HARMONIZATION AND INNOVATION OF COGNITIVE, BEHAVIOURAL AND FUNCTIONAL ASSESSMENT IN NEURODEGENERATIVE DEMENTIAS

Report of a JPND Working Group on Longitudinal Cohorts

October, 2015

JPND research

EU Joint Programme – Neurodegenerative Disease Research

This document is the final report from one of ten working groups commissioned by the EU Joint Programme – Neurodegenerative Disease Research (JPND) in 2014 through a peer-reviewed call for proposals. The working groups were established to address methodological challenges preventing current population- and disease-based cohorts being further exploited for ND research. All ten reports are listed below and are available to download on the JPND website by clicking on the website link at the bottom of this page:

- **HD-READy (High-Dimensional Research in Alzheimer's Disease)**
  Coordinator: Professor M. Afran Ikram, Erasmus University Medical Centre, Rotterdam, Netherlands.

- **Harmonization and innovation of cognitive, behavioural and functional assessment in neurodegenerative dementias**
  Coordinator: Dr Alberto Costa, IRCCS Fondazione Santa Lucia, Rome, Italy.

- **NETCALS (Network of Cohort Assessment in ALS)**
  Coordinator: Professor Leonard van den Berg, University Medical Centre Utrecht, Utrecht, Netherlands

- **21st Century EURODEM**
  Coordinator: Professor Carol Brayne, University of Cambridge, Cambridge, UK

- **Multi-centre cohort-studies in Lewy-body dementia: Challenges in harmonizing different clinical and biomarker protocols**
  Coordinator: Professor Dag Aarsland, Stavanger University Hospital, Stavanger, Norway

- **Developing a methodological framework for trials in presymptomatic neurodegenerative disease – the Presymtomatic Neurodegeneration Initiative (PreNI)**
  Coordinator: Dr Jonathan Rohrer, University College London, London, UK

- **BioLoC-PD: Harmonization of biomarker assessment in longitudinal cohort studies in Parkinson's disease**
  Coordinator: Professor Daniela Berg, Hertie-Institute for Clinical Brain Research and German Center for Neurodegenerative Diseases, Tübingen, Germany

- **Dementia Outcome Measures: charting new territory**
  Coordinator: Professor Gail Mountain, University of Sheffield, Sheffield, UK

- **Body fluid biobanking of longitudinal cohorts in neurodegenerative diseases**
  Coordinator: Dr Charlotte Teunissen, VU University Medical Centre, Amsterdam, Netherlands

- **Realising the potential of cohort studies to determine the vascular contribution to neurodegeneration**
  Coordinator: Professor Joanna Wardlaw, University of Edinburgh, Edinburgh, UK

JPND Website link: http://www.neurodegenerationresearch.eu/initiatives/jpnd-alignment-actions/longitudinal-cohorts/

Table of Contents

Summary

The aim of this project was to provide guidelines for harmonization and innovation of assessment tools to be applied to research in neurodegenerative dementias (NDD), with a special focus on longitudinal cohort studies.

A consensus was achieved on the general recommendations to be followed in developing procedures and tools for neuropsychological assessment. The context, the target population, the cognitive/behavioural variable to evaluate and the exact clinical questions to be answered, should be considered as the main preliminary aspects to address. Validity and reliability are two main psychometric properties that must be known before using a test. Sensitivity, specificity and predictive value are particularly important in the context of NDD, to define normal vs. abnormal conditions and for the purpose of differential diagnosis.

The working group particularly focused on unmet needs and provided specific recommendations on how to design studies that can provide empirical evidence for selecting between methods for diagnosis and monitoring of NDD. These issues refer to the assessment of global functions, memory, language, visual-spatial abilities, executive function, functional abilities, motor and behavioural symptoms. A central issue is represented by the lack of standardized norms for the tests to be used in the different European countries. Another important aspect is the need to improve knowledge about the sensitivity, specificity and predictive value of the currently used tools. For each single domain a consensus was also achieved on the practical recommendations of the assessment tools.

Introduction

The requirement for harmonised assessment of cognitive, behavioural and functional features has been extensively discussed in the field of dementia research (see, for example, the MRC recommendations for a minimal dataset- Wilcock et al., 1989). Surveys of assessment tools used across Europe for the assessment of Alzheimer's disease (AD) have been published by the European Alzheimer's disease consortium (Diaz et al., 2005) and by a task force of the European Federation of Neurological Societies (Maruta et al., 2011). Harmonization recommendations have been proposed in the case of vascular cognitive impairment (Hachinski et al., 2006). In the last few years, the development of new diagnostic criteria for AD (Dubois et al, 2014), posterior cortical atrophy (Crutch et al., 2013), behavioural variant frontotemporal dementia (Raskovsky et al., 2011), progressive aphasia (Gorno Tempini et al., 2011), and corticobasal syndrome (Armstrong et al., 2013) all include a consideration of recommended, disease-specific testing procedures aiming at early diagnosis—

although precise details of specific tests are often lacking, meaning that test selection is open to interpretation and, therefore, to potential inconsistency.

The main aim of this project was to provide guidelines for harmonization and innovation of assessment tools to be applied to research in neurodegenerative disorders affecting cognition across Europe, with a special focus on longitudinal studies. Such guidelines would be an improvement on the current situation for various reasons. First, providing psychometric tests with known validity/reliability is urgent to provide shared quality standards among European professionals. Indeed, the phenomenological manifestations (cognitive, behavioural, motor or functional) of NDD may be differentiated early by accurate clinical and psychometric evaluation. Therefore, good levels of sensitivity and specificity are critical requirements for the tests to be used.

Second, the heterogeneity of the current tools makes it hard to compare results across studies carried out in different countries, therefore impeding research progress. The homogenization of the assessment procedures across European countries would allow researchers to merge data sets from different longitudinal studies aimed at investigating the risk factors for dementia as well as from clinical trials. This information would be crucial in informing early diagnosis and treatment.

