimer Disease. *Arch Neurol* 57, 808-813.
- Meulen, E.F.J., Schmand, B., Campen, J.P. van, Koning, S.J. de, Ponds, R.W., Scheltens, P., et al. (2004). The seven minute screen: a neurocognitive screening test highly sensitive to various types of dementia. *J Neurol Neurosurg Psychiatry* 75, 700-705.
- Michel, J.M., Sellal, F. (2011). Les démences curables en 2011. *Ger Psychol Neuropsychiatr Vieil* 9, 211-25
- Milne, A., Culverwell, A., Guss, R., Tuppen, J., Whelton, R. (2008). Screening for dementia in primary care: a review of the use, efficacy and quality of measures. *Int Psychogeriatr* 20, 911-926.
- Mimori, K., Muneta, T., Nakagawa, T., Shinomiya, K. (1999). A New Pain Provocation Test for Superior Labral Tears of the Shoulder. *Am J Sports Med* 27, 137-142.
- Ministère de la Santé, de la Jeunesse, des Sports et de la Vie associative. Plan « Alzheimer et maladies apparentées » 2008-2012. Paris: Ministère de la Santé, de la Jeunesse, des Sports et de la Vie associative. Novembre 2007.
- Mittelman, M.S., Ferris, S.H., Shulman, E., Steinberg, G., Levin, B. (1996). A family intervention to delay nursing home placement of patient with Alzheimer disease: a randomized controlled trial. *JAMA* 276, 1725-1731.
- Mitchell, A.J., Psych, M.R.C., Malladi, S. (2010a). Screening and Case-Finding Tools for the Detection of Dementia. Part II: Evidence-Based Meta-Analysis of multidomain Tests. *Am J Geriatr Psychiatry* 18, 759-800.
- Mitchell, A.J., Psych, M.R.C., Malladi, S. (2010b). Screening and Case-Finding Tools for the Detection of Dementia. Part II: Evidence-Based Meta-Analysis of Single-Domain Tests. *Am J Geriatr Psychiatry* 18, 783-800.
- Modrego, P.J., Ferrández, J. (2004). Depression in patients with mild cognitive impairment increases the risk of developing dementia of Alzheimer type: a prospective cohort study. *Arch Neurol* 61, 1290-1293.

- Moore, A.R., O’Keeffe, S.T. (1999). Drug-induced cognitive impairment in the elderly. *Drugs Aging* 15, 15-28.
- Morris, J.C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 43, 2412-2414.
- Morrison, A.S. Introduction. In: Morisson A.S. (1992). *Screening in chronic disease*. 2nd edition. New York and Oxford: Oxford University Press, 3-20.
- Muir, S.W., Gopaul, K., Montero Odasso, M.M. (2012). The role of cognitive impairment in fall risk among older adults: a systematic review and meta-analysis. *Age Ageing* 41, 299-308.
- Mulherin, S.A., Miller, W.C. (2002). Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 137, 598-602.
- Murden R.A., McRae T.D., Kanen S., Bucknam M.E. (1991). Mini-Mental Status Exam scores in Blacks and Whites. *J Am Geriatr Soc* 43, 138-145.
- Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 53, 695-699.
- Nierenberg, A.A., Feinstein, A.R. (1988). How to Evaluate a Diagnostic Marker Test: Lessons From the Rise and Fall of Dexamethasone Suppression Test. *JAMA* 259, 1699-1702.
- Ohayon, M.M. (2007). Prévalence et comorbidité des troubles du sommeil dans la population générale. *Rev Prat* 57, 1521-1528.
- Ohayon, M.M., Vecchierini, M.F. (2002). Daytime Sleepiness and Cognitive Impairment in the Elderly population. *Arch Intern Med* 162, 201-208.
- O’Bryant, S.E., Humphreys, J.D., Smith, G.E., Ivnick R.J., Graff-Radford, N.R., Petersen R.C. et al. (2008). Detecting Dementia with the Mini-Mental State Examination (MMSE) in Highly Educated Individuals. *Arch Neurol* 65, 963-967.
- O’Connor, D.W., Pollitt, P., Hyde, J., Felows, J.L., Miller, N.D. Brokk, C.P.B,et al. (1989) The reliability and the validity of the Mini-Mental State in a British community survey. *Journal of Psychiatry research* 23, 87-96.
- Palmer, K., Bäckman, L., Winblad, B., Fratiglioni, L. (2003). Detection of Alzheimer’s disease and dementia in the preclinical phase: population based cohort study. *BMJ* 326, 2-5.
- Panza, F., Capurso, C., D’Introno, A., Colacicco, A.M., Frisardi, V., Santamato, A. et al. (2008). Vascular risk factors, alcohol intake, and cognitive decline. *J Nutr Health Aging* 12, 376-381.

- Parentis, M.A., Glousman, R.E., Mohr, K.S., Yocum, L.A. (2006). An evaluation of the provocative tests for superior labral anterior posterior lesions. *Am J Sports Med* 34, 265-268.
- Parker, C., Philp, I. (2004). Screening for cognitive impairment among older people in black and minority ethnic groups. *Age Ageing* 33, 447-452.
- Pasquier, F., Delacourte, A. (1998). Non-Alzheimer degenerative dementias. *Curr Opin Neurol* 11, 417-427.
- Pasquier, F., Lebert, F., Petit, H. (1999). Organisation des centres de la mémoire et perspectives. *Rev Neurol* 155, 4583-4592.
- Patterson, C.J.S., Gauthier, S., Bergman, H., Cohen, C.A., Feightner, J.W., Fieldman, H., et al. (1999). Diagnostic, évaluation et traitement de la démence: conclusions de la conférence canadienne de consensus sur la démence. *CMAJ* 160, SF1-SF18.
- Perry, R.J., Hodges, J.R. (1996). Spectrum of memory dysfunction in degenerative disease. *Curr Opin Neurol* 9, 281-285.
- Perneger, T., Perrier, A. (2004). Analyse d'un test diagnostique: courbe ROC, ou « receiver operating characteristic ». *Rev Mal Respir* 21, 398-401.
- Petersen, R.C., Smith, G.E, Ivnik, R.J, Kokmen, E., Tangalos, E.G. (1994). Memory function in very early Alzheimer's disease. *Neurology* 44, 867-72.
- Petersen, R.C., Smith, G.E., Waring S.C., Ivnik R.J., Tangalos, E.G., Kokmen E. (1999). Mild Cognitive Impairment: clinical characterization outcome. *Arch Neurol* 56, 303-308.
- Petersen, R.C., Stevens, J.C., Ganguli, M., Tangalos, E.G., Cummings, J.L., DeKosky, S.T. (2001). Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 56, 1133-1142.
- Petersen, R.C. (2007). MCI treatment trials: failure or not? *Lancet Neurol* 6, 473-475.
- Pohjasvaara, T., Mäntylä, R., Ylikoski, R., Kaste, M., Erkinjuntti, T. (2000). Comparison of different clinical criteria (DSM-III, ADDTC, ICD-10, NINDS-AIREN, DSM-IV) for the diagnosis of vascular dementia. National Institute of Neurological Disorders and Stroke-Association Internationale pour la Recherche et l'Enseignement en Neurosciences. *Stroke* 31, 2952-2957.
- Price, J.L, Morris, J.C. (1994). Tangles and Plaques in Nondemented Aging and "Preclinical" Alzheimer's Disease. *Ann Neurol* 45, 358-368.
- Rafii, M.S., Galasko, D. (2008). Primary Care Screening for Dementia and Mild Cognitive Impairment. *JAMA* 299, 1132.

- Ransohoff, D.F., Feinstein, A.R. (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 299, 926-930.
- Reid, M.C., Lachs, M.S., Feinstein, A.R. (1995). Use of methodological standards in diagnostic test research: Getting better but still not. *JAMA* 274, 645-651.
- Reitan, R.M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills* 8, 271-276.
- Reitan RM, Wolfson D. (1985). *The Halstead-Reitan Neuropsychological Test Battery: Therapy and clinical interpretation.* Tucson, AZ: Neuropsychological Press.
- Riedel-Heller, S.G., Schork, A., Matschinger, H., Angermeyer, M.C. (2000). Recruitment procedures and their impact on the prevalence of dementia: results from the Leipzig Longitudinal Study of the Aged (LEILA75+). *Neuroepidemiology* 19, 130-140.
- Royal, D.R., Cordes, J.A., Polk, M. (1998). CLOX: an executive clock drawing task. *J Neurol Neurosurg Psychiatry* 64, 588-594.
- Rutjes, A.W., Reitsma, J.B., Vandenbroucke, J.P., Glas, A.S., Bossuyt, P.M. (2005). Case-Control and Two-Gate Designs in Diagnostic Accuracy Studies. *Clin Chem* 51, 1335-1341.
- Sarazin, M., Berr, C., de Rotrou, J., Fabrigoule, C., Pasquier, F., Legrain, S., et al. (2007). Amnesic syndrome of the medial temporal type identifies prodromal AD: a longitudinal study. *Neurology* 69, 1859-1867.
- Saxton, J. Lopez, O.L., Ratcliff, G., Dulberg, C. Fried, L.P., Carlson, M.C. et al. (2004). Preclinical Alzheimer disease. Neuropsychological test performance 1.5 to 8 years prior to onset. *Neurology* 63, 2341-2347.
- Sasieni, P., Castanon, A., Cuzick, J. (2009). Effectiveness of cervical screening with age: population based case-control study of prospectively recorded data. *BMJ* 339:b3115.
- Signoret, J.L, Allard, M., Benoit, N., Bolgert, F., Bonvarlet, M., Eustache, F. Evaluation des troubles de mémoire et des désordres cognitifs associés. BEC 96. Paris: Ipsen, 1998.
- Steinbauer, J.R., Cantor, S.B., Holzer, C.E., Volk, R.J. (1998). Ethnic and Sex Bias in Primary Care Screening Tests for Alcohol Use Disorders. *Ann Intern Med* 129, 353-362.
- Stepanuik, J., Ritchie, L., Tuokko, H. (2008). Neuropsychiatric impairments as risk factors for Mild Cognitive Impairment and dementia. *Am J of Alzheimers Dis other Demen* 23, 326-333.