Third, many of the tests used in clinical settings were developed several decades ago, often in fields other than NDD (e.g., developmental disorders or stroke). The enormous progress of cognitive neurosciences in the last decades, resulting in novel, translationally relevant information about the organization of cognitive functions in the normal brain, has been exploited only to a very limited extent in proposing new measures with increased sensitivity and specificity to early stages of neurodegenerative diseases. On the other hand, there is a danger of over-simplification in applying this knowledge in NDD where patterns of neuronal loss are often incompletely understood and by far more complex than the focal lesion models that can inform cognitive neuroscience (e.g., Mesulam et al., 2015).

Context

The overall area of application of the guidelines is longitudinal studies, including population-based cohorts of people not selected on the basis of a specific NDD diagnosis as well as targeted and disease-focused cohorts (at-risk, presymptomatic, and manifest disorders).

Terminology

AAT=Aachener Aphasie Test

ACE=Addenbrooke's Cognitive Examination

AD= Alzheimer's Disease

ADCS=Alzheimer Disease Cooperative Study

AML=Amyotrophic Lateral Sclerosis

BADL= Basic Activities of Daily Living

BADS=Behavioral Assessment of Dysexecutive Syndrome

BDAE=Boston Diagnostic Aphasia Examination

BPSD =Behavioral and Psychological Symptoms of Dementia

CBD= Corticobasal Degeneration

DLB= Dementia with Lewy Body

ECAS=Edinburgh Cognitive Assessment

ECog= measurement of everyday cognition

EFNS=European Federation of Neurological Societies

EMAS=Edinburgh Motor Assessment

FCSRT= Free and Cued Selective Reminding Test

FTD=Frontotemporal Dementia

GDS= The Geriatric Depression Scale

GP=General Practitioner

GREFEX = Groupe de Réflexion sur l'Evaluation des Fonctions Exécutives.

IADL=Instrumental Activities of Daily Living

LBD=Lewy Body Disease

MCFSI= Mail-In Cognitive Function Screening Instrument

MMSE= Mini Mental State Examination

MoCA= Montreal Cognitive Assessment

MSA=Multiple System Atrophy

NDD=NeuroDegenerative Dementia

NPI=Neuropsychiatric Inventory

NPS=Neuropsychiatric symptoms.

PCA=Posterior Cortical Atrophy

PD=Parkinson's Disease

PDD=Parkinson's Disease Dementia

PPA=Primary Progressive Aphasia

PSP= Progressive Supranuclear Palsy

PSPRS=Progressive Supranuclear Palsy Rating Scale

UHDRS=Unified Huntington's Disease Rating Scale

UMSARS= Unified Multiple System Atrophy Rating Scale UMSARS

UPDRS=Unified Parkinson's Disease Rating Scale

VaD=Vascular Dementia

WAB = Western Aphasia Battery

Methods

A panel of opinion leaders was identified on the basis of their research productivity in the relevant domains (cognition, motor/perceptual function, behaviour, functional assessment); all have accepted the invitation to take part in the task force. According to the general format of the WG, the plan of activities includes two in-person workshops (marking the beginning and the end of activities).

Additionally, there were web-based conferences during which objectives, tasks and outcomes of the project were discussed and defined.

In detail, the work plan was based on a first workshop, in which participants discussed available evidence and organized small groups (3-4 persons each), who worked together via web based communication systems over the following 5 months. The small groups produced consensus statements about the critical issues of the current methods used to investigate target phenotypic dimensions of dementias. In a final workshop a consensus was reached on the indications proposed by the small groups about the current available assessment tools, and on the points that should be addressed by further research.

All participants contributed to the different phases of the project that followed a stepwise methodology. In detail, target dimensions to be explored were defined according to standard guidelines for the neuropsychological assessment. Therefore, a first "global assessment" area was defined referring to procedures used for the initial screening of the patient's general cognitive status. A second "detailed assessment" area was defined referring to the detailed examination of the neuropsychological profile of the patient. Accordingly in the first workshop, the specific target dimensions of memory, executive functions, language, visual-spatial abilities, motor, behavioural symptoms and functional status were defined, with a plenary consensus.

Subsequently, small groups were created that focused on each of the cognitive dimensions previously mentioned. The task of each small group was to identify unmet needs for each specific domain taking into account psychometric qualities (i.e., validity, reliability, sensitivity, specificity) and feasibility for the current methods and tools of assessment. Within small groups, critical issues in each area were identified based on existing evidence and participants' expertise and clinical practice. For each small group a leader was identified who was responsible for managing all the activities of the small group and for providing first drafts of all consensus documents. The initial indications proposed by small groups were then discussed in web-based conferences involving the whole consortium. In these conferences, statements, procedures and indications by small groups were analytically revised.

A consensus on the tools and unmet needs for each cognitive domain identified as critical in the assessment and monitoring of NDD, was achieved in a final workshop, involving the entire consortium, either by physical participation or skype connection. During the workshop, the leader of each small group presented their outcomes in a plenary assembly. The consensus on each specific issue was achieved with the agreement of all participants.

AGREED GUIDELINES

A preliminary discussion within the WG was about the general format of the work activities, in relation to the expected output (guidelines to inform cohort studies of NDD). Several possibilities were considered:

Systematic review. While a systematic review of assessment tools might be expected to provide useful information and may be worth being promoted, the WG concluded that there is an absence of empirical evidence (especially in comparing different testing methods head-to-head) to underpin strong recommendations. This approach would therefore be of limited usefulness in informing guidelines, in particular because of the major changes in the field of NDD research during the last few years. With a few recent exceptions, studies on the psychometric properties of neuropsychological instruments report sensitivity and specificity measures comparing patients with AD to healthy controls, while the focus of contemporary research is largely on the predictive value of tests applied in the very early (predementia) stages for the development of specific forms of NDD.