- Stephan, B.C., Brayne, C., Savva, G.M., Matthews, F.E. and the Medical Research Council Cognitive Function and Ageing Study (2011). Occurrence of medical co-morbidity in mild cognitive impairment: implications for generalization of MCI research. *Age Ageing* 40, 501-507.
- Steurer, J. (2002). Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 324, 824-826.
- Strange, L.C., Bachman, D. (2012). Neurocognitive impairment in obstructive sleep apnea. *Chest* 141, 1601-1610.
- Tabert, M.H., Manly, J.J., Liu, X., Pelton, G.H., Rosenblum, S., Jacobs, M., et al. (2006). Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Arch Gen Psychiatry* 63, 916-924.
- The Lund and Manchester Group. (1994). Clinical and neuropathological criteria for frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 57, 416-418.
- Thomas Anterion, C., Honore, S., Cougny, H., Grosmaître, C., Laurent, B. (2001). Apport de l'épreuve d'évocation lexicale du Set Test dans le dépistage de la maladie d'Alzheimer. *Rev Neurol* 157, 1377-1382.
- Tian, J., Bucks, R., Haworth, J., Wilcock, G. (2003). Neuropsychological prediction of conversion to dementia from questionable dementia: statistically significant but not yet clinically useful. *J Neurol Neurosurg Psychiatry* 74, 433-438.
- Tielsch, J.M., Sommer, A., Katz, J., Royall, R.M., Quigley, H.A., Javitt, J. (1991). Racial Variations in the Prevalence of Primary Open-angle Glaucoma: The Baltimore Eye Survey. *JAMA* 266, 369-374.
- Tounsi, H., Deweer, B., Ergis, A.M., Van Der Linden, M., Pillon, B., Michon, A., et al., (1999). Sensitivity to semantic cueing: an index of episodic memory dysfunction in early Alzheimer disease. *Alzheimer Dis Assoc Disord* 13, 38-46.
- Van der Linden, M., Coyette, F., Poitrenaud, J., Kalafat, M., Calacis, F., Wyns, C. et al. L'épreuve de rappel libre/rappel indicé à 16 items (RL/RI-16). In: Van der Linden, GREMEM, eds. L'évaluation des troubles de la mémoire. Marseille: Solal, 2004:25-47.
- van der Schouw, Y.T., Van Dijk, R., Verbeek, A.L. (1995a). Problems in selecting the adequate patient population from existing data files for assessment studies of new diagnostic tests. *J Clin Epidemiol* 48, 417-422.
- van der Schouw, Y.T., Verbeek, A.L., Ruijs S.H.J. (1995b). Guidelines for the assessment of new diagnostic tests. *Invest Radiology* 30, 334-340.

- Vanholsbeeck, G., Ferreira, S., Pitard, A., Bouiller, C., Chopard, G., Rumbach, L., et al. (2007). Intérêt d'un dossier national de réseau de santé dans le cadre de la veille sanitaire des maladies d'Alzheimer et syndromes apparentés: le logiciel "rapid". *Filière et Réseaux de Santé* 13, 36-39.
- Vercambre, M.N., Grodstein, F., Berr, C. Kang, J.H. (2012). Mediterranean diet and cognitive decline in women with cardiovascular disease or risk factors (2012). *J Acad Nutr Diet* 112, 816-823.
- Verghese, J., Lipton, R.B., Katz, M.J., Hall, C.B, Derby, C.A., Kuslansky, G. et al. (2003). Leisure activities and the risk of dementia in the elderly. *N Eng J Med* 348, 2508-2516.
- Verghese, J. LeValley, A., Derby, C., Kulansky, G., Katz, M., Hall, C., et al. (2006). Leisure activities and the risk of amnesic mild cognitive impairment in the elderly. *Neurology* 66, 821-827.
- Videnovic, A., Bernard, B., Fan, W., Jaglin, R.N., Leurgans, S., Shannon, K.M. (2010). The Montreal Cognitive Assessment as a screening tool for cognitive dysfunction in Huntington's disease. *Mov Disord* 25, 401-404.
- Visser, P.J. Predictors of Alzheimer type dementia in subjects with mild cognitive impairments. Maastricht: Neuropsych Publishers Maastricht, 2000.
- Visser, P.J. (2006a). Mild Cognitive Impairment. *In: Principles and Practice of Geriatric Medicine*. 4th ed.M.S. John Pathy, Alan J. Sinclair and John E. Morley.
- Visser, P.J., Kester, A., Jolles, J., Verhey, F. (2006b). Ten-year risk of dementia in subjects with mild cognitive impairment. *Neurology* 67, 1201-1207.
- Wald NJ. The epidemiological approach. London, Royal Society of Medicine, 2004.
- Watine, J. (2004). Revues systématiques et méta-analyses en biologie clinique : principes et méthodes. *Annales de biologie clinique* 62, 611-627.
- Wechsler, D. (1945). A standardized memory scale for clinical use. *J Psychol* 9, 87-95.
- Weytingh, M.D., Bossuyt, P.M., van Crevel, H. (1995). Reversible dementia: more than 10% or less than 1%? A quantitative review. *J Neurol* 242, 466-471.
- White, N., Scott, A., Woods, R.T., Wenger, G.C., Keady, J.D., Devakumar, M. (2002). The limited utility of the Mini-Mental State Examination in screening people over the age of 75 years for dementia in primary care. *Br J Gen Pract* 52, 1002-1003.

- Whiting, P., Rutjes, A.W.S., Reitsma, J.B., Glas, A.S., Bossuyt, P.M.M., Kleijnen, J. (2004). Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 140, 189-202.
- Willis, B.H. (2008). Spectrum bias--why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract* 25, 390-396.
- Wilson, J.M.G., Jungner, G. (1970). Principes et pratique du dépistage des maladies. Cahiers de santé publique N°34. Genève. Organisation Mondiale de la Santé.
- Zhou, X.H., Obuchowski, N.A., McClish, D.K. *Statistical Methods in Diagnostic Medicine*. New-York: John Wiley and sons, 2002.

ANNEXES

**ANNEXE 1. CONSIGNES DE PASSATION ET DE COTATION
DU MEMORY IMPAIRMENT SCREEN (MIS).**

MIS
MEMORY IMPAIRMENT SCREEN

CONSIGNES

→ Il s'agit d'un test de mémorisation de 4 mots appartenant à 4 catégories sémantiques différentes (outils de jardin, animal, couleur, ustensile de cuisine). Ces mots sont écrits sur une feuille et présentés au patient.

PROCEDURE

1. Dire au patient :

« Je vais vous lire 4 mots. Il s'agit de mots que vous devrez mémoriser. Retenez-les bien car je vous les redemanderai tout à l'heure ».

2. Présenter la feuille au patient.

3. **Lecture simple.**

Dire au patient :

« Lisez à haute voix les mots que je vous montre ».

Le patient va lire les 4 mots dans un ordre **indiqué par l'examineur** qui les pointe (passoire, violet, chameau, pioche). L'ordre choisi par l'examineur doit être différent de l'ordre habituel de lecture.

4. **Lecture indicée.**

Faire lire les 4 mots avec l'indice catégoriel dans un **ordre différent** de celui du point 3.

Exemple : *« Lisez à haute voix et montrez du doigt le nom de l'animal »*, etc.

Le patient **pointe** du doigt les mots cibles en lecture indicée mais l'examineur peut pointer les mots que le patient n'arrive pas à identifier.

5. Retourner la feuille et dire :

« Je vous redemanderai ces mots dans 2 minutes ».

6. Tâche interférente non sémantique pendant 2 minutes minimum (3 minutes maximum) : faire une tâche de comptage à rebours ou test de l'horloge.

7. **Rappel Libre (RL).**

L'examineur montre du doigt la feuille du test (les mots étant cachés) et demande au patient : *« Pouvez-vous me rappeler dans n'importe quel ordre les mots qui étaient écrits sur cette feuille ? »*

Ne pas corriger la réponse du patient lorsqu'il ne sait pas ou répond par une intrusion.

8. **Rappel Indiqué (RI).**

Faire un rappel indicé pour chaque mot non rappelé lors du rappel libre.

Exemple : *« Quel était le nom de l'animal qui était écrit sur cette feuille ? »*

Ne pas corriger la réponse du patient lorsqu'il ne sait pas ou répond par une intrusion.

COTATION ET CALCUL DU SCORE

↪ 2 points pour chaque mot correctement rappelé en RL (Rappel Libre)

↪ 1 point pour chaque mot correctement rappelé en RI (Rappel Indiqué)

↪ Score total sur 8

Exemple 1 :

Un patient répond « *chameau* » et « *passoire* » en RL

2 mots correctement rappelés en RL ($2 \times 2 = 4$ points)

Ce patient retrouve 1 seul mot en RI « *pioche* » (1 point)

Son score est : $(2 \times 2) + 1 = 5$

Exemple 2 :

Un patient ne retrouve aucun mot en RL « *je ne me rappelle plus des mots* »

Ce patient retrouve les 4 mots à l'aide des indices catégoriels

Son score est : $(2 \times 0) + 4 = 4$

Exemple 3 :

Un patient retrouve 3 mots en RL « *passoire, chameau, violet* »

L'examineur demande : « *Quel était le nom de l'outil ?* »

Le patient répond « *je ne sais plus* »

Son score est : $(2 \times 3) + 0 = 6$

passoire

violet

chameau

pioche

**ANNEXE 2. CONSIGNES DE PASSATION ET DE COTATION
DU SET TEST D'ISAACS (STI).**

STI
SET TEST D'ISAACS

CONSIGNES

Dire au patient « *Donnez-moi le plus possible de couleurs pendant 15 secondes. Allez-y !* »

Idem pour les autres catégories.