Practice survey. The collection of data about the tests used in clinical practice in Europe, which is of course crucial for the harmonization of procedure, has already been achieved by a EFNS task force (Maruta et al., 2012), and is only in need of a minor update.

It was thus agreed that the most useful process was to focus the discussion on the unmet needs in this area, according to two main themes: i) general recommendations on the procedures and tools to be used for the assessment of cognitive disorders in NDD; ii) recommendations referring to specific domains.

General recommendations

In general terms, the usefulness of a test depends on the situation (e.g. screening versus diagnostic use), the person being assessed (e.g. healthy elderly, pre-dementia and dementia patients), and the questions being asked (e.g. discrimination versus the measurement of changes over time or diagnostic sensitivity vs sensitivity to change). Accordingly, the guidelines cannot be general, but need to be tailored in consideration of:

(a) The situation (e.g. GP practice, memory clinic, specialist clinic or research centre)?

(b) The question being addressed (e.g. diagnosis versus longitudinal change; screening for evidence of dementia amongst the healthy, community-dwelling elderly vs symptomatic clinic-based cohorts etc.);

If evaluating healthy individuals, these questions might include:

- Is the aim to screen a population for subtle but not currently clinically relevant cognitive decline as a predictor of future impairment/disease?
- Is it necessary to refer the patient for detailed cognitive assessment? E.g. Is the aim to discriminate between cognitively normal and abnormal [for age])?

In the context for example of a cognitive disorders clinics, these questions might include:

- Can we discriminate between prodromal AD and other reasons for cognitive symptoms?
- What is the underlying pathology or phenotype? e.g. Is the aim to discriminate between phenotypes (e.g. typical vs atypical AD) and/or underlying pathologies (e.g. AD vs DLB)?
- Has the patient changed compared to his/her previous visit?

(c) Which factors may affect cognition or behaviour in a specific population? (e.g. age group, culture, disease state, disease stage, risk factors and related medical problems)

Having established the context, the population(s) of interest and the questions to be addressed, a number of general principles that refers to the psychometric characteristics of the tools to be used can help to guide the selection of specific tests.

a) Validity (i.e. does a task measure what it is intended for?), including:

- Theoretical, neuroscientific and clinical basis of the test design (content validity). This includes several aspects, from the anatomical basis (e.g. matching memory demands to anatomical regions of interest; accepting however that the neural basis of deficits in NDD is typically complex and incompletely understood) to clinical relevance (e.g. involvement of target population for the development of relevant items in behavioural scales).

- Evidence that the instrument is measuring the construct of interest (construct validity), e.g. ensuring that memory test performance reflects memory function, rather than visual function etc.
- Suitability for the target population: Tests are generally developed for a specific situation or purpose. When applied in a new situation or for another purpose, they need to be revalidated. [reference to FDA guidelines for 'new situations']
- Purpose of measurement: The validation performed by the developers should correspond to the measurement purpose of the user – e.g. for screening purposes in the general population, the task should have been validated for discriminating between dementia and healthy ageing. For diagnostic purposes in a memory clinic, the task should have been validated for discriminating between different types of dementia. For AD trials, the tasks/battery should be targeted at cognitive impairments expected in AD. For at-risk or pre-dementia stages, the test needs to have high positive and negative predictive value.

Some general points apply to all measures. The ideal test should be:

- insensitive to the cultural background and literacy levels of those assessed (cross-cultural validity)
- clinically relevant, with a known relationship to real-life impairments.
- have a high discrimination value. This includes more than sensitivity and specificity, including, depending on the measurement purpose:
  - Prediction of AD: be sensitive to cognitive disorders in the very early stages of AD
  - Screening: high sensitivity and specificity for cognitive deficit versus no deficit.
  - Responsiveness: be responsive to changes in cognition and discriminate between patients who change and patients who remain stable.

b) Reliability:

- Provide stable measurements in stable subjects (test-retest reliability) and by different raters (inter-rater reliability)
- Minimise improvement as a result of repeated testing effects in healthy ageing (practice effects).
- Avoid floor- and ceiling effects depending on the testing context (e.g. screening for early deficits necessitates tests where the normal population cannot score perfectly in order to detect the earliest changes).

c) Feasibility:

- Interpretability of scores
- Be non-invasive with minimal discomfort to the patient
- Be quick to administer and inexpensive (depending on the context)

Recommendations referring to specific domains

<u>Global Functioning</u>

General considerations. Due to the high prevalence of Alzheimer's disease (AD) among the wide spectrum of neurodegenerative diseases of the elderly, the literature has focused on this type of dementia, causing an "alzheimerization" of screening tests used in clinical practice, with the consequence that a large proportion of instruments have focused on memory at the expense other

cognitive domains. Furthermore emphasis on cut-off scores rather than on cognitive profile, can be counter-productive for differential diagnosis.

Clinical relevance. An ideal test should include sub-items able to detect cognitive impairments belonging to the following domains (memory, language, visuo-praxic and executive functions) in order to generate specific profiles for the major (in epidemiological terms) neurodegenerative diseases: Alzheimer's disease (AD), Fronto-Temporal Dementia (FTD) or Dementia with Lewy Bodies (DLB).

Challenges and unmet needs. The Mini Mental State Examination (MMSE; Folstein et al., 1975) has a test re-test reliability ranging from 0.80-0.95 with test-retest bias due to ceiling effect for short time interval, has no parallel forms yielding to a misclassification of about 15% of patients. Furthermore it shows low sensitivity in highly educated people who can obtain normal scores despite being impaired. In contrast, false positives may arise in people with low educational level who can obtain low scores even if unimpaired. Moreover MMSE has heavy reliance on memory and language but not abstract thinking and executive functions (Nieuwenhuis-Mark, 2013) and visuospatial ability. However, the intersecting-pentagon copy sub-item, was found to be reliable in differentiating between mild AD and DLB (Caffarra et al, 2013; Mitolo et al, 2014) in an autopsy verified population. The Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) is another brief screening test for cognitive disorders which consists of 30-points test administered in 10 minutes focusing on memory, visuo-spatial, executive, language and orientation to time and place. MoCA has the advantage to be short and easy to administer with a good sensitivity (86%) for Mild Cognitive Impairment (MCI) and AD (97%) and for the post-stroke cognitive screening, being able to detect MCI patients at risk to develop AD at a rate of 90.5% for total score < 20/30. Among tools allowing a more detailed, yet short, screening of cognitive functions the Addenbrooke's Cognitive Examination (ACE-R; Mioshi et al., 2006) appears to be quite promising. Indeed, a recent meta-analysis on dementia studies evidenced that sensitivity and specific for ACE-R were 95.7% and 87.5%, respectively, with higher diagnostic accuracy than the MMSE in contexts with both modest and high prevalence of dementia (Larner and Michell, 2014).

Cognitive screening poses particular problems in patients with motor symptoms, since the test results can be influenced by dysarthria, tremor, rigidity, weakness, apraxia etc. Edinburgh Cognitive Assessment (ECAS), originally designed for patients with Amyotrophic Lateral Sclerosis (ALS), minimizes the impact of motor dysfunction on cognitive performance and can be used in patients with motor difficulties (Abrahams et al 2014).

Recommendations. The MMSE and MoCA were found to be valid as screening measures in primary setting. Therefore, they are recommended. A comprehensive test battery - the ACE-R is a good example of a scale allowing a short screening of several cognitive functions- is more suitable for secondary or tertiary specialist settings, with tasks specific for cognitive networks mostly involved in different dementia syndromes.

Executive functions

General considerations. Executive function encompasses a series of high-level processes, the main aim of which is to facilitate adaptation to complex or new situations, when highly practiced cognitive abilities no longer suffice. Neuroimaging studies have suggested that executive functioning relies on a distributed cerebral network encompassing frontal and posterior associative cortices.

The unitary models of executive functioning propose that executive deficits experienced by brain-damaged patients result from difficulties to create and maintain action plans or by a weakening of

the information temporarily stored in working memory. Multi-component models distinguished between several executive processes on the basis of single-case analyses or exploratory/confirmatory factorial analyses. Among these processes, three were clearly identified: updating, shifting and inhibition (Miyake et al., 2002). Other processes more related to the temporal or sequential organization of cognitive processes were also associated to executive functioning, such as the expression of action plans before the executive, assessment and editing stages of these plans. Above functions are generally defined as "cold" executive functions (i.e., due to the engagement of purely cognitive processes). Abilities that involve emotional or social processing are defined as "hot" executive functions (e.g., affective theory of mind abilities).

Clinical relevance. Differential diagnosis is a matter of debate. Early onset AD and FTD patients have both shown impaired performance on executive tests. However, the two patients' groups show qualitatively distinct profiles (Stopford et al, 2012). In a recent study, letter fluency and assessment of social abilities differentiated the behavioural variant of FTD (bvFTD) from AD (Possin et al, 2013).

Challenges & unmet needs. The divergent viewpoints in cognitive neurosciences partly explain difficulties in selecting executive tests suitable for measuring executive abilities in clinical trials. Short batteries have been designed. The Frontal Assessment Battery consists of six subtests, assessing conceptualization, conflicting instructions, motor programming, sensitivity to interference, motor inhibitory control and prehension behavior. The INECO Frontal screening includes motor programming, conflicting instructions, motor inhibitory control, numerical working memory, verbal working memory, spatial working memory, abstraction capacity (proverbs), and verbal inhibitory control (modified Hayling test). The differential weight (and interest) of the different subtests remains debatable.

Another approach is to combine different tests, and look for the number of failed tests to assess executive dysfunction (see Godefroy and GREFEX, 2008). The GREFEX battery is composed of a questionnaire (Behavioral dysexecutive syndrome inventory) and seven cognitive tasks (Stroop task, modified 6 elements test, Trail-Making test, Adapted version of the Brixton task, Baddeley dual task, verbal fluency and modified card sorting test). Cut-off scores may be derived. For example, impaired performance on three scores have to be observed to consider the presence of dysexecutive syndrome. A similar battery is the Behavioral Assessment of Dysexecutive Syndrome (BADS; Wilson et al., 1998) that includes several subtests exploring the different dimensions of the executive domain.

A relevant matter is represented by the poor explored ecological validity of executive tests used. Another question is that executive batteries do not include assessment of "hot" functions.

Recommendations

Taking into account the multifaceted nature of the executive domain, the use of a battery that includes subtests investigating different executive components is recommended**.**

Functional assessment

General considerations. Functional assessment can be subdivided into 'instrumental activities of daily living' (IADL) and 'basic activities of daily living' (BADL). The latter refers to activities such as eating, dressing and toileting (Lawton & Brody, 1969)) Generally, these activities remain preserved until the later stages of dementia (e.g., see Peres et al., 2008). IADL refer to higher order activities, sensitive to cognitive disorders such as cooking, doing finances and shopping(Nygard et al., 2003).

Clinical Relevance. IADL strongly influence everyday lives of patients and caregivers, and assessment is essential for the planning of care needs, legal and financial matters. In clinical and research practice, objectives include diagnostic use, because the extent to which cognitive disorders influence everyday functioning is part of the diagnostic criteria for dementia.(McKhann et al. 2011). Second, early detection is an objective, as several studies demonstrated subtle IADL impairments in MCI and healthy elderly (e.g. Sikkes et al., 2011). Third, It is relevant for the measurement of change over time. In clinical trials, IADL is an outcome measure required by the Food and Drug Administration, to demonstrate clinically relevant changes (Desai et al., 2004).