Début au chronomètre juste après la consigne et non quand le patient donne le 1^{er} mot.

Nombre de noms et de couleurs

Les dégradés de couleurs (exemple : vert clair, vert foncé) ne comptent que pour une couleur = 1 point.

Exemple : « rouge, bleu, vert, vert clair » : 3 points.

Nombre de noms d'animaux

Les femelles et les petits comptent pour un seul animal (exemple : cheval, jument, poulain = 1 point).

Le générique vaut 1 point s'il n'est pas détaillé par la suite. S'il est détaillé (que ce soit tout de suite ou après) il ne compte plus (exemple : poisson, truite = 1 point ; oiseau, cheval, canard = 2 points).

Nombre de noms de fruits

Tomate et avocat sont des fruits.

Nombre de noms de villes et villages

On accepte aussi les villages ainsi que les villes étrangères.

TOTAL

1 point pour chaque item correctement donnée dans les 4 catégories.

ANNEXE 3. FEUILLE DE COTATION DU MINI-MENTAL STATE EXAMINATION (MMSE).

Mini Mental State Examination (MMSE) (Version consensuelle du GRECO)

Orientation / 10

Je vais vous poser quelques questions pour apprécier comment fonctionne votre mémoire. Les unes sont très simples, les autres un peu moins. Vous devez répondre du mieux que vous pouvez. Quelle est la date complète d'aujourd'hui ? _____

Si la réponse est incorrecte ou incomplète, posez les questions restées sans réponse, dans l'ordre suivant :

- | | |
|----------------------------------|--------------------------|
| 1. En quelle année sommes-nous ? | <input type="checkbox"/> |
| 2. En quelle saison ? | <input type="checkbox"/> |
| 3. En quel mois ? | <input type="checkbox"/> |
| 4. Quel jour du mois ? | <input type="checkbox"/> |
| 5. Quel jour de la semaine ? | <input type="checkbox"/> |

Je vais vous poser maintenant quelques questions sur l'endroit où nous trouvons.

- | | |
|--|--------------------------|
| 6. Quel est le nom de l'hôpital où nous sommes ?* | <input type="checkbox"/> |
| 7. Dans quelle ville se trouve-t-il ? | <input type="checkbox"/> |
| 8. Quel est le nom du département dans lequel est située cette ville ?** | <input type="checkbox"/> |
| 9. Dans quelle province ou région est située ce département ? | <input type="checkbox"/> |
| 10. A quel étage sommes-nous ? | <input type="checkbox"/> |

Apprentissage / 3

Je vais vous dire trois mots ; je vous voudrais que vous me les répétiez et que vous essayiez de les retenir car je vous les redemanderai tout à l'heure.

- | | | | |
|------------|--------|----------|--------|
| 11. Cigare | Citron | Fauteuil | |
| 12. Fleur | ou | Clé | ou |
| 13. Porte | | Ballon | ou |
| | | | Canard |

Répéter les 3 mots.

Attention et calcul / 5

Voulez-vous compter à partir de 100 en retirant 7 à chaque fois ?*

- | | | |
|-----|----|--------------------------|
| 14. | 93 | <input type="checkbox"/> |
| 15. | 86 | <input type="checkbox"/> |
| 16. | 79 | <input type="checkbox"/> |
| 17. | 72 | <input type="checkbox"/> |
| 18. | 65 | <input type="checkbox"/> |

Pour tous les sujets, même pour ceux qui ont obtenu le maximum de points, demander :

Voulez-vous épeler le mot MONDE à l'envers ?**

Rappel / 3

Pouvez-vous me dire quels étaient les 3 mots que je vous ai demandés de répéter et de retenir tout à l'heure ?

- | | | | |
|------------|--------|----------|--------|
| 11. Cigare | Citron | Fauteuil | |
| 12. Fleur | ou | Clé | ou |
| 13. Porte | | Ballon | ou |
| | | | Canard |

Langage / 8

Montrer un crayon. 22. Quel est le nom de cet objet ?*

Montrer votre montre. 23. Quel est le nom de cet objet ?**

24. Ecoutez bien et répétez après moi : « PAS DE MAIS, DE SI, NI DE ET »***

Poser une feuille de papier sur le bureau, la montrer au sujet en lui disant : « Ecoutez bien et faites ce que je vais vous dire :

- | | |
|--|--------------------------|
| 25. Prenez cette feuille de papier avec votre main droite, | <input type="checkbox"/> |
| 26. Pliez-la en deux, | <input type="checkbox"/> |
| 27. Et jetez-la par terre. »**** | <input type="checkbox"/> |

Tendre au sujet une feuille de papier sur laquelle est écrit en gros caractère : « FERMEZ LES YEUX » et dire au sujet :

28. « Faites ce qui est écrit ».

Tendre au sujet une feuille de papier et un stylo, en disant :

29. « Voulez-vous m'écrire une phrase, ce que vous voulez, mais une phrase entière. »*****

Praxies constructives / 1

Tendre au sujet une feuille de papier et lui demander : 30. « Voulez-vous recopier ce dessin ? »

ANNEXE 4. CARACTERISTIQUES DEMOGRAPHIQUES ET NEUROPSYCHOLOGIQUES DE L'ECHANTILLON D'ETUDE (ARTICLE 1).

Demographic and neuropsychological characteristics of participants.

Characteristic	Original sample		Validation sample	
	Controls (n = 206)	Patients with dementia (n = 282)	Controls (n = 105)	Patients with dementia (n = 139)
Age, mean ± SD	75.3 ± 8.3	76.2 ± 6.2	75.2 ± 8.3	76.8 ± 6.1
Male, n (%)	85 (41.3)	121 (42.9)	43 (40.9)	59 (42.4)
Education ≥ 12 years, n (%)	36 (17.5)	39 (13.8)	19 (18.1)	18 (12.9)
Memory Impairment Screen score, mean ± SD (range 0-8)	7.3 ± 0.9	3.6 ± 2.4	7.5 ± 0.8	3.8 ± 2.3
Isaacs Set Test score, mean ± SD	33.8 ± 6.7	21.2 ± 6.9	34.2 ± 6.1	20.5 ± 6.8
Mini-Mental State Examination score, mean ± SD (range 0-30)	27.2 ± 2.4	19.3 ± 4.5	27.6 ± 1.9	19.5 ± 4.5
Buschke Cued Recall Test score, mean ± SD (range 0-48)				
Free Recall	24.1 ± 6.4	8.7 ± 5.5	23.3 ± 6.8	8.8 ± 4.8
Total Recall	44.2 ± 4.3	26.3 ± 10.9	43.8 ± 5.3	27.1 ± 10.7
Trail-Making Test score, mean ± SD				
Part A (range 0-150)	53.2 ± 25.6	111.9 ± 78.7	55.8 ± 27.0	107.9 ± 74.1
Part B (range 0-300)	157.2 ± 89	263.5 ± 111.6	170.1 ± 113.6	293.9 ± 182.1
Picture naming score, mean ± SD (range 0-30)	29.1 ± 1.6	25.2 ± 5.0	28.9 ± 1.6	23.8 ± 4.4
Copy score, mean ± SD (range 0-6)	5.4 ± 1.3	3.9 ± 2.3	5.5 ± 1.1	4.1 ± 2.1

Abbreviations: SD = Standard Deviation.

ANNEXE 5. SENSIBILITE, SPECIFICITE ET VALEUR DE PREDICTION POSITIVE DU MIS, DU STI, DU MMSE ET DU COUPLE MIS-STI SUR L'ECHANTILLON DE VALIDATION (ARTICLE 1).

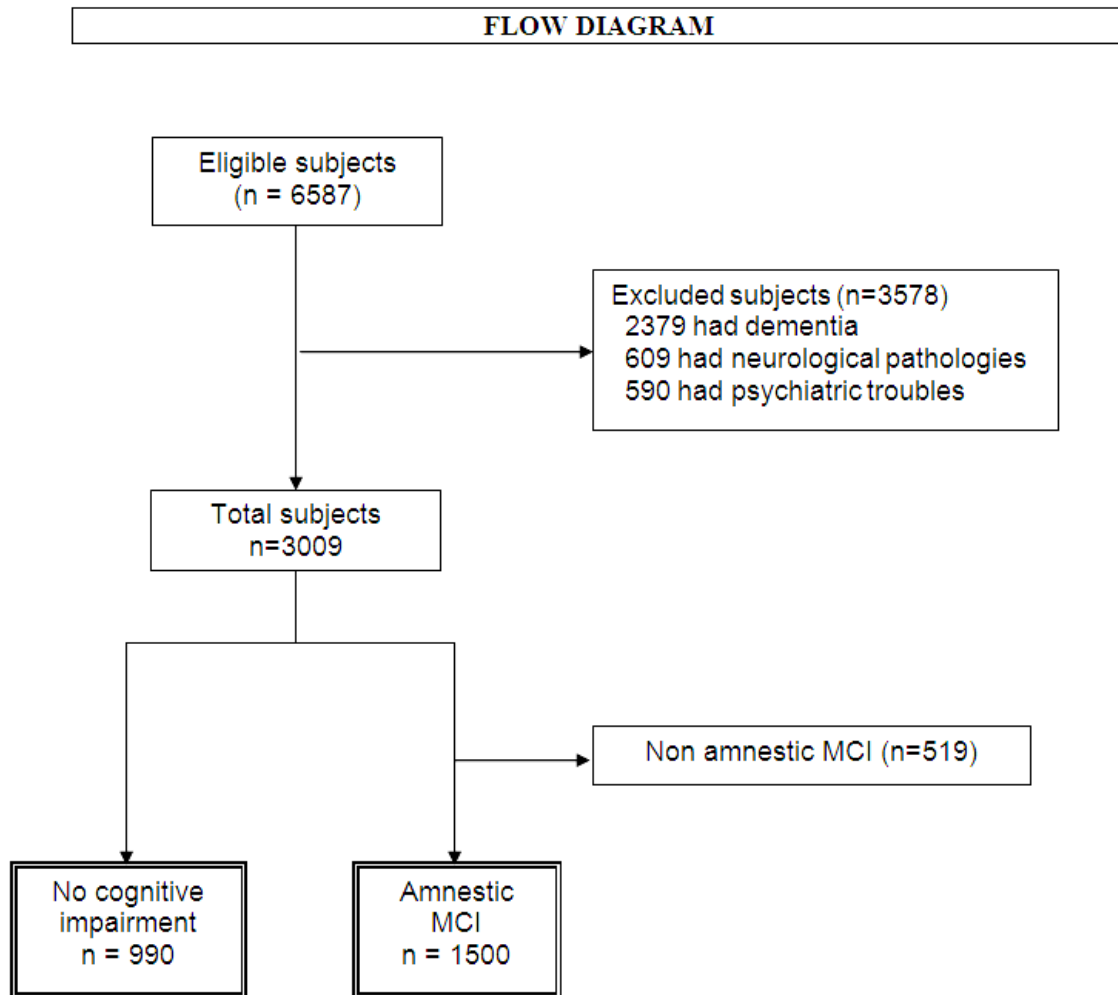
Sensitivity, Specificity and Probability of Dementia of the MMSE, the MIS, the IST, the « or » Rule and the Weighted Sum of the Validation Sample.