Challenges & unmet needs. The relevance of a functional assessment resulted in an abundance of IADL instruments, ranging from self-report questionnaires and informant-based questionnaires to performance-based instruments (Sikkes et al. 2009; Sikkes et al., 2014) . A first unmet need is that there are large content differences. Instruments differ in whether they combine IADL with BADL and/or cognition, such as the Disability Assessment for Dementia (Gelinas et al. 1999). This is problematic, as combining IADL and BADL into a single instrument results in instruments less sensitive to cognitive impairments (e.g. Royall et al., 2007). Second, the quality of many IADL instruments is limited. In several (systematic) reviews it was shown that published IADL instruments lacked evidence on content validity, reliability, construct validity and responsiveness (e.g. Marshall et al. 2012, Sikkes et al. 2009; Burton et al. 2009; Law et al. 2012).

Recommendations. Quality limitations currently hamper functional assessment. Although further research is needed to ensure quality, promising results have been found for several tools. Including the measurement of everyday cognition (ECog) (Farias et al. 2008), the Mail-In Cognitive Function Screening Instrument (MCFSI) (Walsh et al. 2006) and the Alzheimer Disease Cooperative Study (ADCS). The Amsterdam IADL Questionnaire is a tool addressing many issues outlined above (Sikkes et al. 2012; Sikkes et al. 2013; Sikkes et al. 2014). In general, functional assessment could benefit from modern psychometric approaches such as item response theory, adaptive and computerized assessments.

<u>Language Assessment</u>

General considerations. Language tasks are probes of the structural and functional condition of the large-scale network responsible for language processing. This network is more extended than the classical language areas of the left hemisphere (Price, 2012). Most language tasks require additional cognitive resources (such as visual perception, sustained attention and working memory) for an adequate performance. These issues should be kept in mind when considering the diagnostic contribution of language task in current usage for NDD.

Clinical Relevance. The indications for language assessment in NDD are multiple, including early diagnosis, follow-up and assessment of treatment outcomes. Early diagnosis requires considerations of sensitivity, specificity and predictive value. Lexical semantic tests are sensitive tools for screening, and are included in minimal datasets for AD. In contrast, the analysis of sensorimotor, phonological and morphosyntactic levels is crucial for its role in the specific diagnosis of NDD involving the perisylvian region of the language dominant hemisphere (primary progressive aphasias, PPA).

Challenges & unmet needs. Many language tasks have been developed for the assessment of language disorders (aphasia) due to focal brain lesions (most commonly stroke). Only a minority have been developed to assess those NDD in which language impairment is either one of the manifold neuropsychological features (AD, PD) or the main determinant of clinical presentation (PPA). A basic distinction is between tasks assessing the ability to use single words (lexical-semantic

tasks) and evaluations of sentence processing. The most widely used single word tasks are based on picture naming, where performance is influenced by a wide array of factors (e.g., Akinina et al., 2014). Cross linguistic databases are available (e.g., Bates et al., 2003) and there have been effort to standardize naming tests for several European languages (Kremin et al. 2003).

The clinical use of extensive language examinations is required only in the case of PPA. Again, the usual practice has been to apply tests developed for stroke aphasia, such as the Boston Diagnostic Aphasia Examination (BDAE), the Western Aphasia Battery (WAB), or (in Europe) the Aachener Aphasie Test (AAT) and the Montreal-Toulouse Aphasia battery. Only more recently has there been an interest in the development of specific testing procedures. These include the Cambridge semantic memory test (Adlam et al, 2010) and the SYDBAT (Savage et al., 2013).

The importance of harmonization in the area of language assessment is evident. Language tasks typically require not only translation but also adaptation. The cross-linguistic differences (for example, the morphological structure or the impact of orthography on written language tests) need to be considered in the comparison among patients from different linguistic backgrounds. The requirements for clinical use are different in disorders not primarily characterized by language dysfunction ("typical" AD, PD, ALS), and in the group of PPA. While in the former case the main need is a harmonization of existing core measures (naming, verbal fluency), in the latter field a short yet more detailed examination is needed allowing early and differential diagnosis. An additional need is a standardized functional measure of communication abilities.

Recommendations. The extent of language assessment should be closely linked to the aim of the investigation. In most setting, picture naming is the test of choice. A consensus on a high-quality cross-naming task is heavily needed. In investigations of progressive language disorders, a more comprehensive evaluation including an analysis of extended speech production as well as sentence-level tasks is recommended. Communication abilities should be assessed also as a key component of the functional profile

Motor Assessment

General considerations. The practice of motor assessment in memory clinics across the world is highly variable; according to a recent survey (337 responses from 33 countries, presented as a poster at the EAN meeting in Berlin on June 20-23, 2015;Symonds & Bak, 2015) 25% of neurologists and 70% of psychiatrists do not assess motor functions regularly. Most clinicians underestimate the frequency of motor symptoms in memory clinic (according to the above-mentioned survey, most clinicians believe that only 0-20% of patients in cognitive clinic have motor problems, while the literature points to a much higher frequency of 30-50%). Indeed, only a small proportion of clinicians uses motor scales.

Clinical Relevance. In subcortical dementias such as, for instance, PDD, PSP, CBD, HD and VAD, motor involvement is a key sign (Bak et al., 2005; Bak, 2010). In some cases the onset of a cognitive-motor feature can be pathognomonic, e.g. apraxia in CBD) (Bak and Hodges, 2008), although it can also occur in AD (Lesourd et al., 2013). Systematic studies also show that motor signs are common even in classical cortical related cognitive diseases such as AD (Scarmeas et al., 2004) and PCA (Ryan et al., 2014); their presence is usually associated with a worse prognosis (Scarmeas et al., 2005). In MCI, it can herald a later conversion into the Dementia with Lewy Bodies (Molano et al., 2010).