Variable	MMSE (0 - 30)	MIS (0 - 8)	IST	“or” rule	Weighted sum
Cutpoint	< 25	< 7	< 27	< 6 or < 25	0
Specificity %	93	90	89	92	93
Sensitivity %, all dementia	87	85	80	92	98
Positive Predictive Value* %	39	31	28	38	42
Sensitivity %, all AD	91	96	80	93	100
Sensitivity %, AD and CDR score 1	78	96	68	86	100
Sensitivity %, all NAD	82	71	81	90	95
Sensitivity %, NAD and CDR score 1	77	67	72	85	92

Abbreviations: MMSE = Mini-Mental State Examination; MIS = Memory Impairment Screen; IST = Isaacs Set Test; CDR = Clinical Dementia Rating; AD = Alzheimer’s Disease; NAD = Non-Alzheimer Dementia.

* Based on a 5% prevalence of dementia

ANNEXE 6. ORGANIGRAMME DE SELECTION DES SUJETS
(ARTICLE 3).



ANNEXE 7. VARIATION DE LA SENSIBILITE ET DE LA SPECIFICITE DU COUPLE MIS-STI (VALEURS SEUILS FIXES) SELON LE TYPE DE MCI EN FONCTION DE L'AGE ET DU NIVEAU SCOLAIRE (ARTICLE 3).

Variation of sensitivity and specificity of MIS-IST combination single cut-off scores according to type of a-MCI by age and education level.

	NCI (n=990) Spec %	very aMCI-s (n = 133) Sens %	very aMCI-m (n = 407) Sens %	aMCI-s (n = 366) Sens %	aMCI-m (n = 594) Sens %
Age, education					
60 - 69 years					
High	95	0	16	40	54
Middle	94	20	41	31	74
Low	83	5	64	47	69
70 - 79 years					
High	90	12	52	53	68
Middle	84	25	60	50	78
Low	76	39	73	84	84
80 - 89 years					
High	81	71	88	71	85
Middle	59	50	90	84	94
Low	52	61	87	91	93

Abbreviations: MIS = Memory Impairment Screen; IST = Isaacs Set Test; Spec = Specificity; Sens = Sensitivity; NCI = No Cognitive Impairment; very aMCI-s = very amnesic Mild Cognitive Impairment single domain; very aMCI-m = very amnesic Mild Cognitive Impairment multi-domain; aMCI-s = amnesic Mild Cognitive Impairment single domain; aMCI-m = amnesic Mild Cognitive Impairment multi-domain.

ANNEXE 8. VARIATION DE LA SENSIBILITE ET DE LA SPECIFICITE DU COUPLE MIS-STI (VALEURS SEUILS AJUSTEES) SELON LE TYPE DE MCI EN FONCTION DE L'AGE ET DU NIVEAU SCOLAIRE (ARTICLE 3).

Variation of sensitivity and specificity of MIS-IST combination adjusted cut-off scores according to type of a-MCI by age and education level.