Challenges & unmet needs. The only commonly used scale is the UPDRS, which has been designed specifically for PD and does not cover, therefore, many phenomena which can be of importance in dementia patients (e.g. amyotrophic features pointing to ALS). There is no brief motor screening tool

with standard norms and validity data that would not be disease-specific and could be used in a wide range of diagnoses in a memory clinic (a kind of motor equivalent of MMSE, ACE or MoCA). This tool should contain items investigating areas of motor function relevant to dementias: Parkinsonian symptoms (e.g. PD. DLB, PSP, CBD, MSA); amyotrophic features (ALS/FTD); Cerebellar features (e.g. MSA, alcoholism, SCA-17 etc); complex movements such as praxis, motor sequencing etc (e.g. CBD, PCA)

Recommendations. At the current state of the art, a global tool that allows the screening of motor symptoms associated to different dementias is not available. The use of the following disease-specific tools that show good validity level is recommended: UPDRS (Unified Parkinson's Disease Rating Scale); UHDRS (Unified Huntington's Disease Rating Scale); PSPRS (Progressive Supranuclear Palsy Rating Scale); UMSARS (Unified Multiple SystemAtrophy Rating Scale UMSARS); Clinical examination according to El Escorial criteria for ALS. The development of a brief, clinically applicable and easy to use non-disease specific motor screening tool should be a matter of high priority. The recently developed Edinburgh Motor Assessment (EMAS) (Bak et al., 2015) could fill the gap, but it will require collection of normative data in patients and healthy controls of different ages and a validation against the above-mentioned established motor tests.

Memory

General Considerations. While it is evident that memory is particularly affected in AD, there is still very little understanding on how the progression of AD pathology maps onto specific memory circuits. The hallmarks of AD, amyloid and tau pathology, begin in and progress along partly distinct brain circuits (Braak et al., 2013; Lace et al., 2009; Perani, 2014). Meanwhile the pattern of synaptic and neuronal loss, which are the putative substrates for cognitive impairment in AD, do not map completely onto either of histopathological hallmarks but appear to affect distributed networks. Therefore, instead of domain-specific memory tests, assessments that reach across domains, such as spatial navigation tasks which usually require integrating spatial information and object information, may have the advantage that they are sensitive to the complexity of AD pathology and hence may be particularly suitable for the purposes of staging disease severity.

Clinical Relevance. Memory impairment is the hallmark of typical AD. However, memory disorders are common to several NDD and occur in other non-neurologic conditions including depression and even healthy normal aging. Hence, memory tests should have good level of sensitivity and specificity to assist differential diagnosis. Indeed, different memory deficits profiles have been described in the early stages of NDD according to the underlying neuropathologial processes. AD patients show a memory profile due to consolidation failure whereas FTD and subcortical dementias typically show memory disorders due to encoding and retrieval failures (Dubois et al., 2007; 2014). The assessment of the qualitative characteristics of memory disorders is, thus, crucial for the differential diagnosis.

Challenges and unmet needs.

"Classic" memory tests (e.g., the various forms of auditory verbal learning test; Delis et al., 1991; Vakil et al., 1997) may lack specificity for AD diagnosis. Although sensitive, many reach floor performance early in the disease course making them unsuitable for longitudinal assessment. There is considerable interest in theory-driven tests like the Free and Cued Selective Reminding Test (FCSRT; Grober et al., 2010; Lemos et al., 2014) or the Temporary Memory Binding (Della Sala et al., 2012; Parra et al., 2009) though more work is required to investigate whether their theoretical basis translates to real-world advantages s over existing methods. Another new research focus is on prospective memory procedures (Thompson et al., 2010) a (Costa et al., 2015). The FCSRT has a high

predictive validity for identifying patients with biomarkers of Alzheimer pathology (Wagner et al, 2012). The central issue of also these new approaches is their predictive value in very early/prodromal AD and pre-symtomatic carriers. Over and above sensitivity and specificity, issues to consider include: age effects, practice effects, ceiling and floor effects, replicability, ease of administration, and correlation with bio-markers.

There are also a number of memory tests which are incorporated in various computerized batteries (Weintraub et al., 2013; Troyer et al., 2014). These are interesting alternatives for the existing memory tests, incorporating novel strategies for the study of memory (eg. evaluation of visuo-spatial memory with the help of navigation ability as it is studied with virtual reality methods). However, the Psychometric properties of these tests have not been well studied yet and only few comparisons of between them exist.

Recommendations. Memory testing is a critical area with several new lines of investigation that need validation and empirical evidence to assess how they may advance the field over existing methods. Among extant tools, the FCRST and the Temporary Memory Binding tests are promising candidates to discriminate between AD-related memory deficits and memory disorders occurring in other NDD.

Visual-spatial functions

General Considerations. Dementia-related visual impairment tends to be neglected in both clinical and research practice, partly owing to assumptions that visual problems are due to primary ocular age-related changes or disease, and because it sometimes occurs at a point when language and other skills are too impaired for the person with dementia to explain the perceptual problems they are having.

Clinical relevance. Visual impairments have a profound impact upon everyday life, with previous research demonstrating spatial perception is more strongly associated with activities of daily living than episodic and verbal short-term memory (Perry and Hodges, 2000; Glosser et al., 2002). Visual impairment also frequently contributes to problems as diverse as falls, poor diet (e.g. Dunne et al., 2004) and challenging behaviours, hallucinations and delusions (e.g. agitation because curved patterns on curtains are misperceived and misinterpreted as snakes slithering down the wall).

Visuoperceptual and/or visuospatial impairments (and/or a clinical history of object and space perception problems) are central to the diagnostic criteria for Alzheimer's disease (e.g. McKhann et al., 2011, Dubois et al., 2014), Dementia with Lewy Bodies (McKeith et al., 2005), corticobasal syndrome (Mathew et al., 2011) and are the primary impairment in posterior cortical atrophy (PCA; Crutch et al., 2012).

Challenges and unmet needs. One limitation in many evaluations of cortical visual function is that assessments are limited to higher order visuoperceptual and visuospatial processing without consideration of the integrity of fundamental, basic visual processes supported by striate and extrastriate occipital cortex (e.g., basic form, color, motion, and location processing) and concomitant deficits in the spatial orienting of attention. Similarly ocular disease (e.g. macular degeneration, cataract) can influence symptomatology and cortical visual test performance and so must be assessed. Otherwise it is difficult to determine whether higher-order object and space perception deficits are attributable to parietal temporal tissue loss, or to a more basic deafferentation of such areas owing to occipital lobe or ophthalmological disease.

As with all cognitive domains, there is a need to consider the validity of certain tasks for specific populations. For example, certain VOSP visuospatial subtests are better suited to individuals with moderate typical AD as the instructions for the Number Location subtest are less readily comprehended than those for the Dot Counting subtest. Conversely, there is a need to consider the impact of cortical visual dysfunction on other domains of cognitive and motor test performance, particularly given the number of neuropsychological assessments with visually-mediated instructions, stimulus presentation or response formats. For example, for patients with prominent cortical visual impairment, tests of episodic memory with explicit visual demands in encoding and/or retrieval (e.g. Rey-Osterrieth figure copy) are unsuitable, and a majority of tests of executive functions are partly or wholly visually mediated. Less obvious are the more implicit visual demands of tests such as verbal paired associate learning that often draw on mental imagery. Practical implications here include the selection of participants and outcome measures for clinical trials. For example, some AD clinical trials in the past have permitted recruitment of individuals with PCA attributable to probable AD despite using visually-mediated episodic memory tasks (visual paired associate learning) as one of the cognitive outcome measures.

Recommendations. Commonly used batteries to detect deficits in early visual processing (e.g. figure-ground discrimination) and object and space perception include the Visual Object and Space Perception battery (VOSP; Warrington and James, 1991), Birmingham Object Recognition Battery (BORB; Riddoch and Humphreys, 1993). The following Individual tests can be used for a short screening : Rey-Osterrieth figure (Osterrieth, 1944), and Benton Judgement of Line Orientation (Benton et al., 1978). Presently, most available tests are limited by ceiling effects.

Behavioural symptoms

General Considerations. Neuropsychiatric symptoms (NPS) or Behavioral and Psychological Symptoms of Dementia (BPSD) are heterogeneous and include the following: agitation, aberrant motor behavior, anxiety, depression, elation, irritability, apathy, disinhibition, delusions, hallucinations, elations, stereotypic behavior, sleep or appetite changes and abnormal sexual behavior. The assessment of BPSD usually requires a long examination to elicit information regarding the patient's habits, the subjective experience and the objective behavior and for this reason the caregiver's collaboration is crucial to obtain reliable data. There is evidence of a stage effect of BPSD in Alzheimer's disease. Affective symptoms such as depression, anxiety and apathy in mild AD (Feldman et al., 2004); psychotic signs as delusions, hallucinations, disinhibition, disturbances in motor function and aberrant vocalizations in moderate to severe AD (Mayer et al., 2006). In contrast, in FTD prominent and disabling BPSD are the hallmark of even early stages. Even though BPSD are considered expression of diffuse brain impairment, a consolidated literature assigns mood and psychotic symptoms to different subcortico-cortical (i.e., right hippocampus, amygdala, anterior cingulate gyrus, frontal, temporal, limbic regions and visual association; Cerejeira et al 2012). The detection of BPSD is currently possible by means of a range of questionnaires focused on single or multiple signs.

Clinical relevance. BPSD affect the majority of persons with dementia being present in up to 90% of all dementia subjects over the disease course (Cerejeira et al, 2012). BPSD are also frequent in MCI (35-85%; Monastero et al 2009). The prevalence trends is higher in institutionalized subjects (up to 96%) than in community-dwelling subjects (56-98 %), and BPSD might be present individually or co-occur simultaneously in the same patient. Moreover, BPSD are strongly associated with functional and cognitive decline in the affected patient.

Challenges and unmet needs. Most of the tests designed to detect BPSD encompass all behavioral domains, few tests however allow detailed examination of one specific domain or show limited sensitivity to change when compared to measures that incorporate clinician judgment. An important issue in standardizing new BPSD scales relates to the choice of single or multiple domains scales. Moreover, many the profile of BPSD occuring in Alzheimer's disease (AD), Fronto-temporal Dementia (FTD), Vascular Dementia (VaD) and dementia with Lewy Body (DLB) varies greatly posing a further challenge. Even with a single disease there may be considerable heterogeneity. For instance, recently in a large population of AD subjects five distinct neuropsychiatric syndromes were found: the apathetic syndrome (as unique syndrome) was the most frequent, followed by affective syndrome (anxiety and depression), psychomotor (agitation, irritability, and aberrant motor behavior), psychotic (delusions and hallucinations) and manic (disinhibition and euphoria) syndromes (Spalletta et al, 2010). More than 75% of AD patients presented with one or more of the syndromes. However, apathy, delusions and agitation occur also in the other NDD.

Recommendations. The Neuropsychiatric Inventory (NPI) is a reliable scale for the screening of a wide range of BPSD (Cummings et al. 1994) and, thus, is recommended. The Dimension Apathy Scale

(Radakovic et al 2014; Levy and Dubois, 2006) and The Geriatric Depression Scale (GDS-15; Yesavage et al. 1982) are suitable for the assessment of individual affective dimensions.