	NCI (n=990) Spec %	very aMCI-s (n = 133) Sens %	very aMCI-m (n = 407) Sens %	aMCI-s (n = 366) Sens %	aMCI-m (n = 594) Sens %
Age, education					
60 - 69 years					
High	93	14	37	48	72
Middle	94	20	41	31	74
Low	87	5	55	47	67
70 - 79 years					
High	90	25	47	40	76
Middle	89	10	53	45	71
Low	87	17	62	70	69
80 - 89 years					
High	93	28	53	52	77
Middle	92	50	73	68	79
Low	85	43	53	55	71

Abbreviations: MIS = Memory Impairment Screen; IST = Isaacs Set Test; Spec = Specificity; Sens = Sensitivity; NCI = No Cognitive Impairment; very aMCI-s = very amnesic Mild Cognitive Impairment single domain; very aMCI-m = very amnesic Mild Cognitive Impairment multi-domain; aMCI-s = amnesic Mild Cognitive Impairment single domain; aMCI-m = amnesic Mild Cognitive Impairment multi-domain.

ANNEXE 9. VARIATION DE LA SENSIBILITE ET DE LA SPECIFICITE DU MMSE (*VALEUR SEUIL FIXE*) SELON LE TYPE DE MCI EN FONCTION DE L'AGE ET DU NIVEAU SCOLAIRE (ARTICLE 3).

Variation of sensitivity and specificity of MMSE single cut-off scores according to type of a-MCI by age and education level.

	NCI (n=990) Spec %	very aMCI-s (n = 133) Sens %	very aMCI-m (n = 407) Sens %	aMCI-s (n = 366) Sens %	aMCI-m (n = 594) Sens %
Age, education					
60 - 69 years					
High	98	14	16	19	42
Middle	94	23	28	28	57
Low	78	22	53	33	79
70 - 79 years					
High	92	0	28	40	48
Middle	85	20	51	39	60
Low	66	39	69	66	78
80 - 89 years					
High	86	57	59	46	67
Middle	73	50	67	89	86
Low	46	69	80	74	93

Abbreviations: MMSE = Mini-Mental State Examination; Spec = Specificity; Sens = Sensitivity; NCI = No Cognitive Impairment; very aMCI-s = very amnesic Mild Cognitive Impairment single domain; very aMCI-m = very amnesic Mild Cognitive Impairment multi-domain; aMCI-s = amnesic Mild Cognitive Impairment single domain; aMCI-m = amnesic Mild Cognitive Impairment multi-domain.

ANNEXE 10. VARIATION DE LA SENSIBILITE ET DE LA SPECIFICITE DU COUPLE MMSE (*VALEURS SEUILS AJUSTEES*) SELON LE TYPE DE MCI EN FONCTION DE L'AGE ET DU NIVEAU SCOLAIRE (ARTICLE 3).

Variation of sensitivity and specificity of MMSE adjusted cut-off scores according to type of a-MCI by age and education level.

	NCI (n=990) Spec %	very aMCI-s (n = 133) Sens %	very aMCI-m (n = 407) Sens %	aMCI-s (n = 366) Sens %	aMCI-m (n = 594) Sens %
Age, education					
60 - 69 years					
High	68	57	58	58	69
Middle	80	33	49	54	69
Low	78	22	53	33	79
70 - 79 years					
High	69	50	68	87	82
Middle	85	20	51	39	60
Low	81	22	59	48	65
80 - 89 years					
High	74	71	82	73	79
Middle	73	50	67	89	86
Low	79	39	58	41	55

Abbreviations: MMSE = Mini-Mental State Examination; Spec = Specificity; Sens = Sensitivity; NCI = No Cognitive Impairment; very aMCI-s = very amnesic Mild Cognitive Impairment single domain; very aMCI-m = very amnesic Mild Cognitive Impairment multi-domain; aMCI-s = amnesic Mild Cognitive Impairment single domain; aMCI-m = amnesic Mild Cognitive Impairment multi-domain.

ANNEXE 11. EVALUATION DU BIAIS SPECTRAL POUR UNE VALEUR SEUIL FIXE DU MMSE DANS LE GROUPE DES CAS MCI AMNESIQUES AVEC DEFICITS MULTIPLES (ARTICLE 3).

Ratio of likelihood ratio for MMSE in subgroups of aMCI-m.

Groups	N	LR +	Ratio of LR+	95% CI	LR -	Ratio of LR -	95% CI
Overall		2.78	1	Referent	0.46	1	Referent
Age							
60 - 69 y	401	6.75	2.43 ^a	1.66-3.55	0.48	1.03	0.90-1.18
70 - 79 y	597	2.65	0.95	0.81-1.13	0.49	1.05	0.95-1.16
80 - 89 y	328	2.05	0.74 ^a	0.62-0.87	0.31	0.68 ^a	0.56-0.82
Education							
High	378	4.25	1.53 ^a	1.08-2.16	0.57	1.22 ^a	1.08-1.38
Middle	490	4.14	1.49 ^a	1.17-1.89	0.45	0.96	0.85-1.08
Low	458	2.04	0.73 ^a	0.64-0.84	0.31	0.67 ^a	0.58-0.77

Abbreviations: MMSE = Mini-Mental State Examination; aMCI-m = amnesic Mild Cognitive Impairment multi-domain; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; CI = Confidence Interval.

^a = statistically significant spectrum bias.

ANNEXE 12. ALGORITHME DE REPERAGE DES TROUBLES COGNITIFS.