Table 1. The table summarises the key guidelines of the Working Group

Guidelines for assessment in NDD	
Aims definition ("horses for courses")	
Context	e.g., GP practice; memory clinic; specialist clinic; research centre
Target population	e.g., age, education, cultural factors, disease state, disease stage, risk factors and related medical problems
Definition of the clinical/research question in the target population	Healthy subjects: risk for future cognitive decline; screening or detailed cognitive assessment. Patient populations: discrimination between prodromal and clinical phases; underlying pathology or phenotype; monitoring of changes; differential diagnosis
Psychometric issues	
Validity (i.e. does the task actually measure what it is intended to	Theoretical, neuroscientific and clinical basis of the test design; Sensitivity and specificity in relation to the

measure?)	clinical/research question; positive/negative predictive value	
Reliability	Stable measurements intra-subject and inter-raters; minimize floor- and ceiling effects	
Feasibility	non-invasive with minimal discomfort to the patient; quick to administer and inexpensive (depending on the context)	

Guidelines for specific assessment domains		
Assessment domain	*Unmets needs*	*Recommendations*
Global Functioning	For initial screening: - effect of education - ceiling effect - do not cover all cognitive domains - lack of parallel forms Batteries for secondary examination: - do not cover all cognitive domains - low validity	The MMSE and MoCA were found to be valid as screening measures in primary setting. Therefore, they are recommended. A comprehensive test battery -the ACE-R is an example of a scale allowing a short screening of several cognitive functions- is more suitable for secondary or tertiary specialist settings, with tasks specific for cognitive networks mostly involved in different dementia syndromes.
Executive abilities	-Poor ecological validity of executive tests used -Executive battery do not include validated tests for assessment of orbito-frontal (ventromedial) functions -Lack of a full agreement on the modular organization of executive domain -Feasibility limitations	Taking into account the multifaceted nature of the executive domain, the use of battery that includes subtests investigating the different components is recommended.
Functional abilities	Many IADL tools lack evidence on: -validity, -reliability -responsiveness	In general, functional assessment could benefit from modern psychometric approaches such as item response theory, adaptive and computerized assessments.
Language	-Lack of short and detailed examination for differential diagnosis in NDD that are not primarily characterized by language disorders. , -There are not standardized tools for the functional assessment of	In general terms, a comprehensive evaluation including an analysis of extended speech production as well as sentence-level tasks is recommended. In most setting, however, picture naming is the test of choice

	communication -Cross linguistic differences should be taken into account in choosing the test -Lack a high-quality cross-naming task.	
Motor symptoms	There is no brief a global tool that allows the screening of motor symptoms associated to different dementias	The use of the following disease-specific tools that show good validity level is recommended: UPDRS ; UHDRS PSPRS; UMSARS; UMSARS; Clinical examination according to El Escorial criteria for ALS.
Memory	- Some memory measures are highly sensitive but lowly specific - Many measures are affected by ▫ age effects, ▫ practice effects, ▫ ceiling and floor effects ▫ low replicability, ▫ administration difficulty ▫ poor predictive validity.	Memory testing is a critical area with several new lines of investigation that need validation and empirical evidence to assess how they may advance the field over existing methods. Among extant tools, the FCRST and the Temporary Memory Binding tests are interesting candidates to discriminate between AD-related memory deficits and memory disorders occurring in other NDD.
Visual-spatial functions	-Relatively lak of specificity: performance on some tests could be attributable to basic visual processes and concomitant deficits in the spatial orienting of attention. -Validity for diagnosis and monitoring -Ceiling effect	Comprehensive batteries such as VOSP and BORB are recommended. For a short screening of visual-spatial functioning the Rey-Osterrieth figure and Benton Judgement of Line Orientation can be used.
Behavioural symptoms	-Many scales include symptoms that are not disease-specific -Few validated scales are available for detailed assessment of individual dimensions	The Neuropsychiatric Inventory (NPI) is a reliable scale for the screening of a wide range of BPSD and, thus, is recommended. The Dimension Apathy Scale and the Geriatric Depression Scale, GDS-15 are suitable for the assessment of individual affective dimensions.

Contributors

Name	Institute	Country
Alberto Costa (project manager)	IRCCS Fondazione S. Lucia and Niccolò Cusano University	**Italy**
Thomas Bak	University of Edinburgh	**Scotland, United Kingdom**
Mathieu Ceccaldi	University Hospital La Timone and Aix Marseille University	**France**
Paolo Caffarra	University of Parma	**Italy**
Carlo Caltagirone,	University of Rome and IRCCS Fondazione S.	**Italy**

	Lucia	
Stefano Cappa	IUSS Pavia and San Raffaele Scientific Institute of MIlano	**Italy**
Sebastian Crutch	University College of London	**United Kingdom**
Sergio Della Sala	University of Edinburgh	**Scotland, United Kingdom**
Bruno Dubois	Institut de la Mémoire et de la Maladie d'Alzheimer (IMMA), Hôpital de la Pitié-Salpêtrière	**France**
Emrah Duzel	German Center for Neurodegenerative Diseases (DZNE) and University College of London	**Germany and United Kingdom**
Peter Nestor	German Center for Neurodegenerative Diseases (DZNE)	**Germany**
Sokratis Papageorgiou	University of Athens Medical School	**Greece**
Martin Rossor	University College of London	**United Kingdom**
Eric Salmon	University Hospital of Liege	**Belgium**
Philip Scheltens N	Alzheimer Center VU University Medical Center of Amsterdam	**The Netherlands**
Wilhelem van de Flier	Alzheimer Center, VU University Medical Center, Amsterdam	**The Netherlands**
Pieter Jelle Visser	Maastricht University Medical Centre	**The Netherlands**
Gunhild Waldemar	University of Copenhagen	**Denmark**
Pietro Tiraboschi	Fondazione Carlo Besta	**Italy**
Sietske Sikkes	Alzheimer Center, VU University Medical Center, Amsterdam	**The Netherlands**
Fabienne COLLETTE	Cyclotron Research Centre and University of Liège	**Belgium**
Jean François Démonet	**Leenaards Memory Centre CHUV Lausanne University Hospital**	**Switzerland**

www.jpnd.eu

JPND research

EU Joint Programme – Neurodegenerative Disease Research